

LEXICONORDICA

18 · 2011

ONOMASIOLOGISKE
ORDBØKER I NORDEN

NORDISK FORENING FOR LEKSIKOGRAFI

LEXICONORDICA

LEXICONORDICA

18 · 2011

ONOMASIOLOGISKE
ORDBØKER I NORDEN

NORDISK FORENING FOR LEKSIKOGRAFI

LexicoNordica 18 · 2011

Onomasiologiske ordbøger i Norden

Hovedredaktører

Henrik Lorentzen (ansvarshavende)

Ruth Vatvedt Fjeld

Nasjonale redaktører

Sturla Berg-Olsen

Ken Farø

Jón Hilmar Jónsson

Nina Martola

Emma Sköldberg

© 2011 LexicoNordica og forfatterne

Omslag og sats: Laurids Kristian Fahl

Trykk: Tarm Bogtryk a-s, Danmark

LexicoNordica trykkes med støtte fra

Nordplus Nordiske Sprog og Kultur

ISSN 0805-2735

INNHold

Ruth Vatvedt Fjeld & Henrik Lorentzen

Onomasiologiske ordbøker i Norden9

Tematiske bidrag

Lars Borin & Markus Forsberg

Swesaurus – ett svenskt ordnät

med fria tyglar 17

Ken Farø

Onomasiologiskhed og leksikografi 41

Anna Helga Hannesdóttir & Sofia Tingsell

Svensk fraseologi för islänningar – om den isländsk-svenska
ordboken i ISLEX och en onomasiologisk svensk resurs:

Frasbanken 61

Jón Hilmar Jónsson & Þórdís Úlfarsdóttir

Íslenskt orðanet: Et skritt mot en allmennspråklig

onomasiologisk ordbok 87

Nina Martola

FinnWordNet och kulturbundna ord 111

Sanni Nimb

Semantiske relationer i en ny dansk begrebsordbog:

genbrug på tværs af ordbøger.....135

Klaas Ruppel

Ontologi, begreppssystem och WordNet157

Lars Trap-Jensen
 Dansk begrebsordbog: baggrund og metode.....183

Lars S. Vikør
 Den første norske tesaurus:
 Ivar Aasens Norsk Maalbunad..... 199

Ikke-tematiske bidrag

Marcin Overgaard Ptaszynski, Mikołaj Sobkowiak
 & *Patrick Leroyer*
 Hvad hedder det på dansk?
 Polske studerendes brug af ordbøger
 mellem polsk og dansk 227

Leiv Inge Aa
 Partiklar inn i ordbøkene?
 Om tradisjonell og nyare kategorisering av preposisjonar249

Anmeldelser

Lars Brink & Ken Farø
 Den svenske forbindelse, version 2010 271

Ilse Cantell & Riitta Korhonen
 Finska ord på finska för invandrare 293

Louise Holmer
 Norstedts ordboksappar 307

Kalevi Koukkunen
 En modern ordbok för alla som läser och forskar i Kalevala..... 323

Sven-Göran Malmgren
 En formidabel metaordbok335

<i>Ole Michael Selberg</i> En etterlengtet ordbok	345
Meldinger	
<i>Oddrun Grønvik</i> Rapport frå den 11. konferansen om leksikografi i Norden	359
<i>Ruth Vatvedt Fjeld</i> Invitasjon til den 15. Euralex-konferansen i Oslo 7.-11. august 2012.....	367
<i>Jan Hoel</i> Nytt fra styret for Nordisk forening for leksikografi	369
Redaksjonelt	373

Onomasiologiske ordbøker i Norden

Ruth Vatvedt Fjeld & Henrik Lorentzen

Nordisk forening for leksikografi presenterer herved sitt 18. nummer i rekken av årsskriftet *LexicoNordica*. Som vanlig inneholder det en tematisk del, denne gang med ni artikler, en ikke-tematisk del med to artikler, og gledeligvis seks anmeldelser av leksikografiske verker.

Den tematiske delens ni artikler handler alle om onomasiologiske ordbøker, i sin alminnelighet og i Norden spesielt. Temaet har vært lite behandlet i den nordiske faglitteraturen tidligere, men med dette nummer av *LexicoNordica* har vi fått en lang rekke presentasjoner av nordiske onomasiologiske ordbøker og leksikografiske prosjekter, samt grundig drøfting av forholdet mellom onomasiologi og semasiologi. Og ikke minst har Ken Farø tatt på seg en stor oppgave ved å drøfte sitt nyslåtte begrep *onomasiologiskhet*, eller med et nordisk ord, *betegnelse*, både i dybden og bredden, nettopp fordi man i onomasiologiske ordbøker tenker ut fra innholdet eller begrepet til den betegnelsen det har fått i et språk. Eller sagt med Dornseiff, en tysk pioner i fagfeltet: Hvilke ord og uttrykk brukes når vi vil uttrykke forskjellig innhold? Farø skisserer noen overordnede perspektiver på fenomenet, samt gir en rekke eksempler på forskjellige typer onomasiologiske ordbøker som er gitt ut i Norden og drøfter deres funksjoner. Artikkelen er en utmerket inngang og overbygning til de øvrige artiklene om onomasiologiske ordbøker.

Ruppel har også tatt et par skritt tilbake og forsøkt å se likheter og forskjeller mellom ontologier, begrepssystem og ordnett både utenfra og ovenfra, og gir en historisk oversikt, samt en oppklaring av de forskjellige funksjoner og termer brukt på feltet. Ulikheten

mellom ordnett og begrepssystem demonstreres tydelig med eksempler fra samisk. Noe av det samme viser Martola med eksempler fra det finske ordnettet, som er oversatt fra engelsk: en onomasiologisk beskrivelse av et ordforråd lar seg vanskelig oversette fra ett språk til et annet uten at det blir skjevheter i målspråket, særlig for kulturrelative ord og uttrykk.

Borin & Forsberg presenterer arbeidet med det svenske ordnettet Swesaurus, samt de ressursene som ligger til grunn for det. Artikkelen drøfter også det sentrale synonymibegrepet i dette og liknende arbeider i et historisk perspektiv. Hannesdóttir & Tingsell drøfter problemet med fraseologi for andrespråksinnlærere med utgangspunkt i ISLEX-prosjektet, som er ment for islendinger som vil lære skandinaviske språk, i denne sammenhengen svensk. De viser bl.a. hvordan en onomasiologisk tilnærming til fraseologien er særlig nyttig for andrespråksinnlærere. Jónsson & Úlfarsdóttir refererer fra arbeidet med et islandsk ordnett med utgangspunkt i hovedforskjellene mellom semasiologiske og onomasiologiske ordbøker. De har endt opp med et ordboksverk som skal omfatte mye mer enn ordnett vanligvis gjør, særlig er fraseologien godt ivaretatt. Målet er å dokumentere alle sider ved det islandske ordforrådet i ett og samme elektroniske ordboksverk, der det er mange forskjellige innganger; slik skal den også kunne fylle flere ordboksfunksjoner.

Både Trap-Jensen og Nimb skriver om arbeidet med den nye danske begrepsordboken, der de begge er redaktører. De viser hvordan man kan bygge nye ordboksressurser med utgangspunkt i tidligere utarbeidede ordbøker i elektronikkens tidsalder. Den danske begrepsordboken bygger først og fremst på *Den Danske Ordbog* og *DanNet*, men også på andre begrepsordbøker. Prosjektet er en måte å gjøre maskinleselige data utnyttbare også for menneskelige brukere. Artikkelen til Trap-Jensen gir en oversikt over forskjeller og likheter mellom begrepsordbøker og ordnett og presenterer det metodiske grunnlaget for begrepsordboken. Nimb ut-

dyper synergien mellom tidligere utarbeidede ordbøker og denne nye, og hvordan man rent praktisk benytter tidligere ordbøkers opplysninger og så beriker dem med ny, ontologisk informasjon. Begge artiklene viser tydelig hvor rikt det leksikalske inventaret i et språk er, og hvor mange perspektiver man kan legge på det i en beskrivelse, pluss at en onomasiologisk vinkling gir helt andre leksikografiske innsikter enn de semasiologiske.

Vikør beskriver bakgrunn og tilblivelse av den eldste norske onomasiologiske ordbok, Ivar Aasens *Norsk Maalbund*, og dens noe triste resepsjonshistorie. Aasen utarbeidet denne ordboka for å få bedre oversikt over forskjellen mellom dansk og norsk semantikk eller begrepssystem, da hans hovedmål var å lage en ordbok over det norske språket uten å måtte gå veien om dansk. Gjennomgangen viser ellers hvordan ordboka avspeiler det rurale samfunnet i Norge i Aasens samtid, noe som samtidig viser hvordan begrepssystem og kultur er tett forbundet.

Den tematiske delen beskjeftiger seg med onomasiologiske ordbøker ut fra en nordisk synsvinkel, og ut fra pågående forskningsprosjekter, slik planen var for *LexicoNordica*-symposiet, som ble holdt i København i januar 2011.

LexicoNordica 18 har også to ikke-tematiske bidrag: Aas artikkel er en grammatisk orientert drøfting av ordklasseinndelingen og ordklassenes funksjon i ordbøker, særlig for klassene preposisjoner, adverb og partikler, og refererer til Faarlunds artikkel om ordklassemarkering i *Norsk Ordbok* i *LexicoNordica 17*. Ptaszynski, Sobkowiak & Leroyer refererer en brukerundersøkelse av polsk-danske ordbøker, med funksjonsteorien i sentrum.

Redaksjonen er også svært glad for å kunne presentere seks anmeldelser av utgitte ordbøker og andre leksikografiske produkter: av tospråklige ordbøker er en svensk-dansk allmennordbok og en polsk-norsk-polsk juridisk ordbok omtalt. Ellers blir en verkordbok om *Kalevala*-eposet og en enspråklig finsk ordbok for innvandrere anmeldt. Og vi har en svensk artikkel om såkalte apper,

små ordbøker og liknende som kan lastes ned til mobiltelefoner og annet kommunikasjonsutstyr, og endelig en omtale av et metaleksikografisk produkt: en omfattende bibliografi over danske ordbøker utgitt mellom 1990 og 2010.

Årskriftet avsluttes med referat fra Nordisk konferanse i leksikografi i Lund i mai 2011, informasjon om kommende Euralex-konferanse i Norge i 2012 og nytt fra Nordisk forening for leksikografi.



Med dette nummer av *LexicoNordica* takker Ruth Vatvedt Fjeld for seg som nasjonal redaktør for perioden 2006–2009, og som en av to hovedredaktører i 2010 og 2011. Fra neste nummer blir det tospannet Henrik Lorentzen fra Danmark og Emma Sköldberg fra Sverige som tar hovedansvaret med tidsskriftet videre.

Og mye mer må vi takke av Nina Martola og Jón Hilmar Jónsson, som har holdt ut som nasjonale redaktører helt siden første nummer av *LexicoNordica* kom ut i 1994, for henholdsvis Finland og Island. Begge to har gjort en formidabel innsats for at nordisk skal omfatte mer enn skandinavisk i leksikografisamarbeidet. Særlig må vi nevne Ninas store innsats med både å fungere som nasjonal redaktør og som oversetter av en lang rekke artikler som opprinnelig ble skrevet på finsk. Det omfatter også de bidrag som er blitt avvist og som ingen har fått se. Ninas innsats har åpnet et vindu for oss andre nordister inn i den finsk-finske faglitteraturen, og det har vært av uvurderlig verdi for et faglig forent Norden. Jón Hilmar har på liknende måte i alle disse årene trukket det litt mer fjerntliggende Island inn i det nordiske fagfellesskapet.

Vi ønsker samtidig nye nasjonale redaktører velkommen, Veturliði Óskarsson fra Island, Mariann Skog-Södersved fra Finland og Annika Karlholm fra Sverige. Hun erstatter Emma Sköldberg, som nå rykker opp som hovedredaktør.

Symposiet på Schæffergården i januar 2011 har fått støtte fra Fondet for dansk-norsk samarbejde med stipendiatpriser på Schæffergården, og utgivelse av tidsskriftet har fått støtte fra Nordplus Nordiske Sprog og Kultur samt Nordisk Sprogkoordination.

Avslutningsvis vil vi melde at temaene for de to neste *LexicoNordica*-symposiene er

2012: Betydningsbeskrivelser i nordiske ordbøker

2013: Leksikografi og lingvistik i Norden

Vi ønsker forslag til foredragsholdere velkommen, det er bare å ta kontakt med de enkelte lands nasjonale redaktør, som sammen med hovedredaktørene setter opp et program for symposiene.

Til sist vil vi takke Birgit Eaker i egenskap av leder for Nordisk forening for leksikografi for stort arbeid med søknader og rapporter i forbindelse med finansiering av symposium og utgivelse av årsskriftet. Og endelig takker vi nok en gang Rikke E. Hauge for utrettelig innsats til beste for *LexicoNordica*-redaksjonen og hele Nordisk forening for leksikografi.

Ruth Vatvedt Fjeld
professor, dr.philos.
Institutt for lingvistiske
og nordiske studier
Universitetet i Oslo
Postboks 1001 Blindern
NO-315 Oslo
r.e.v.fjeld@iln.uio.no

Henrik Lorentzen
seniorredaktør, cand.mag.
Det Danske Sprog- og
Litteraturselskab
Christians Brygge 1
DK-1219 København K
hl@dsl.dk

TEMATISKE BIDRAG

Swesaurus – ett svenskt ordnät med fria tyglar

Lars Borin & Markus Forsberg

Swesaurus is a free Swedish wordnet currently under construction in Språkbanken, University of Gothenburg. Swesaurus is constructed by reusing information about lexical-semantic relations in a number of pre-existing freely available lexical resources: SALDO, SDB, Synlex and Swedish Wiktionary. In this article we describe how this existing lexical-semantic information is reused and enriched by utilizing the connections among the lexical resources. We focus in particular on our experiments with turning the graded synonymy relations of Synlex into *fuzzy synsets* in Swesaurus. The introduction of fuzziness into a wordnet raises many intricate methodological and theoretical questions, which are discussed and illustrated in the article.

1. Introduktion

Nedan beskriver vi ett pågående arbete för att skapa ett fritt svenskt ordnät. Detta görs till stor del genom återanvändning av ett antal befintliga fria lexikonresurser. På grund av strukturen hos dessa resurser får det nya ordnätet några i sammanhanget ovanliga och intressanta karakteristika, bland annat associativa relationer mellan ord och variabel, kvantifierad synonymi, av vilka den sistnämnda står i fokus i denna artikel.

1.1. Ordnät

De s.k. *ordnäten* har på relativt kort tid blivit centrala lexikonresurser för språkteknologitillämpningar. Det första ordnätet var det

engelska *Princeton WordNet* (PWN), som började utvecklas i mitten av 1980-talet av en forskargrupp vid Princeton-universitetet i USA under ledning av George A. Miller (Fellbaum 1998; PWN 2011). Idag finns ordnät för ett antal språk (se GWN 2011), och många av dem är fritt tillgängliga under samma typ av licens som det ursprungliga PWN, som tillåter alla slags användningar liksom vidareutveckling och återpublicering utan några restriktioner.

1.2. Ordnärens idé och organisation

Ordnäten är onomasiologiska lexikonresurser. Deras uppbyggnad motiverades ursprungligen inte av lexikografiska eller teoretisklingvistiska hänsyn, utan av psykolingvistiska och kognitionsvetenskapliga idéer om det mentala ordförrådets organisation. Grundenheterna i ett ordnät är något som man skulle kunna kalla *lexikaliserade begrepp*, men som man normalt utan förlust av precision kan tänka på som ordbetydelser – ett Saussureanskt språkligt tecken, alltså en kombination av språkligt uttryck (ett lexikonord) och innehåll. Ordbetydelserna organiseras i synonymmängder (synsets) enligt en speciell definition av synonymi (se avsnitt 2). Synonymmängder och lexikonord kopplas samman med traditionella – i litteraturen om ordnät ofta kallade ”klassiska” – lexikalisk-semantiska relationer, av vilka den viktigaste, utöver synonymi, är hyponymi och därmed automatiskt dess omvändning hyperynymi. En lexikonresurs måste minimalt ha synonymmängder hierarkiskt organiserade av hyponymirelationer för att räknas som ett ordnät (GWN 2011).

PWN innehåller enbart de fyra öppna ordklasserna substantiv, verb, adjektiv och adverb. På grund av vad som verkar vara en specifik anglosaxisk lexikografisk tradition (Apresjan 2002) finns dock även räkneord med i PWN, klassificerade som substantiv och adjektiv.

1.3. Mot ett fritt svenskt ordnät: *Swesaurus*

För svenska finns för närvarande inte något ordnät som är tillgängligt med en liknande typ av licens som PWN, vilket hämmar utvecklingen av en rad språkverktyg för svenska som är självklart tillgängliga för engelska.¹

Swesaurus syftar till att fylla den luckan för svenskans vidkommande. Det är ett svenskt ordnät som för närvarande är under utveckling i Språkbanken (SB 2011) vid Göteborgs universitet och som görs tillgängligt under en typ av öppen källkodslicens. *Swesaurus* är delvis ett resultat av kombinationen av redan existerande, fria resurser. De fyra centrala befintliga resurserna är SALDO, SDB, Synlex och Wiktionary. I avsnitt 3 beskriver vi dessa resurser och hur informationen i dem används för att bygga upp *Swesaurus*. Först behöver vi dock diskutera den i det här sammanhanget centrala termen *synonymi* lite närmare, vilket vi gör i nästa avsnitt.

2. Synonymi

Synonymi är den centrala lexikalisk-semantiska relationen i ett ordnät. Synonymi definieras så här i PWN: Om två ord är utbytbara i minst en mening utan att meningens betydelse – i termer av sanningsvillkor – förändras, så uttrycker de två ordbetydelsorna samma begrepp och bildar därmed en synonymmängd. Det grundläggande arbetsmomentet när ett ordnät ska byggas är alltså att bestämma vilka ordbetydelser som är synonyma i den här bemärkelsen.

1 I början av 2000-talet bedrevs i Sverige ett projekt finansierat med offentliga medel för att bygga en svensk motsvarighet till Princeton WordNet. Projektet resulterade i ett svenskt ordnätfragment med i storleksordningen 25.000 synonymmängder. Detta ordnät verkar inte ha utvecklats vidare sedan mitten av 2000-talet och det är explicit inte tillgängligt under villkor som skulle göra det möjligt att bygga vidare på det i arbetet med *Swesaurus*.

Det här är ett rätt speciellt sätt att förstå synonymi och en intressant fråga som uppkommer i det sammanhanget är huruvida man i olika ordnäsprojekt förstår synonymi på samma sätt som PWN:s skapare. Svaret är faktiskt nej, åtminstone när det gäller det danska ordnätet DanNet (Pedersen et al. 2009) och det ena av de två polska ordnäten (Piasecki et al. 2009). En förklaring till det kan vara att de två nämnda projekten – i motsats till PWN och i motsats till ordnät som har skapats genom översättning av PWN, som det finska ordnätet (Lindén & Carlson 2010) – har utgått ifrån starka lexikografiska traditioner där man redan av hävd har en annan idé än i PWN-projektet om vad som inryms i begreppet synonymi. PWN-definitionen av synonymi – eller snarare dess konsekvenser – att synonymi används för att definiera den grundläggande lexikonenheten, så att man får lika många lexikonenheter som man kan hitta (på) utbyteskontexter – känns i själva verket lite främmande mot en sådan bakgrund. Dessutom kan man notera en vacklan även vad beträffar PWN. Även om man i litteraturen alltid framhåller synonymmängden (synset) som den grundläggande enheten i PWN och andra ordnät, kan detta inte vara hela sanningen: mer än hälften av alla synonymmängder i PWN 3.0 (54 %) har bara en medlem – som därmed rimligen inte uppvisar utbytbarhet mot något annat ord i en enda kontext – varför man måste tänka sig att det finns ytterligare sätt att avgöra vilka ordbetydelser man har att räkna med i ett ordnät. Inget annat ordnät torde ha lika stor betydelsesplittring som PWN, där (grundformen) *break* i PWN 3.0 innehar rekordet med 75 betydelser, fördelade på 59 verb och 16 substantiv.² En sådan stor mängd

2 Vår jämförelsegrund är dock uttryckligen andra ordnät och andra lexikonresurser för språkteknologi. Jämfört med konventionella engelska lexikon är PWN inte anmärkningsvärt. Trots sina psykologvristiska anspråk följer PWN här som i en del andra avseenden helt enkelt en engelskspråkig lexikografisk tradition: stora engelska konventionella lexikon anger ännu fler betydelser för *break* (107 betydelser i DICTIONARY.COM 2011, varav 71 verb och 36 substantiv).

betydelser är konstaterat svår att differentiera maskinellt, vilket har konsekvenser för automatisk textanalys som använder sig av PWN.

Synonymordlistor har gjorts lika länge som man överhuvudtaget har sammanställt ordlistor och ordböcker; redan sumerer och akkadier lämnade efter sig synonymlistor (Civil 1990) och latinsynonymiken var en på sin tid blomstrande lexikongenre som gick tillbaka till det första århundradet av vår tideräkning och som fortfarande var levande under det sena 1800-talet, en tidsrymd om nästan två årtusenden (Svensén 2010). Det är alltså inte ägnat att förvåna att synonymi med tiden har kommit att förstås på många olika sätt. Ytterligheterna illustreras väl med citat ur två olika latinsynonymiker från artonhundratalets mitt:

Still it ought to be remembered, that there are actually words which differ simply as to form or sound, and by which the scholar is strangely misled, if he starts with the axiom, that there are no two words meaning exactly the same thing, in the same language, – an error, it seems, which may be perceived in far the greater number of works on synonymes. That there are equivalent words may be seen at once, if we remember, that some Latin words end both in *is* and *us*, without a shadow of difference in meaning. (Lieber 1841:vii)

If, therefore, it be admitted, that words identical in meaning do not exist, and that it is morally impossible, if I may use the expression, that they should exist, the only questions are, whether, in such cases, it is worth while to search out their differences, and whether it is possible to find them out. Science will answer the first question, without hesitation, in the affirmative; and with respect to the second, there can at least be no presumption in making the attempt. (Döderlein 1863:xii)

Å ena sidan är det självklart att det finns ord med exakt samma betydelse (Lieber), å den andra är det ”moraliskt omöjligt” att sådana ord skulle existera (Döderlein). I det här avseendet har ingenting ändrats i sak sedan 1800-talet. Även om den lingvistiska argumentationen har blivit mycket mer sofistikerad, finns det lika många bud nu som då om hur synonymi ska förstås och extremståndpunkterna är fortfarande att synonymi är antingen icke-existerande eller ofrånkomlig (se Murphy 2003, kap. 4).

I praktiken verkar både lexikografer och många ordnätssmakare – som i de ovan nämnda danska och polska ordnätprojekten – använda en mer traditionell definition av synonymi än den som används i PWN, nämligen en där lexikonets ord(betydelser) fastställs först, på andra, oberoende grunder, och synonymi sen blir en relation mellan sådana lexikonord, som oftare i själva verket motsvarar partiellt snarare än totalt sammanfall i betydelse. Mot den bakgrunden kunde det vara intressant att försöka kvantifiera graden av överensstämmelse i betydelse mellan synonymer i denna bemärkelse. Det är möjligt att man skulle kunna företa en sådan kvantifiering med PWN, genom att räkna antalet synset som två ord uppträder i samtidigt (om det finns sådana synset). Såvitt vi känner till har detta dock inte gjorts.

En sådan kvantifiering öppnar för möjligheten att skapa ordnät där synonymmängderna kan bli större eller mindre beroende på hur strikt man förstår synonymi. Därmed skulle man kunna komma åt något som ofta har påpekats som ett praktiskt problem med PWN, nämligen att det har så fin betydelseindelning att de metoder man för närvarande förfogar över för automatisk textanalys inte klarar av att tilldela PWN-betydelser till ordförekomster i texter. Ofta efterfrågas i sådana sammanhang ett principiellt välgrundat och pålitligt sätt att slå ihop PWN-betydelser (och därmed få färre synonymmängder). PWN:s struktur erbjuder ingen hjälp därvidlag, eftersom de betydelse som hör till samma grundord inte är hierarkiskt ordnade i huvudbetydelser och underbetydelser. Se t.ex. Palmer et al. (2006) och referenser där.

3. Från SALDO, SDB, Synlex och Wiktionary till Swesaurus

I detta avsnitt presenterar vi de fyra resurser som ligger till grund för Swesaurus, tillsammans med en översiktlig beskrivning av arbetet med att inkludera de lexikalisk-semantiska relationerna från dessa resurser i Swesaurus.

Vi nämnde i avsnitt 1 ovan att dessa resurser är fria, med vilket vi helt enkelt menar att de har användningsvillkor (licenser) som tillåter vem som helst att använda resurserna till vilket ändamål som helst (inklusive kommersiell användning), samt även modifiera eller vidareutveckla och återpublicera dem, med det enda villkoret att de modifierade och återpublicerade resurserna inte får beläggas med restriktivare användningsvillkor än de ursprungliga resurserna. Detta är vad som i mjukvaruvärlden kallas *öppen källkod* ('open source') och vad man ibland för andra typer av resurser än datorprogram kallar *öppet innehåll* ('open content').³ Om man som vi vill att slutprodukten ska vara fritt tillgänglig på detta sätt, måste även alla ingående delresurser vara det. Detta begränsar naturligtvis vilka andra lexikonresurser vi kan använda i arbetet med Swesaurus, men det är essentiellt för vårt forskningsområde att vi kan bygga den typ av öppen språkteknologiinfrastruktur som PWN och andra fria resurser är goda exempel på. Medaljens baksida kan tyckas vara att vi därmed måste avstå från att använda högvärdiga lexikaliska databaser som av kommersiella och andra skäl inte är fritt tillgängliga i den mening som avses här. Vi måste naturligtvis respektera att många vill ha kontroll över sina lexikondata av olika skäl. Samtidigt behöver svensk språkteknologi fria resurser för att växa, vilket betyder att det egentligen i praktiken inte finns något val: ur vårt forskningsområdes synvinkel existerar helt enkelt inte dessa resurser.

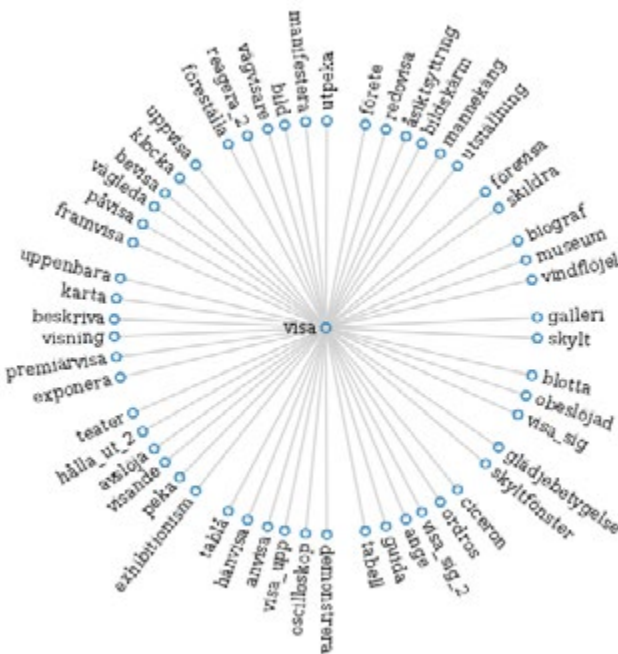
3 Det finns en hel rad sådana licenser; de som vi använder för våra lexikonresurser heter GNU LGPL och Creative Commons BY-SA.

3.1. SALDO

SALDO är en fullskalig (>100.000 ordbetydelser) lexikalisk-semantic – onomasiologisk – resurs för svenska. SALDO uppdateras och utvidgas kontinuerligt. Arbetet med SALDO sker väsentligen i öppenhet: utvecklingsversionen publiceras dagligen dels som nedladdningsbar resurs, dels via en uppsättning sökgränssnitt (SALDO 2011). De nedladdningsbara utvecklingsversionerna görs tillgängliga via ett versionshanteringssystem, ett system som möjliggör att varje enskild version går att återskapa. Detta är en viktig egenskap i en resurs även avsedd för användning inom forskning, där reproducerbarhet är ett centralt metodologiskt krav.

Grundenheterna – (lexikon)ingångarna – i SALDO är ordbetydelser som ordnas genom en grundläggande semantisk relation som vi kallar *association*. I likhet med PWN innehåller SALDO en stor mängd flerordsenheter och egennamn, men till skillnad från PWN beskriver SALDO även ord ur de slutna ordklasserna (pronomen, prepositioner, subjunktioner, etc.). Varje ordbetydelse beskrivs genom en association i form av en *primär deskriptor* eller ’moder’. Den primära deskriptorn ska vara den semantiskt mest närliggande ordbetydelse som också är centralare än den ordbetydelse som ska beskrivas. Det senare kriteriet tillsammans med kravet att alla deskriptorer i sin tur ska ha en beskrivning i SALDO (med ett undantag; se nedan) ger SALDO en hierarkisk struktur. Kriterierna är inte formellt definierade, utan den som arbetar med att definiera nya ingångar i SALDO får använda sin språkliga intuition och sitt omdöme. Vad gäller centralitet finns dock några typiska kännetecken, såsom att en mer central ordbetydelse har *högre korpusfrekvens*, är *morfologiskt enklare*, är *stilistiskt neutralare*, är *mindre ämnesbunden* och *uppträder tidigare i barns språkutveckling* än en mindre central betydelse. Några exempel på primära deskriptorer är *pojke* för *kille*, *sannolik* för *sannolikhet*, och *leva* för *liv*.

När det gäller att avgöra vilken den mest närliggande ordbetydelsen är, visar praktiken att den ständiga interaktionen med lexikonmaterialet ger en bra vägledning. Ordbetydelser med samma primära deskriptor bör när man ser dem samlade bilda en tydlig 'semantisk familj', en gestalt, och det är relativt lätt att se om något eller några ord sticker ut ur mängden (se figur 1). SALDO kan ses som resultatet av många lokala beslut, som dessutom ständigt revideras.



Figur 1: Ordbetydelser i SALDO med *visa* som primär deskriptor.

Det finns ett litet antal – 43 stycken – ordbetydelser som inte har någon primär deskriptor i SALDO. Dessa förses med en artificiell 'rotbetydelse' eller 'nollbetydelse' kallad *PRIM* för att SALDO formellt ska bilda en enda trädstruktur (istället för 43 trädstrukturer). *PRIM* har ingen beskrivning i SALDO.

Över hälften av alla ordbetydelser i SALDO har enbart den primära deskriptorn, men för att differentiera betydelsebeskrivningen kan man ange en eller flera *sekundära deskriptorer* eller 'fäder'. Exempelvis har både *menageri* och *anekdotflora* den primära deskriptorn *samling*, men de skiljs åt i den sekundära deskriptorn: *djur* för *menageri*, *anekdot* för *anekdotflora*. Av de sekundära deskriptorerna krävs inget annat än just att de hjälper till att skilja åt ordbetydelser som delar primär deskriptor. Därför är de sekundära deskriptorerna ofta semantiskt mindre centrala än den betydelse som ska beskrivas. Ungefär en femtedel (strax över 9.000) av de sekundära deskriptorerna är mindre centrala – har en primär deskriptor längre bort från PRIM – än den betydelse de beskriver.

En primär deskriptor är ofta en hyperonym eller synonym (eller i sällsynta fall en hyponym) till den beskrivna ordbetydelsen. Därför kan SALDO användas vid uppbyggnaden av Swesaurus. Det är enkelt att ta ut alla par i SALDO där lexikoningången och dess primära deskriptor har samma ordklass. Specialfallet sammansättningar, där efterledet i sammansättningen typiskt är dess primära deskriptor, kan särbehandlas genom att mekaniskt kontrollera att lexikoningångens grundform slutar med den primära deskriptorns grundform. På detta sätt samlar man synonym/hyponymkandidatlistor som snabbt kan gås igenom för hand. Detta är ett arbete som är påbörjat.

Swesaurus kommer som en slags bonus även att innehålla de associativa relationerna från SALDO, vilka kan inkluderas utan något extra arbete eftersom en enda uppsättning formella ordbetydelseidentifierare – nämligen SALDOs – används i alla våra lexikonresurser (Borin 2010). Det betyder att Swesaurus redan från början får en typ av relationer gratis som man nu, intressant nog, lägger ner stort arbete på att lägga till i Princeton WordNet och som man i det sammanhanget kallar för "framkallande" (eller "frammanande": "evocation"; Boyd-Graber et al. 2006; se även Morris & Hirst 2004).

SALDO följer i stort sin föregångare SAL när det gäller den semantiska organisationen. Den som vill ha en mer ingående beskrivning av SALDO hänvisar vi till Borin (2005) för den semantiska organisationen och Borin et al. (2008) för övriga aspekter av SALDO.

3.2. SDB

SDB (semantisk databas; Järborg 2001) är en lexikalisk databas med 68.000 ordbetydelser, av vilka ett antal är mycket detaljerat beskrivna med avseende på sin semantiska valens, som anges med användning av ett fyrtiotal semantiska roller. Många av ordbetydelserna är också länkade till ordförekomster i en svensk standardkorpus. Delar av informationen i SDB ska användas vid uppbyggnaden av ett svenskt frasnät (Borin et al. 2010).

För vårt arbete med Swesaurus är dock det intressantaste med SDB att resursen tillhandahåller information om en rad klassiska lexikalisk-semantiska relationer mellan ett stort antal av sina lexikonringångar: synonymi, hyponymi, kohyponymi, antonymi och en ’allmän semantisk relation’ (”se”). Den sistnämnda anger helt enkelt något slags semantiskt samband, utan att försöka specificera det närmare, inte olikt SALDOs associationer, men inte nödvändigtvis med samma krav på att peka på en centralare betydelse. Ett exempel är golftermen *albatross* i betydelsen ”tre slag under par” som står i ”se”-relation till en annan golfterm, *par*.

En relation kan vara explicit angiven eller logiskt härledbar. Exempelvis betraktas ofta synonymi som en *diskret* – till och med *binär* – *transitiv relation*. Att relationen är binär betyder att den antingen råder eller inte; man räknar inte med några grader av synonymi.

Transitivitet betyder i den tekniska bemärkelse som vi använder den här följande. Om vi vet att en viss transitiv relation råder mellan objekten A och B och att samma relation råder mellan ob-

jekten B och C, så kan vi därmed sluta oss till att relationen råder även mellan A och C. Transitivitet kan alltså sägas vara den formella versionen av en utsaga som ”Min väns vän är min vän”. En prototypisk transitiv relation i den fysiska världen är ”större än”. Om vi får veta att en elefant är större än en kamel och att en kamel är större än en koala, så vet vi också automatiskt att en elefant är större än en koala, utan att detta behöver sägas explicit.

Att synonymi är transitiv innebär alltså att om vi vet att A är synonymt med B och B med C, så följer med automatik att A är synonymt med C, eller med andra ord, synonymparet A–C kan härledas från de explicit angivna synonymparen A–B och B–C. Den mängd man får om man gör detta uttömmande för en transitiv relation kallas *det transitiva höljet*, ett begrepp som kommer att vara centralt när vi nedan diskuterar användningen av graderade synonympar ur Synlex (se avsnitt 4).

Det transitiva höljet är helt enkelt den mängd av objekt som man får om man explicit räknar ut alla transitivetsrelationer givet något eller några av objekten. Om vi igen använder exemplet ”Min väns vän är min vän”, så får jag det transitiva höljet genom att först räkna upp mina vänner, för var och en av dem hennes eller hans vänner, för var och en av dem hans eller hennes vänner, etc., tills jag har uttömt alla möjligheter. Mängden av individer som jag har fått fram på det sättet är det transitiva höljet med avseende på relationen ”vän med” och utgående från mig själv.

I vårt fall kan man tänka på det transitiva höljet som en ’kedja’ av ordbetydelser där varje länk i kedjan ges av ett synonympar, eller med ett konkret exempel, det transitiva höljet för synonymparen *barn–parvel*, *parvel–pys* och *pys–knatte* är synonymmängden {*barn, parvel, pys, knatte*}.

Synonymi är även en *symmetrisk* relation: om A är synonymt med C, så följer även att C är synonymt med A. Hyponymi och hyperonymi är varandras *inverser*: om A är en hyponym till B, så är B en hyperonym till A.

Med kännedom om hur de lexikalisk-semantiska relationerna förhåller sig logiskt till varandra, kan man använda dem och deras kombinationer för att 'fylla på' ofullständig information hämtad ur en lexikonresurs som SDB, där inte alla logiskt möjliga synonym- och hyponym/hyperonympar anges explicit, utan bara ett litet fåtal av dem.

Den allmänna strategin för att använda den lexikalisk-semantiska informationen i SDB blir då följande: (1) extrahera relevanta ordbetydelsepar i SDB; (2) beräkna transitiva höljet med avseende på synonymi, samt beräkna symmetri och invers för att därmed utöka till alla möjliga par; (3) undersöka resultatet manuellt för att hitta iögonenfallande fel. Detta arbete är påbörjat.

3.3. Synlex

Synlex (Kann & Rosell 2006; SYNLEX 2011) är en graderad svensk synonymlista skapad av Viggo Kann och hans kollegor vid KTH i Stockholm. De har utvecklat ett system där användare av ett svenskt-engelskt nätlexikon ombeds att bedöma graden av synonymi för ordpar som genererats automatiskt ur tvåspråkiga lexikon och ur korpusar, på en skala från 0 (ej synonyma) till 5 (fullt synonyma). Paren är ordformer utan ordklassangivelse, huvudsakligen grundformer, men även böjda former och flerordsuttryck förekommer. Några exempel på ordpar tagna direkt från Synlex system är *stark–högaktiv*, *fullständig–rätt*, *ram–gränsområde* och *design–mönster*.

En fritt nedladdningsbar delmängd av Synlex publiceras med oregelbundna mellanrum på projektets hemsida (SYNLEX 2011). Datamängden innehåller ordpar som har fått ett minimum av tre bedömningar och som har en genomsnittlig synonymigrad om minst 3.⁴ Versionen daterad 2010-09-19 innehöll c:a 19.000 grade-

4 Synlexgraderingen 3–5 multipliceras med 20 i Swesaurus, vilket ger ett intervall 60–100 som man med fördel kan tänka på som en procentangivelse, där 100 representerar full synonymi.

rade synonympar (det dubbla om man tar hänsyn till synonymi-relationens symmetri).

Hittills har vårt arbete med Swesaurus fokuserat mest på just Synlex som informationskälla. Detta arbete beskrivs utförligare i avsnitt 4.

3.4. Wiktionary

Wiktionary är en ansats liknade Wikipedia men för kollaborativt författande av nätlexikon snarare än av nätuppslagsverk. Svenska Wiktionary (WIKTIONARY 2011) är en nedladdningsbar fri resurs som – i tillägg till mycket annat – innehåller en hel del lexikalisk-semantiska relationer. Arbetet med att extrahera sådana relationer ur Wiktionary försvåras dock av att datamängden är enbart delvis formellt strukturerad. Det kommer nämligen an på författaren av en lexikoningång att använda det givna wikiformatet för att koda de olika informationskategorierna i ingången på det sätt som Wiktionarys upphovsmän har avsett, men det görs ingen kontroll av att så faktiskt sker. Eftersom resultatet av en felaktig kodning kan se rätt ut för det mänskliga ögat, finns det i praktiken en rad formella fel i Wiktionary som komplicerar den automatiska informationsutvinningen.

Vi har experimenterat med att extrahera synonymirelationer mellan ord, vilket hittills resulterat i 4.012 synonympar, varav 1.514 är ordpar där vardera medlemmen i paret motsvarar endast en betydelse i SALDO. Därmed behövs ingen manuell disambiguering, utan dessa kan införlivas i Swesaurus direkt. En del av synonymparen i Wiktionary är felaktiga, vilket kommer sig av att vissa ordbeskrivningar även innehåller information om andra språk och relationer inom dessa. Detta resulterar i ett fåtal fall där exempelvis ett svenskt ord länkas till ett polskt. I praktiken innebär detta inget problem, eftersom länkningen till SALDO även innebär att ord som inte förekommer i SALDO filtreras bort.

Synonymirelationerna i Wiktionary är generellt av högre kvalitet än i Synlex, vilket är att förvänta sig eftersom författaren av en lexikongång i Wiktionary gör ett aktivt val när hon anger synonymer till det ord som beskrivs, medan Synlex bygger på bedömningar av automatiskt genererade ordpar, vilket resulterar i att ord som vanligtvis inte anses vara synonyma ändå kan komma att få en gradering större än noll. Ta ord som *förlovning* och *förpliktelse*, vilka vanligen inte anses synonyma, men när man får dem presenterade tillsammans med en uppmaning att kvantifiera synonymigraden mellan dem, lockas man nog till att ange ett visst mått av synonymi. Medaljens baksida är att det finns betydligt färre synonympar i Wiktionary, vilket också är det förväntade.

Det svenska Wiktionary har som tidigare nämnts ordbeskrivningar som innehåller lexikal information om andra språk, i huvudsak engelska. Det finns översättningslänkar mellan svenska och engelska som kan utnyttjas för att skapa ett svenskt-engelskt lexikon. Ett arbete som är knappt påbörjat är att undersöka om den svensk-engelska informationen kan användas för att skapa en första approximation till länkning av Swesaurus till Princeton WordNet.

4. Graderad synonymi och luddiga synonymmängder

Som vi nämnde i avsnitt 2 är det som kallas synonymi normalt närsynonymi, så ett försök att gradera synonymi som i Synlex tycks intuitivt korrekt. Det finns i princip ett otal tänkbara sätt att beräkna denna synonymigrad beroende på vår uppfattning om vad synonymi är och vad vi anser vara en ordbetydelse. Men varför inte försöka använda synonymigraderna i Synlex, eftersom de är fritt tillgängliga?

Frågan är hur man förenar Synlex graderade synonymi med

det synonymibegrepp som konstituerar ett ordnät som PWN, som ju består av synonymmängder och lexikalisk-semantiska relationer mellan dessa mängder. Som vi redan har nämnt så ses synonymi i många sammanhang – inklusive PWN – som en binär, transitiv relation. Därmed ges en synonymmängd i normala fall av det *transitiva höljet* (se avsnitt 3.2).

Frågan kompliceras dock av graderingen, alltså att vi oftast har att göra med närsynonymi snarare än total synonymi, eftersom det inte är självklart att närsynonymi är transitiv. Om A är en närsynonym till B med graden x och B till C med graden y , följer det då med nödvändighet att A är närsynonym till C, och i så fall med vilken grad? Ett exempel taget från Synlex är ordparen *precis-absolut* med graden 60, och *absolut-fullkomlig* med graden 62, där frågan vi ställer oss är om det därmed är givet att *precis-fullkomlig* är närsynonymer, och om så är fallet, med vilken grad? Införandet av graderad synonymi i ett ordnät får också konsekvenser för andra relationer, exempelvis för hyperonymi–hyponymi och holonymi–meronymi, eftersom dessa relationer i PWN:s lexikalisk-semantiska modell råder mellan synonymmängder och inte mellan ord. En annan fråga man därmed måste ställa sig är hur graderad synonymi påverkar dessa andra relationer. Exempelvis, om A är synonym med B med graden x , och C är en hyperonym till A, är det då nödvändigtvis så att även C är en hyperonym till B? Det enda sättet att tackla dessa frågor är att empiriskt testa olika antaganden och granska konsekvenserna.

4.1. Synonymmängder som kedjor av synonympar

Vår första ansats var att experimentera med transitiva höljet tillsammans med olika brytpunkter, där enbart synonympar med en gradering vid eller ovan brytpunkten tilläts bilda länkar i kedjan. Till exempel skulle en brytpunkt på 90 endast tillåta ordpar med en gradering på 90 eller mer. En intressant följd av att använda brytpunk-

ter är att det ger upphov till en mängd ordnät, ett ordnät per brytpunkt, eller vad man skulle kunna kalla ett parametriserat ordnät.

Resultatet blev å ena sidan en uppsättning av rimliga synonymmängder, men å den andra även en återstående ensam synonymmängd med flera tusen ordbetydelser. Denna onormalt stora synonymmängd uppstår på grund av: (1) vi har ett betydelsekontinuum; (2) Synlex innehåller ordpar som kanske inte ska räknas som synonymrelationer; (3) vissa ordbetydelser saknas i SALDO. Antagandet om entydighet mellan ordform och ordbetydelse i SALDO är naturligtvis giltigt endast under förutsättning att alla betydelser kan antas beskrivna i SALDO. Så är det förstås inte, utan i praktiken ser vi en växelverkan mellan de lexikonresurser som vi arbetar aktivt med (främst SALDO, Swesaurus och det svenska frasnätet), så att nya ordbetydelser för befintliga grundformer ständigt får läggas till i SALDO.

Vi lade sedan till kravet att en synonymmängd endast får bestå av ord i samma ordklass – något som ofta anses gälla för synonymi och som alltid gäller i PWN-modellen – vilket minskade den största synonymmängden, men antalet ordbetydelser i den kan fortfarande räknas i tusental.

4.2. Synonymmängder där alla känner alla

Vårt nästa experiment var en mer konservativ ansats: Vi krävde att en synonymmängd ska vara en *klick*. En synonymmängd är en klick när alla ingående ordbetydelser är (på förhand angivna som) synonyma med varandra, vilket sammanfaller med vad vi intuitivt menar med synonymmängd. För det transitiva höljet krävde vi ju bara en kedja av (på förhand givna) synonymirelationer från den första till den sista medlemmen i synonymmängden, och härledde sedan resterande relationer, t.ex. den mellan kedjans första och sista länk. Här kräver vi istället att alla synonymirelationer i mängden redan är explicit givna.

Om vi igen använder exemplet med ”Min väns vän är min vän”, så måste vi i detta fall lägga till ”men bara om hon själv säger både att hon är min väns vän och att hon är min vän”; vi får alltså nu bara samla ihop individer som alla säger sig vara vänner med alla andra i gruppen. Det inses lätt att den gruppen i normalfallet blir betydligt mindre än det transitiva höljet.

Beräkningen av klickar gav inga onormalt stora synonymmängder, men däremot ett par andra problem. Det första problemet är att ordbetydelser stundtals förekommer i fler än en synonymmängd, vilket strider åtminstone mot vad man i ordnätssammanhang brukar mena med ordbetydelse. Typiskt skiljer sig dessa synonymmängder åt endast med avseende på en eller ett par ordbetydelser – en indikation antingen på saknade synonymrelationer eller på saknade ordbetydelser i SALDO.

Naturligt nog uppstår då frågan om hur vi härleder saknade (när)synonympar, så att en ordbetydelse endast förekommer i en synonymmängdsklick. En del av det arbetet är att samtidigt identifiera saknade ordbetydelser.

En möjlig tanke är att härleda alla implicita par i det transitiva höljet som vi därefter går igenom manuellt, men det ger oss en ohanterlig mängd par. Den största synonymmängd som vi fick genom att använda det transitiva höljet innehöll 5.770 ordbetydelser. När varje ordbetydelse paras ihop med varje annan ordbetydelse i synonymmängden motsvarar det $5.770 \cdot 5.769 = 33.287.130$ synonympar. Antalet par kan förvisso halveras, eftersom om vi har A–B så behövs inte den omvända relationen B–A, men att ’bara’ behöva gå igenom knappt 17 miljoner synonymparskandidater manuellt hjälper oss inte mycket i sammanhanget.

En mer försiktig härledning av nya synonympar är att utgå från de klickar som har en eller flera ordbetydelser gemensamma och härleda de par som, om de existerade, skulle förena klickarna till en klick. Metoden kan upprepas över de sålunda sammanslagna klickarna, och därmed härleda nya par. Strategin ger upphov till

par av god kvalitet, speciellt om vi begränsar oss till det relativt vanliga fallet med klickar där alla medlemmar utom en är gemensam. Då leder begränsningen dock givetvis till att endast klickar av samma storlek kan förenas genom denna strategi.

5. Sammanfattning och utblick mot framtiden

Arbetet med Swesaurus bygger vidare på vad andra åstadkommit, vilket gör att vi hoppas kunna skapa en omfattande resurs av godtagbar kvalitet med en i sammanhanget relativt blygsam arbetsinsats.

Swesaurus blir således ett fritt svenskt ordnät med ett par i dessa sammanhang ovanliga egenskaper: Det kommer till en del att ha graderade synonymmängder, vilket i praktiken gör det till lika många ordnät som man väljer brytpunkter för graden av synonymi, och det kommer att ha SALDOs associativa relationer i tillägg till de klassiska lexikalisk-semantiska relationerna som vi förknippar med ett ordnät.

Det återstår ett antal teoretiska och metodologiska frågor att ta ställning till som har med graderingen att göra, som vi ännu inte rört vid. Särskilt kan här nämnas vilka effekter luddiga synonymmängder har på andra relationer, om några, och hur vi ska åstadkomma någon form av gradering för synonympar hämtade från någon annan resurs än Synlex. Empiriskt verkar åtminstone en tvågradig skala fungera för det praktiska arbetet med synonymparskandidater både ur SALDO (se avsnitt 3.1) och ur SDB (se avsnitt 3.2); det visar sig vara naturligt och relativt enkelt att klassificera dessa i *närsynonymer* (som tentativt åsätts 90 % synonymi) och (fulla) *synonymer* (och naturligtvis par som inte är någondera).

Vi har börjat utforska automatiska metoder för att göra kvalificerade gissningar om kvaliteten av existerande och härledda

par. En metod vi ska titta närmare på är om avståndet mellan två ordbetydelser i SALDOs hierarki kan användas för att göra detta. Detta är ett dubbelriktat arbete, eftersom om vi har ett stort avstånd, men synonymparet visar sig vara av god kvalitet, kan det i sin tur leda till en revidering av SALDO.

Vi planerar att släppa Swesaurus 1.0 i slutet av 2011, men redan nu är utvecklingsversionen sökbar och nedladdningsbar via Språkbankens lexikala system SBLEX (SBLEX 2011). Utvecklingsversionen uppdateras varje natt, vilket inbjuder till insyn och deltagande i vårt dagliga arbete. Om vårt arbete av någon anledning skulle upphöra, så kan dessutom därmed någon annan ta upp tyglarna och fortsätta precis från den punkt där vi släppt dem.

Litteratur

- Apresjan, Yuri D. 2002: Principles of systematic lexicography. I: Marie-Hélène Corréard (red.): *Lexicography and natural language processing. A Festschrift in honour of B. T. S. Atkins*. Grenoble: Euralex, 91–104.
- Borin, Lars 2005: Mannen är faderns mormor: Svenskt associationslexikon reinkarnerat. I: *LexicoNordica* 12: 39–54.
- Borin, Lars 2010: Med Zipf mot framtiden – en integrerad lexikonresurs för svensk språkteknologi. I: *LexicoNordica* 17: 35–54.
- Borin, Lars & Markus Forsberg 2009: All in the family: A comparison of SALDO and WordNet. I: *Proceedings of the Nodalida 2009 Workshop on WordNets and other Lexical Semantic Resources – between Lexical Semantics, Lexicography, Terminology and Formal Ontologies*. Odense: NEALT, 7–12.
- Borin, Lars, Markus Forsberg & Lennart Lönngren 2008: The hunting of the BLARK – SALDO, a freely available lexical database for Swedish language technology. I: Joakim Nivre, Mats Dahllöf and Beáta Megyesi (red.): *Resourceful language tech-*

- nology. *Festschrift in honor of Anna Sågvald Hein*. Acta Universitatis Upsaliensis: Studia Linguistica Upsaliensia 7. Uppsala: Uppsala University, 21–32.
- Borin, Lars, Dana Danélls & Markus Forsberg, Dimitrios Kokkinakis & Maria Toporowska Gronostaj 2010: The past meets the present in Swedish FrameNet++. I: *Proceedings of the 14th EURALEX International Congress*. Leeuwarden: EURALEX, 269–281.
- Boyd-Graber, Jordan, Christiane Fellbaum & Daniel Osherson, Robert Shapire 2006: Adding dense, weighted connections to WordNet. I: *Proceedings of the Third International Wordnet Conference, GWC 2006*. Brno: Masaryk University, 29–35.
- Civil, Miguel 1990: Sumerian and Akkadian lexicography. I: Franz Josef Hausmann, Oskar Reichmann, Herbert Ernst Wiegand, Ladislav Zgusta (utg.): *Wörterbücher: Ein internationales Handbuch zur Lexikographie. Zweiter Teilband / Dictionaries: An international encyclopedia of lexicography. Second volume / Dictionnaires: Encyclopédie internationale de lexicographie. Tome second*. Berlin: Walter de Gruyter, 1682–1686.
- Döderlein, Ludwig 1863: The author's preface. I: *Döderlein's handbook of Latin synonymes*. Translated by Rev. H.A. Arnold, B.A., with an introduction by S.H. Taylor, LL.D. Andover: Warren F. Draper, ix–xvi.
- Fellbaum, Christiane (utg.) 1998: *WordNet: An electronic lexical database*. Cambridge, Massachusetts: MIT Press.
- Järborg, Jerker 2001: *Roller i Semantisk databas*. Research Reports from the Department of Swedish, No. GU-ISS-01-3. University of Gothenburg: Dept. of Swedish.
- Kann, Viggo & Magnus Rosell 2006: Free construction of a free Swedish dictionary of synonyms I: *Proceedings of the 15th NO-DALIDA conference*. Joensuu: University of Eastern Finland, 105–110.

- Lieber, Francis 1841: Preface of the translator. I: *Dictionary of Latin synonymes, for the use of schools and private students, with a complete index*. By Lewis [Ludwig] Ramshorn. From the German by Francis Lieber. Boston: Charles C. Little and James Brown, iii–viii.
- Lindén, Krister & Lauri Carlson 2010: FinnWordNet – WordNet på finska via översättning. I: *LexicoNordica 17*: 119–140.
- Morris, Jane & Graeme Hirst 2004: Non-classical lexical semantic relations. I: *HLT-NAACL 2004: Workshop on Computational Lexical Semantics*. Boston, Massachusetts: ACL, 46–51.
- Murphy, M. Lynne 2003: *Semantic relations and the lexicon*. Cambridge: Cambridge University Press.
- Palmer, Martha, Hoa Trang Dang & Christiane Fellbaum 2006: Making fine-grained and coarse-grained sense distinctions, both manually and automatically. *Natural Language Engineering 13*(2): 137–163.
- Pedersen, Bolette Sandford, Sanni Nimb, Jørg Asmussen, Nicolai Hartvig Sørensen, Lars Trap-Jensen & Henrik Lorentzen 2009: DanNet: the challenge of compiling a wordnet for Danish by reusing a monolingual dictionary. I: *Language Resources & Evaluation 43*:269–299.
- Piasecki, Maciej, Stanisław Szpakowicz & Bartosz Broda 2009: *A wordnet from the ground up*. Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej.
- Svensén, Bo 2010: Liktydingar med anor. Om den latinska synonymlexikografins utlöpare i Sverige. I: K. Jóhannesson et al. (red.): *Bo 65. Festskrift till Bo Ralph*. Göteborg: Meijerbergs arkiv för svensk ordforskning, 372–382.

Internethänvisningar

DICTIONARY.COM 2011 = <http://dictionary.reference.com>
(juni 2011)

GWN 2011 = <http://www.globalwordnet.org> (juni 2011)
PWN 2011 = <http://wordnet.princeton.edu> (juni 2011)
SALDO 2011 = <http://spraakbanken.gu.se/saldo> (juni 2011)
SB 2011 = <http://spraakbanken.gu.se> (juni 2011)
SBLEX 2011 = <http://spraakbanken.gu.se/sblex> (juni 2011)
SYNLEX 2011 = <http://folkets2.nada.kth.se/synlex.html> (juni
2011)
WIKTIONARY 2011 = <http://sv.wiktionary.org> (juni 2011)

Lars Borin
professor
Språkbanken
Institutionen för svenska språket
Göteborgs universitet
Box 200
SE-405 30 Göteborg
lars.borin@svenska.gu.se

Markus Forsberg
forskare
Språkbanken
Institutionen för svenska språket
Göteborgs universitet
Box 200
SE-405 30 Göteborg
markus.forsberg@gu.se

Onomasiologiskhed og leksikografi

Ken Farø¹

What is *onomasiologicalness*? How does it interact with (other) lexicographical concepts? Is it a relevant lexicographical genre criterion? The newcoined term *onomasiologicalness* makes it easier to more clearly graduate the traditional binarity of the lexicographical genre distinction between semasiological dictionaries on the one hand and onomasiological dictionaries on the other. In this vein, the onomasiologicalness of a number of quite different dictionary types is looked at.

1. Indledning

Denne artikel beskæftiger sig med den onomasiologiske tilgang til ordforrådet i en leksikografisk kontekst. Jeg forsøger at trække nogle mere overordnede linjer, uden at være bundet af en tilknytning til et konkret ordbogsprojekt. Artiklen indeholder følgende:

- en nødtørftig begrebsafklaring af det, der her kaldes *onomasiologiskhed*
- en refleksion over forholdet til andre leksikografisk relevante kategorier
- en stillingtagen til “onomasiologisk” som leksikografisk genrebetegnelse

Ligesom bidraget ikke er rent leksikografisk, er dets empiri heller ikke udelukkende nordisk. Der indgår herudover tyske, engelske

1 Tak til Lars Brink for både diskussioner og grundige kommentarer.

og vietnamesiske ordbøger. Der er tale om en åbning af en diskussion, snarere end om et helstøbt teoretisk bidrag.²

2. Begreb og term

Vi lægger ud med et forsøg på et lille overblik over det, vi kalder det “onomasiologiske”. Selve termen er i hvert fald i dansk sammenhæng ikke synderligt udbredt. Det fremgår bl.a. af *Den Store Danske*, som alene taler om “begrebsordbøger” og “synonymordbøger” og ikke om “onomasiologiske ordbøger”. Ikke desto mindre er det en velkommen metaterm, som derfor med rette indgår i titlen på dette nummer af *LexicoNordica*, bl.a. fordi den åbner for en bredere metaleksikografisk diskussion. Men begrebet er på den anden side også så abstrakt, at det kan være vanskeligt at se nytten af det. Jeg gør et ydmygt forsøg i dette bidrag.

Et muligt synonym til termen er *betegnelselære* (jf. tysk *Bezeichnungslehre*, se Rehbock 2000), fordi det er en tilgang til ordforrådet, der går fra indholdet til de aktuelle betegnelser for det samme indhold. Der spørges altså med Dornseiff (1959): “Hvilke ord og ordforbindelser bruger vidt forskellige mennesker, typer, grupper, områder og tider, når de vil udtrykke forskellige indhold, og hvorfor?” (min oversættelse). Det interrogative element i Dornseiffs formulering er ikke uvæsentligt, for som leksikologisk disciplin er onomasiologien i sit historiske udgangspunkt erkendelsessøgende³, hvor den i leksikografisk sammenhæng er et struktureringsprincip.

Dornseiffs skudsmål vækker minder om sociolingvistikken, og det eksplicitte onomasiologiske perspektiv har da også sin oprin-

2 At der heller ikke i alle tilfælde er tale om de nyeste udgaver af ordbøgerne, spiller ingen rolle for artiklens pointer.

3 Interessen retter sig såvel mod rent sproglige problemer som betydningsforandringer og mere kulturelle fænomener.

delse i den franske og tyske dialektgeografi, hvor koblingen mellem sprog og kultur spiller en væsentlig rolle. Fra onomasiologiens ståsted er sprogtegnets *betydning* undersøgelsens udgangspunkt og konstant (Nöth 2000:148). Onomasiologi handler grundlæggende om leksikalsk selektion, nemlig: “the selection of a name for a referent” (Geeraerts/Grondelaers 2002:314f.).

Onomasiologiens antonym, semasiologien, kaldes også *betydningslære* (af tysk *Bedeutungslehre*, Bußmann 2008:618). Den tager afsæt i de sproglige udtryk, altså de leksikalske former, og undersøger primært deres semantiske egenskaber eller betydningsrelationerne imellem dem.

Man kan så stille spørgsmålet, hvilken rolle *fremmedsprog* spiller i denne begrebssystematik? For onomasiologien er i sit historiske udgangspunkt nok varietetsorienteret, men den er samtidig også noget modersmålsfokuseret. Ifølge Nöth (2000:160) er den interlingvale oversættelses semantik altid onomasiologisk. I retningen $L1 \rightarrow L2^4$ er der tale om en onomasiologisk problemstilling. Vi står jo med et indhold X i sprog Y, hvortil vi søger et udtryk i sprog Z. Jo mere sproglig diversitet, der er på ækvivalentsiden, desto mere onomasiologisk er ordbogen. Derfor er polylingvale ordbøger ofte onomasiologiske; det samme er sådanne, der angiver dialektale, gruppesproglige eller andre sociolektale ækvivalenter.

Den nordisk baserede *Förbjudna ordboken* (1998) er i høj grad polylingval (den behandler 9 sprog) og angiver ækvivalenter til en lang række tabuiserede ord fra fire kildeområder: sygdom, afføring, religion og sex – ikke tilfældigt de samme semantiske domæner, som leverer stof til bandeord.

Man kan også anskue forskellen mellem det semasiologiske og det onomasiologiske perspektiv ved hjælp af to andre leksikologiske begreber, synonymi og polysemi: Synonymi er et fokusom-

4 L1 skal her læses som ‘modersmål’, mens L2 repræsenterer fremmedsproget.

råde i onomasiologien, mens polysemi er et af semasiologiens centrale problemer (Geeraerts/Grondelaers 2002:304). For uden flere betegnelser for (næsten) samme indhold ingen onomasiologi og uden flere indhold(svianter) til samme betegnelser i praksis ingen eller kun en fattig semasiologi.

I dansk almensprog har vi verbumsparret *hedde* og *betyde*, som repræsenterer henholdsvis det onomasiologiske og det semasiologiske perspektiv.

Uanset denne skelnen kommer man ikke uden om et vist “interplay of semasiology and onomasiology” (d.s.:314), som det vil fremgå af det følgende.

3. Typer af onomasiologiske ordbøger⁵

3.1. Synonymordbøger

Synonymordbøger (fx *Politikens Synonymordbog*) nævnes ofte som eksempel på onomasiologiske ordbøger. I en typisk synonymordbog er det genuine formål (Wiegand 1999), at brugeren ved hjælp af et ham allerede kendt leksem skal kunne finde frem til et andet leksem med en beslægtet betydning, som fx passer stilistisk, pragmatisk eller semantisk bedre til den aktuelle kommunikationssituation. I nogle tilfælde vil det være et, der “ligger på tungen”, som Dornseiff allerede i 1933 formulerede det fyndigt i sit første forord til den onomasiologiske klassiker *Deutscher Wortschatz nach Sachgruppen* (Dornseiff 1959).

Sidsteleddet *-schatz* fører os videre til *tesaurusser*, der netop rummer etymonet “skatkammer”. Som term er *tesaurus* lidt vaklende i sin brug. I nogle sprog, som fx engelsk, er begrebet nærmest

5 Diskussionen afstår fra at inddrage den ellers oplagte genre *begrebsordbog*, fordi den får en meget udførlig omtale i hhv. Nimb (2011) og Trap-Jensen (2011).

synonymt med *synonymordbog*. *Collins Cobuild Dictionary and Thesaurus* indeholder således i sin fod en løbende undertekst med synonymer til ordene fra hovedteksten. *Nonevent* suppleres således med *flop* (trykt med fed), *disappointment*, *failure*, *fiasco*, *dud*, *washout*. Hertil kommer enkelte genus- og stiloplysninger.

Selv om synonymordbøger har status som et af flere paradeeksempler på onomasiologiske ordbøger, så er de ikke nødvendigvis konsekvent onomasiologiske. Dels er de ikke enestående: Synonymifunktioner eller snarere angivelse af betydningsrelationer i mere bred forstand er et element i mange såkaldte betydningsordbøger; fx indeholder DDO ud over synonymer også hyponymer (fx *cykel* > *tandem*), hyperonymer og antonymer, samt mere uspecifikt betydningsbeslægtede ord (*fremmedsprog* > *andetsprog*), dels er synonymordbøgerne typisk kun onomasiologiske i deres mikrostruktur. Makrostrukturelt er de helt traditionelt semasiologiske: Lemmapræsentationen er som regel rent alfabetisk ordnet. I nogle tilfælde er den ganske vist suppleret med et stikordsregister, men det er i reglen også alfabetisk.

3.2. Fraseologiske ordbøger

Er fraseologiske ordbøger onomasiologiske? Spørgsmålet er relevant, fordi fraseologien af visse forskere udlægges som et tillæg til enkeltordene, der sikrer afsender et kommunikativt variationspotentiale. Det er ikke vanskeligt at finde eksempler på, at fraser supplerer ordene stilistisk, konnotativt og ikonografisk. Fx er de fraseologiske muligheder for at udtrykke indholdet '(om mand) få sex' næsten uendelige på dansk. Her er et lille udvalg:

- (1) få noget
- (2) få noget på den dumme
- (3) score
- (4) fedte rotten

En anden onomasiologisk topscorer er indholdet ‘dø’, der manifesterer sig i et utal af fraser på dansk. Nogle få eksempler:

- (5) tage billetten
- (6) stille (træ)skoene
- (7) få en etværelses med låg
- (8) gå til de evige jagtmarker
- (9) gå hjem til Gud
- (10) gå bort
- (11) sove (stille) ind
- (12) tabe kampen mod kræften/...

Frasernes funktion er (i det aktuelle eksempel) eufemistisk at omskrive det unævnelige, i (5) til (8) ved at spøge med det, eller i (9) til (12) ved hjælp af det modsatte: sproglig diskretion. Selv om fraseologien således er højrelevant for onomasiologien, gør det ikke per automatik fraseologiske ordbøger onomasiologiske. De kan placere sig vidt forskelligt på det *onomasiologisk-semasiologiske kontinuum*. Dette begreb indfanger det forhold, at både leksikografiske enkeltelementer og hele ordbøger kan placeres et sted på en skala, hvor “fuldt semasiologisk” udgør det ene ekstrem, og “fuldt onomasiologisk” det andet.

Netop det er tilfældet for følgende fire eksempler: (a) *Danske Talemåder*, (b) *Idiomordbog*, (c) *Talemåder i dansk* og (d) *Synonymwörterbuch der deutschen Redensarten*.

(a) *Danske Talemåder* bevæger sig overhovedet ikke over i den onomasiologiske ende af kontinuummet. Dens særkende er nemlig alene dokumentationen og forklaringen af de (farverige) idiomer, kompilatoren har fundet under sin daglige læsning. Der er ikke meget leksikografisk raffinement eller brugerfunktionalitet at hente i ordbogen. Dens væsentligste værdi består i de mange autentiske konteksteksempler, der demonstrerer frasersnes opførsel i sprogbrugen.

(b) Trods sit visuelle element (se afsnit 3.3.) – ordbogen indeholder stregtegninger, der illustrerer idiomernes ikonografi – er heller ikke *Idiomordbog* fra Gyldendal onomasiologisk i sit koncept. Idiomets betydning forklares, og tegningerne har ikke andet end en lidt primitiv underholdningsværdi: Det er meget forudsigeligt, hvordan en tegner håndterer et idiom som *have hovedet under armen* eller *stå med skægget i postkassen*. Og det er trods alt den side af idiomet, der er lettest at forstå og beskrive; den har ikke nogen onomasiologiske implikationer. For det er *udtryksside-semantik* (jf. Farø 2006), eller med en term af Brink (2006): *pålydendebetydning*, hvorved vi ikke bevæger os afgørende væk fra det semasiologiske perspektiv.

(c) Det er først med *Talemåder i dansk*, at det fra et onomasiologisk standpunkt bliver interessant. I første omgang kan man fokusere på det faktum, at bogen kategoriserer sine fraser ud fra 30 forskellige tematiske grupper. Det er ikke, som man skulle tro, på indholds niveau, men derimod på ikonografisk plan. Men det er faktisk ikke noget onomasiologisk greb, for det orienterer sig igen mod idiomets udtryksside i bred forstand. Under emnet *Sygdom, sundhed og død* figurerer fx *der ligger hunden begravet*, skønt idiomets betydning snarere har noget med årsager og motiver at gøre. Men ordbogen indeholder et andet element, som med rette kan kaldes onomasiologisk, og det er synonymangivelser. Ganske vist forekommer der både synonymer i definitionsdelen og i en særlig frasedel. Fx defineres *det regner i stride strømme* både ved hjælp af indholdsparfrasen “det regner meget stærkt” og synonymet “det øsregner” i samme del af artikelstrukturen. Men generelt gives de (fraseologiske) synonymer på en særskilt plads nederst i artiklen, efter evt. fremmedsprogede ækvivalenter – her mærkeligt nok kun det tyske eksempel *es gießt/regnet in Strömen*. I det aktuelle tilfælde gives der synonymer som “(det er) et herrens vejr; det regner skomagerdrenge; det øser i spande” m.fl. Der er blevet gjort meget lidt brug af henvisninger.

(d) En virkelig onomasiologisk idiomordbog er *Synonymwörterbuch der deutschen Redensarten*. Den indeholder 20.000 fraser, som er inddelt systematisk efter et begrebssystem, der begynder med *Tid, rum, sansedata* (fx *bis spät in die Nacht*) over *Holdning til medmennesker (jm den Kopf verdrehen)* til *Præferencer* (fx *weiß der Kuckuck!*, et udtryk for ligegyldighed eller opgiven). Værket indeholder desuden et stikordsregister, som gør det muligt at søge målrettet efter idiomer m.m., der udtrykker bestemte indhold. Fx kan man søge udtryk for *fræk, forargelse, vrede og begravelse*. Dette element er ikke alene nyttigt i forbindelse med brugerens produktion på tysk, men også for en leksikograf, der indædt søger efter ikonografiske frase-ækvivalenter til sin dansk-tyske ordbog. Forfatteren gør i sin indledning opmærksom på, at ordbogen tillader hele tre forskellige brugsmåder: 1. en semasiologisk, 2. en onomasiologisk og 3. en associativ. Hvis alle nordiske sprog havde et lignende værktøj, ville den nordiske idiomleksikografi ikke alene være et langt skridt videre. Det ville også være et solidt grundlag for en internordisk fraseleksikografi.⁶

Fraseologiske ordbøger er altså ikke nødvendigvis onomasiologiske. Men de lægger i høj grad op til en onomasiologisk behandling, i hvert fald i modersmålsammenhæng, hvor fraseologien kommunikativt først og fremmest repræsenterer en mulighed for udtryksvariation⁷. Hermed ser vi bort fra mere kognitive, meta-lingvistiske og underholdningsmæssige interesser for idiomatik, som ordbøger også kan vælge at honorere.

6 Tættest kommer vel Jón Hilmar Jónssons *Stóra orðabókin um íslenska málnotkun* (2005), som indeholder en liste med danske ækvivalenter til de islandske begrebslemmaer, hvilket giver danske og danskkyndige brugere en bedre tilgang til ordbogsteksten.

7 Men altså ikke udelukkende, for selvfølgelig repræsenterer idiomer også i nogen grad og ikke mindst for nogle modersmålstalere et receptionsproblem.

3.3. Visuelle ordbøger

Er *visuelle* ordbøger onomasiologiske? Der ser i hvert fald ud til at være en vis sammenhæng mellem visualisering og onomasiologiskhed, som det vil fremgå af dette afsnit.

Man kan naturligvis ophøje alle ordbogens kendetegn til genrekriterium. Det gælder også for et træk som en illustration. Nu kan illustrationer være et mere eller mindre bærende element ved en ordbog. I en del tilfælde har illustrationen bare en understøttende funktion i forhold til receptionen af de pågældende leksemer. Det gælder eksempelvis i nogle fremmedordbøger. Decideret visuelle ordbøger, dvs. ordbøger hvor illustreringen udgør et bærende element i ordbogens funktionalitet, kendes især fra lærerleksikografien, både den der understøtter modersmålstilegnelsen, og den der potentielt bidrager til fremmedsprogstilegnelsen. Børnebogs-markedet bugner af gennemillustrerede ordbøger. Man føler her instinktivt en indskydelse til at sætte "ordbøger" i citationstegn. Men "ordbog" er ingen beskyttet titel og ingen fagterm⁸. De fuldt illustrerede børneordbøger findes både i overvejende onomasiologiske på den ene side og semasiologiske varianter på den anden, som overfladisk kan ligne hinanden, men som strukturelt er helt forskellige.

ABC-Bog er fx rent semasiologisk, eftersom den alene er struktureret ud fra det alfabetiske princip. De tableauer, barnet får serveret på siderne, er usandsynlige pseudotableauer uden nogen ontologisk sammenhæng mellem de illustrerede entiteter. De knyttes alene sammen af det fælles begyndelsesbogstav. På et tableau sidder fx en lænket løve på en lastbil med en lineal mellem benene og en lampe i ryggen, mens det lyner, og en loppe springer fra dens manke. Formålet er her at træne den lydige bevidsthed, ud over selvfølgelig at lære ord og stavning.

8 En anden udlægning er at betragte termen som repræsenterende både et almen- og et fagsprogligt indhold.

Andre børneordbøger er mere onomasiologiske. Merchandise-udgivelsen fra Danmarks Radio *Bamse fuld af ord* består af dobbeltsider med hvert sit tema og tableau, som honorerer et børneunivers, fx en skovtur, udklædning m.m. Ud fra disse indholdsuniverser kan barnet lære ord i sammenhæng.



Figur 1: Tilnærmelsesvis onomasiologisk tableau fra børneordbogen *Bamse fuld af ord*.

Lad os nu bevæge os fra modersmålet og til fremmedsproget. Det rudimentære tableau-koncept i Bamse-ordbogen går igen på højere niveau i den bilingvale *The New Oxford Picture Dictionary English/Vietnamese* (TNOPDEV). Her er 2.400 engelske leksemer, som optræder i 82 såkaldte “contextualized illustrations”, blevet oversat til vietnamesisk. Bogen er progressivt indrettet, således at den begynder med familierelationer og kroppen, mens sportsgrene og fritidsaktiviteter udgør afslutningen. På den måde får den snarere en lærebogs struktur, end den er et begrebssystem. Desuden bliver asymmetrien mellem udgangs- og målsprog meget tydelig i så-

danne tilfælde, hvor de kulturelle forskelle er særligt store. Relevansen af de genstande og aktiviteter m.m., tableauerne afbilder, er meget forskellig i Vietnam og den vestlige engelsktalende verden, hvis en sådan findes.

Det er relativt tydeligt at spore, hvilke ækvivalenter til de engelske ord der ikke dækker over hjemlige koncepter, men som må antages at være ad hoc-oversættelser. Hvis vi ser bort fra egentlige, synlige importord som *ga* (banegård, fra fransk *gare*) og *phim* (film), så er de indimellem meget lange morfem-syntagmer et vidnesbyrd om, at der her er et fænomen, der er (traditionel) vietnamesisk kultur fremmed. Det er vanskeligt at lave velfungerende onomasiologiske ordbøger på tværs af omfattende kulturskel.

Foreign learner-ordbogen *MacMillan English Dictionary for Advanced Learners* (MEDAL) rummer også forskellige visuelle elementer, om end mindre konceptuelt end i den netop omtalte engelsk-vietnamesiske. Dels finder man under opslag som *bed* illustrationer af eksemplarer af den genstand, lemmaet denoterer. Et andet eksempel er opslaget *shape*, hvor der ofres en side på diverse geometriske standardformer og deres engelske betegnelser. Der behøver dog ikke at være tale om et eksplicit overbegreb – eller om et overbegreb overhovedet: Lemmaet *sharp* giver fx anledning til en sammenstilling af en række betydningsbeslægtede ord med illustrationer, som må formodes at kunne volde problemer for lærerens produktion: *pointed* (stiletter), *jagged* (glasskår), *spiky* (frisur), *sharp* (kniv), *serrated* (brødkniv), *blunt* (kniv) og *prickly* (blad). Ordbogens håndfuld tableauer (*House, Kitchen, Office, Clothes and patterns, Motorway and types of Transport* og *Car*) virker i sammenligning lidt tilfældige.

Politikens visuelle ordbog (PVO, 2007) er samlingens mest konsekvent onomasiologiske visuelle leksikografiske produkt. Den består af 920 sider med billeder og tableauer med udgangspunkt i et dansk ord, der tilsyneladende oversættes til 4 sprog: engelsk, tysk, fransk og spansk. Hertil kommer et overordentligt omfattende

register på næsten 200 sider. I modsætning til den netop analyserede ordbog med engelsk og vietnamesisk er værket struktureret ontologisk-hierarkisk. På makroniveau er inddelingen fagområder eller grupper af genstande, fx *Kunst og arkitektur* eller *Gør det selv og havearbejde*. Makroniveauet er ofte yderligere underinddelt; fx er *Billedkunst* inddelt i bl.a. *Maling og tegning*, der herudover omfatter fx *Vigtigste materialer* og *Udstyr*. På det underliggende mellemniveau kommer så selve de afbildede genstande (fx *synthesizer*) eller i nogle tilfælde tableauer som *badminton(bane)*. På (det optionelle) mikroniveau følger endelig genstandenes eller tableauernes enkeltdele.

Det onomasiologiske perspektiv åbner for en sondring, som ofte overses, nemlig distinktionen mellem faglig/ikke-faglig. Skellet mellem disse to funktionelle kontekster fremstår ret tydeligt i PVO. Dels er det generelt sådan, at mikroniveauet er der, hvor det faglige vokabular udfoldes. En *mejetærsker (Transport og maskiner / Tunge maskiner)* er jo ikke i sig selv fagsprog, men så snart man får behov for en reference til genstandenes enkeltdele, så aktiveres fagsproget i form af termer som *indføringskanal / feeding tube / Schrägförderer* og *tømmesnegl / unloading tube / Auslaufrohr*.

Visualisering er altså både potentielt og tendentielt et element, der spiller sammen med et onomasiologisk greb om ordforrådet. I forhold til de fraseologiske ordbøger er de to undersøgte hovedrepræsentanter for de visuelle ordbøger, TNODEV og PVO, i meget høj grad encyklopædiske – de forsøger jo både med deres billedudvalg og begrebssystematik at gengive verden sprogligt, mens den fraseologiske type så godt som ikke besidder encyklopædiske træk overhovedet: Her handler det udelukkende om sproglig ageren. Samtidig er visualisering meget lidt oplagt i sidstnævnte. Det er der gode grunde til. I det omfang, man opererer med en relativt snæver idiomdefinition, og det gør både *Synonymwörterbuch der deutschen Redensarten*, *Talemåder i dansk*, *Danske talemåder* og *Idiomordbog* i hvert fald i praksis, er

kategoriens repræsentanter netop karakteriseret ved at være “non-encyklopædiske” (Farø 2006), fordi de ikke refererer til verden udenfor, men til menneskets subjektive oplevelser og relationer. De er desuden vanskelige at illustrere, fordi deres indhold i reglen er abstrakt (jf. forordet til *Talemåder i dansk*).

4. Den onomasiologiske ordbogs funktion(er)

Ifølge standardopfattelsen er den modersmåls-onomasiologiske ordbogs funktion at *aktivere* allerede kendt, men glemt viden hos brugeren. Det er også Dornseiffs udgangspunkt i hans originale forord til *Deutscher Wortschatz nach Sachgruppen* fra 1933 (1959). Men dels er opfattelsen ikke helt inkompatibel med den leksikografiske virkelighed, dels er målsætningen bag den ikke nødvendigvis hensigtsmæssig.

Selv om en del såkaldte onomasiologiske ordbøger er ganske nøgne i deres angivelsesstruktur, så er der selv i de mest karige af dem sprækker i nøgenheden. Under *parre sig* i *Politikens Synonymordbog* får man således immervæk oplysninger om henholdsvis *knalde* og *kendes* stilleje eller kontekst, der er henholdsvis lavsproglig og bibelsk. Men andre ordbøger går væsentligt videre. I den tidligere omtalte *Sinn- und sachverwandte Wörter* stilmarkeres der temmelig konsekvent og udførligt. Der gives også semantiske og pragmatiske distributionsoplysninger som fx “mht. mad”⁹ om *jmdm. die Bissen in dem Mund zählen* som synonym til *geizig* (‘nærig’).

Så synspunktet, at en onomasiologisk ordbog nødvendigvis skal aktivere allerede erhvervet viden, er relativt bart og måske endda problematisk. Det forudsætter en homogen og kompetent brugerprofil, som kun få brugere i virkeligheden kan leve op til. I mange tilfælde rækker oplysningerne i onomasiologiske ordbøger

9 “in bezug auf Essen”.

kun til krydsordsløsning, som med god vilje kan kaldes for semi-produktion. Reel sprogpåvirkning kræver væsentligt mere nuancerede produktionsoplysninger (jf. Hannesdóttir & Tingsell 2011).

Man behøver ikke engang at pege på de ikke-modersmålstalende som en i øvrigt meget oplagt brugergruppe, der ikke har så megen viden at aktivere. Også børn, unge, mindre belæste og i nogen grad også veluddannede vil indimellem opleve, at visse synonym-ækvivalenter repræsenterer ny viden og ikke bare aktivering af gammel.

Der kan naturligvis være tekniske og ikke mindst ressourcemæssige grunde til at praktisere princippet om det nøgne ækvivalentleksem i en onomasiologisk ordbog. Og der er tradition for at gøre det på den måde, selv om den altså ofte ikke er helt konsekvent gennemført. Men det helt grundlæggende spørgsmål er, hvorfor man skal ophæve noget til princip eller ligefrem dyd, som i mange tilfælde slet ikke er i brugerens interesse? For den er at få svar på alle potentielle spørgsmål, og dem er der mange af, når man anvender de foreliggende onomasiologiske bøger. Herudover er der det elektroniske argument: En elektronisk onomasiologisk ordbog er ikke tynget af de ekstra data, en sådan funktionel komplettering af "aktiveringsdelen" kræver; her kan der *i princippet* bare hældes på, som Aarhuskolen ved flere lejligheder har fremført. Men selv her anses det vist for optimalt ikke at belaste brugeren med (for meget) ikke-efterspurgt information, fordi det jo kan hindre, at han overhovedet finder det, han reelt er interesseret i. Så alting med måde.

I nogle tilfælde kan man pege på helt andre funktioner. Brødrene Melin (1998) nævner, at deres allerede omtalte *Förbjudna ordboken* som en af sine hovedfunktioner har at sætte brugeren i stand til ikke at producere, men for en gangs skyld at *undgå* de pågældende former, fx i international erhvervskommunikation: "Fula ord är nämligen ofta korta och har en enkel ljudstruktur – precis de egenskaper man önskar sig för namn på företag och

produkter i international markedsføring. Och där är orden definitiva omöjliga” (fra bagteksten). Den store danske energiproducent *DONG Energy* kunne med fordel have rådført sig med denne ordbog og ville så måske ikke have fastholdt sit historiske danske førsteled (en forkortelse for *Dansk Olie og NaturGas*), der i den internationale kommunikation skriftligt giver associationer til kønslivet (*dong* er engelsk slang for ‘penis’), mens en lydlig afkodning suppleres med indholdet ‘afføring’ (*dung*).

Onomasiologiske ordbøger er langt overvejende orienteret mod de *kommunikative funktioner* (1) tekstproduktion og (2) oversættelse til L2, mod (3) indlæring samt mod den *kognitive funktion* (4) viden eller metalingvistisk information (jf. Tarp 2007). Der er ingen principielle grænser for, hvor mange informationer ordbogen kan give for at understøtte disse funktioner. Giver der kun få eller næsten ingen, er det et udtryk for tradition eller ressourceknaphed, ikke iboende genrekrav.

5. Sammenfatning

Onomasiologiskhed er et perspektiv i kommunikationssituationen, der bevæger sig fra tegnets indhold til dets udtryk. Samtidig er det et leksikografisk struktureringsprincip. Det onomasiologiske perspektiv har ændret sig fra primært at tjene erkendelsesinteresser som betydningsforandring til i højere grad at understøtte praktiske kommunikative behov. Nogle ville sige fra leksikologi til leksikografi, andre fra én form for leksikografi til en anden.

Begrebet har affinitet til både den fraseologiske, encyklopædiske, dokumentariske, pædagogiske og den produktionsleksikografiske genre. I sig selv er *onomasiologisk* ikke nogen funktionel genrebetegnelse, selv om det er praktisk i værkstedssproget at kunne skelne groft mellem de to grundlæggende strukturerings-

principper *semasiologisk* og *onomasiologisk*. Men det er derudover principielt altid muligt at placere en ordbog på det onomasiologisk-semasiologiske kontinuum, både på elementplan og når den enkelte ordbog anskues ud fra sin samlede grad af onomasiologiskhed. Dermed er også begrundelsen for den terminologiske neologisme *onomasiologiskhed* givet: I leksikografisk sammenhæng er der altid tale om en grad af samme.

Ud over at begrebet er relativt og graduerbart, og ikke binært, er det relevant for forskellige leksikografiske niveauer og funktioner. En ordbog kan altså indeholde forskellige onomasiologiske elementer snarere end blot og bart at være onomasiologisk eller semasiologisk. To ordbøger, der set med det binære blik formentlig ville blive betegnet som semasiologiske, er *Macmillan English Dictionary for Advanced Learners* og *Den Danske Ordbog*, der dog begge indeholder væsentlige onomasiologiske karakteristika. Førstnævnte er imidlertid som accentueret produktionsværktøj noget mere sofistikeret og pragmatisk end DDO, som skal leve op til sin metafunktion som videnskabelig nationalordbog, der især fokuserer på betydningsforklaring. MEDAL gør således brug af typiske onomasiologiske virkemidler som billeder, synonymik og sprogbrugsoplysninger præcis der, hvor de skønnes at gøre gavn, og i et omfang, der ligeledes er funktionelt betinget.

De leksikografiske elementer, man tendentielt kan kalde onomasiologiske, eller som ofte indgår i samspil med det onomasiologiske, er:

1. betydningssystematisk ordning af lemmata
2. produktionsorientering
3. visualisering (i form af tableauer eller plancher)
4. synonymik (samt antonymik og hyperonymik og hyponymik)
5. begrebslig universalisme
6. fokus på det encyklopædiske frem for det snævert sproglige

Sidstnævnte finder man bl.a. i *Wörter und Sachen*-orienterede løbende dialektstovværker som de danske *Ømålsordbogen* og *Jysk Ordbog*.

Det onomasiologiske perspektiv er i sin grundide altså universalistisk, ligesom ideen om interlingval ækvivalens i fx bilingvale ordbøger grundlæggende er det. Men analysen af ordbøgerne i dette bidrag demonstrerer, at en sådan vision i praksis er umulig, både på grund af kulturelle forskelle og de samfundsændringer, tiden er ansvarlig for. Det universelle begrebssystem uafhængigt af tid og rum findes ikke (jf. også Trap-Jensen 2011).

Litteratur

Ordbøger og andre opslagsværker

ABC-Bog. København: Giraf Produkter 2008.

Bamse fuld af ord. Red. Gitte Meldgaard. København: Danmarks Radios Forlag 2009.

Bußmann, Hadumod 2008: *Lexikon der Sprachwissenschaft*. Stuttgart: Kröner.

Collins English Dictionary and Thesaurus. Red. J.M. Sinclair. Harper Collins, 2. udg. 2000.

Danske talemåder. Red. Allan Røder. København: Gads Forlag 1998.

Den Danske Ordbog 1-6. Hovedred. Ebba Hjorth & Kjeld Kristensen. København: Gyldendal/DSL 2003-05.

Den Store Danske. *Gyldendals åbne encyklopædi*. <http://www.denstoredanske.dk>

Dornseiff, Franz (red.): *Der deutsche Wortschatz nach Sachgruppen*. Berlin/New York: Walter de Gruyter 1959.

Förbjudna ordboken. Red. Lars og Martin Melin. Stockholm: Norstedts 1998.

- Gyldendals Fremmedordbog*. Red. Sven Brüel og Niels Åge Nielsen. København: Gyldendal 1987.
- Idiomordbog. 1000 talemåder*. Red. Christian Michelsen. København: Gyldendal 1993.
- Jónsson, Jón Hilmar 2005: *Stóra orðabókin um íslenska málnotkun*. Reykjavík: Forlagið.
- Jysk Ordbog*. <http://www.jyskordbog.dk/>
- Macmillan English Dictionary for Advanced Learners*. Red. Michael Rundell m.fl. Oxford: Macmillan Education.
- New Oxford Picture Dictionary English-Vietnamese*. Red. E.C. Parnwell. Oxford: OUP 1989.
- Nöth, Winifried 2000: *Handbuch der Semiotik*. Stuttgart: Metzler.
- Politikens synonymordbog*. Red. Allan Karker. København: Politiken 1988.
- Politikens visuelle ordbog. Dansk – engelsk – tysk – fransk – spansk*. Red. Jean-Claude Corbeil & Ariane Archambault. København: Politiken 2007.
- Sinn- und sachverwandte Wörter. Wörterbuch für den treffenden Ausdruck*. Red. Dudenredaktion. Mannheim: Duden 1986.
- Synonymwörterbuch der deutschen Redensarten*. Red. Hans Schemann. Stuttgart: Klett 1993.
- Talemåder i dansk*. Red. Stig Toftgaard Andersen. København: Gyldendal 2001.
- Variantenwörterbuch des Deutschen*. Red. Ulrich Ammon. Berlin/New York: de Gruyter 2005.
- Ømålsordbogen. En sproglig-saglig ordbog over dialekterne på Sjælland, Lolland-Falster, Fyn og omliggende øer*. København: Afdeling for Dialektforskning og C.A. Reitzels Forlag 1992-.

Anden litteratur

- Brink, Lars 2006: Den fraseologiske terminologi. I: *Studier i Nordisk* 8, 39-51.

- Cruse, D. Alan, Franz Hundsnurscher, Michael Job & Rolf Lutzeier 2002 (udg.): *Lexikologie. Ein internationales Handbuch zur Natur und Struktur von Wörtern und Wortschätzen*. Bd. 1. Berlin & New York: Walter de Gruyter.
- Cubber, Walter De 2002: Onomasiologiske Fallstudien. I: Cruse m.fl.: 752-763.
- Farø, Ken 2006: *Idiomatizität – Ikonizität – Arbitrarität. Beitrag zu einer Theorie der funktionalistischen Idiomäquivalenz*. (Upubliceret ph.d.-afhandling). København: Københavns Universitet.
- Geeraerts, Dirk & Grondelaers, Stefan 2002: Structuring of word meaning: An overview. I: Cruse et al.: 304-318.
- Hannesdóttir, Anna Helga & Sofia Tingsell 2011: Svensk fraseologi för islänningar – om den isländsk-svenska ordboken i ISLEX och en onomasiologisk svensk resurs: Frاسبanken. I: *LexicoNordica* 18 (i dette bind).
- Hausmann, Franz-Josef 1990: Das onomasiologische Wörterbuch: Ein Überblick. I: Franz-Joseph Hausmann, Oskar Reichman, Herbert Ernst Wiegand, Ladislav Zgusta (udg.): *Wörterbücher*, bd. 2. Berlin/New York: de Gruyter, 1057-1067.
- Nimb, Sanni 2011: Semantiske relationer i en ny dansk begrebsordbog: genbrug på tværs af ordbøger. I: *LexicoNordica* 18 (i dette bind).
- Rehbock, Helmut 2000: Onomasiologie. I: Helmut Glück (Red.): *Metzler Lexikon Sprache*, 492f. Stuttgart: Metzler.
- Schmidt-Wiegand, Ruth 2002: Die onomasiologische Sichtweise auf den Wortschatz. I: Cruse m.fl.: 738-751.
- Tarp, Sven 2005: The concept of a bilingual dictionary. I: *Schreiben, Verstehen, Übersetzen und Lernen : Zu ein- und zweisprachigen Wörterbüchern mit Deutsch*. Red. Irmhild Barz; Henning Bergenholtz; Jarmo Korhonen. Frankfurt a.M.: Peter Lang, 27-41. (Finnische Beiträge zur Germanistik).

Tarp, Sven 2007: *Leksikografien i grænselandet mellem viden og ikke-viden. Generel leksikografisk teori med særlig henblik på lærerleksikografi*. Doktorafhandling. Aarhus: Center for Leksikografi.

Trap-Jensen, Lars 2011: Dansk begrebsordbog: baggrund og metode. I: *LexicoNordica* 18 (i dette bind).

Wiegand, Herbert Ernst 1999: *Wörterbücher*. The Hague: Mouton.

Ken Farø
adjunkt, ph.d.
Københavns Universitet
Institut for Engelsk, Germansk og Romansk
Njalsgade 128
DK-2300 København S
kenfaroe@hum.ku.dk

Svensk fraseologi för islänningar – om den isländsk-svenska ordboken i ISLEX och en onomasiologisk svensk resurs: Frasbanken

Anna Helga Hannesdóttir & Sofia Tingsell

The ISLEX projects aim at developing an on-line dictionary site with Icelandic as the source language and the three Scandinavian languages as the target languages. This article presents the ISLEX projects, focusing on the phraseological aspects of the bilingual Icelandic-Swedish dictionary. It also describes the work on Frasbanken.se, an onomasiological electronic resource for presenting phrasemes in Swedish. Frasbanken.se is being developed to give additional phraseological information to users of the Icelandic-Swedish version of ISLEX, and is also meant to be used as a self-contained phraseological component for L2 and L1 speakers of Swedish.

In developing Frasbanken.se, our main concern has been to meet the lexicographical and pedagogical needs of the L2 speaker with respect to phraseology. L2 speakers are known to acquire phraseological information at a late stage in their linguistic progress. Lack of phraseological competence can lead to complications for the learner; he or she needs to rely heavily on grammatical rules, failing to use prefabricated chunks, and the L2 speaker appears less idiomatic in his or her language production. Frasbanken.se is thus a well-directed resource for competent L1 speakers of Icelandic with Swedish as their L2.

1. Inledning

Hösten 2011 kommer de ordböcker som nu arbetas fram inom ramen för ISLEX att bli allmänt tillgängliga på Internet. ISLEX är ett samarbete mellan fyra ordboksprojekt: ett isländskt källspråks-

projekt och tre separata målspråksprojekt, ett danskt, ett norskt och ett svenskt. Den isländska redaktionen tillhandahåller ett material som av målspråksredaktionerna analyseras ur ett kontrastivt perspektiv och förses med adekvata motsvarigheter på respektive målspråk. Arbetet utförs vid Stofnun Árna Magnússonar í íslenskum fræðum vid Háskóli Íslands, Det Danske Sprog- og Litteraturselskab i Köpenhamn, Institutt for lingvistiske, litterære og estetiske studier vid Universitetet i Bergen och Institutionen för svenska språket vid Göteborgs universitet. Nyss har också Fróðskaparsetur Føroya anslutit sig till arbetet i syfte att åstadkomma en isländsk-färöisk ordbok.

Arbetet med ordböckerna utförs i en gemensam databas som skapats inom ramen för det isländska källspråksprojektet (se Jónsdóttir & Úlfarsdóttir 2008). Den isländska institutionen ansvarar för redigeringsystemet och databasen. Varje enskilt projekt har separat, nationell finansiering och uppdraget för de olika projekten skiljer sig åt. De färdiga ordböckerna blir därför av naturliga skäl inte helt identiska. En första version av den isländsk-danska, isländsk-svenska och de isländsk-norska ordböckerna är från hösten 2011 tillgängliga bland annat på islex.se.

Den isländska källspråksbasen bygger på en strikt särspråklig analys av isländskan, en analys som är avsedd att vara neutral i förhållande till vart och ett av de aktuella målspråken. Därigenom kan den tjäna som utgångspunkt även för andra framtida ordboksprojekt, oavsett målspråk.

I anslutning till det isländsk-svenska ordboksarbetet pågår också ett projekt som syftar till att upprätta en elektronisk, onomasiologiskt ordnad samling över svenska *frasem*, dvs. fraseologiska enheter av olika slag (se Hannesdóttir, Jónsson & Tingsell 2010). Denna samling ska göras tillgänglig direkt från den isländsk-svenska ordboken.

Här ges först en kort beskrivning av ISLEX i allmänhet och särskilt den isländsk-svenska ordboken. Därefter följer en diskussion kring fraseologi ur ett tvåspråkigt perspektiv och arbetet med fra-

seologin i den isländsk-svenska ordboken i ISLEX. Med utgångspunkt i den relevans fraseologi har i inlärningen av ett andraspråk beskrivs sedan uppbyggnaden av Frasbanken.

2. ISLEX: syfte och innehåll

Ordböckerna i ISLEX vänder sig till flera olika målgrupper. Till skillnad från ordböcker mellan marknadsmässigt stora språk är det inte realistiskt att renodla de här aktuella ordböckerna till *en* målgrupp, dvs. enbart källspråkstalare eller målspråkstalare. Det är heller inte aktuellt att begränsa ordböckernas funktion till enbart produktion eller avkodning. Tvärtom försöker ordboken i själva verket vara så bidirektionell och multifunktionell som möjligt. Som en isländsk → dansk/norsk/svensk ordbok ska den tjäna isländska användare, i första hand som redskap vid produktion av olika slag på respektive skandinaviskt språk. Genom de möjligheter till flexibla sökningar som det elektroniska formatet erbjuder, kommer denna användargrupp rimligen att använda ordboken också som redskap vid avkodning av texter på de skandinaviska språken, även om den inte i första hand är avsedd att fylla den funktionen. I förhållande till de danska, norska och svenska användarna är däremot avkodningsfunktionen primär; ordboken är avsedd att användas för översättning i vidaste bemärkelse från isländska och vid studier i isländska. Det är dock sannolikt att även skandinaverna kommer att utnyttja ordboken på sätt som den inte primärt är avsedd för, nämligen som verktyg för produktion av texter på isländska. Trots att ambitionerna med ordböckerna kan förefalla utopiska (se Tarp 1995) begränsas deras faktiska funktioner av att det bara är isländskan som beskrivs utifrån ett källspråksperspektiv.

Det isländska ordförrådet i ISLEX-databasen utgörs av ca 50 000 lemman. I första hand är det det moderna isländska ord-

förrådet som beskrivs. Från källspråksredaktionen läggs också särskild vikt vid att presentera ett rikt urval av frasem av olika slag. Frasemen redovisas dels i makrostrukturen som artikelgångar, dels i mikrostrukturen under någon eller några av de enskilda komponenterna.

2.1. Den isländsk-svenska ordboken i ISLEX

Här är det av naturliga skäl den isländsk-svenska ordboken som står i fokus. Ordbokens multipla syfte beträffande användargrupper och de olika funktioner ordboken är avsedd att kunna fylla ställer den svenska målspråksredaktionen inför en rad intressanta problem. Redaktörerna försöker ta största möjliga hänsyn till den isländske användarens behov av stöd för produktion på svenska. I det avseendet är förebilderna inom den svenska lexikografen inte särskilt många. De tvåspråkiga ordböcker som utarbetas med svenska som ett av språken är i princip avsedda för svenska användare, dvs. användare som har svenska som modersmål (L1). Främst gäller detta ordböcker där svenskan är målspråk. Ett undantag utgör Norstedts nederländsk-svenska pappersordbok som enligt baksidestexten är ”avsedd att vara lika användbar för svenskspråkiga som för nederländkspråkiga användare” (Norstedts 2008); liknande formulering finns också på baksidan av Norstedts rysk-svenska ordbok.

Den lexikografiska beskrivningen av det språk som i en tvåspråkig ordbok beskrivs ur ett L1-perspektiv kan utelämna sådana drag som förutsätts ingå i användarnas modersmålskompetens. Dit hör exempelvis de lexikaliska enheternas grammatiska och syntagmatiska egenskaper. Också den fraseologi som presenteras antas vara känd för användarna. Inom svensk lexikografi har det mest renodlade som gjorts för andraspråkstalare (L2) av svenska tillkommit inom ramen för Lexin-projektet (LEXIN 2011). I de tvåspråkiga Lexin-ordböckerna utgör emellertid svenskan en-

bart källspråk. Beskrivningen av svenskan som målspråk för källspråkstalare (L1→L2) är således inte speciellt välutvecklad.

2.2. Fraser och konstruktioner i ISLEX

Den isländska fraseologi som finns i ISLEX har i stor utsträckning hämtats ur Jón Hilmar Jónssons fraseologiska arbeten: *Orðastaður* (Jónsson 1994) där perspektivet är syntagmatiskt, *Orðaheimur* (Jónsson 2002) där det är onomasiologiskt och *Stóra orðabókin um íslenska málnotkun* (Jónsson 2005) där de båda perspektiven kombineras.

I samband med att den isländsk-svenska ordboken planerades var det ur svenskt perspektiv naturligt att försöka finna sätt att dra nytta av det isländska arbetet med fraseologin dels i ett svenskt, dels i ett kontrastivt isländskt-svenskt perspektiv (se Hannesdóttir & Jónsson 2001). Till skillnad från den lexikografiskt noggrant presenterade isländska fraseologin har svenskans fraseologi inte på samma sätt varit föremål för en samlad lexikografisk beskrivning. För att fullt ut kunna realisera de planer som presenterats av Hannesdóttir & Jónsson (2001) krävs därför att en onomasiologiskt ordnad samling av svenska fraser av olika slag skapas. Arbetet med att upprätta en databas med svenska fraser och att ordna dessa fraser i begreppsfält har pågått parallellt – om än i blygsam omfattning – med arbetet med den isländsk-svenska ordboken i ISLEX. Denna databas kommer att komplettera beskrivningen av det svenska målspråket i form av en självständig, onomasiologiskt ordnad komponent (se Hannesdóttir, Jónsson & Tingsell 2010).

3. Onomasiologiska system i ett kontrastivt perspektiv

Onomasiologiskt ordnade ordböcker har långa anor, betydligt längre än de alfabetiska. Den ordbok där svenska förekommer för

första gången är just en sådan, *Variarum rerum vocabula* (1538). Utgångsspråket är latin och de latinska orden förses med svenska motsvarigheter. Arbetet inleds med ett avsnitt om gudliga och bibliska ting och om tiden: ”De deo coelo et temporibus”. Dessa ordböcker över allmänspråket var en väl etablerad och omfattande lexikografisk genre till, för svenskans del, långt in på 1700-talet. Ofta var de flerspråkiga, som exempelvis den fyrspråkiga *Wörter-Büchlein* från 1705 som nyligen publicerats i en ny, vetenskaplig utgåva. Medan sådana två- eller flerspråkiga onomasiologiskt ordnade allmänspråkiga ordböcker hör historien till är emellertid flerspråkiga taxonomiska nomenklaturer och standarder ett viktigt område inom bland annat terminologin (se exempelvis Aitchison, Gilchrist & Bawden 1997:135 ff.).

Också inom språkteknologin har man arbetat med onomasiologiska system. På sina håll pågår ett arbete med att bygga upp sarspråkliga motsvarigheter till det taxonomiskt ordnade engelskspråkiga WordNet (se bl.a. Lindén & Carlson 2010 och Vossen & Fellbaum 2009). Vidare har det arbetats med en flerspråkig motsvarighet inom ramen för EuroWordNet-projektet. Det är dock inte arbeten av detta slag som står i fokus här.

De efterföljare som Roget's engelska tesaurus (1852 och senare) har fått är enspråkiga och i de flesta fall avsedda för L1-talare (se dock Schottmann & Petersson 2004 och Burkhanov 2005). Den lexikografiska forskningen har, i den mån den uppmärksammat denna typ av ordböcker, hittills inte i någon större utsträckning anlagt ett kontrastivt perspektiv på indelningen i begreppsält.

4. Fraseologi i ett kontrastivt perspektiv

De senaste decennierna har den fraseologiska forskningen i allt större utsträckning kommit att omfatta ett kontrastivt perspektiv. Inom den tvåspråkiga lexikografin har fraseologins plats foku-

serats (Cowie [1994] 2008, Lubensky & McShane 2007) liksom inom främmandespråks- och andraspråkspedagogiken.

4.1. Lexikografiska aspekter

Fraseologi i ordböcker över allmänspråket är utrymmeskrävande. Det är därför först med de elektroniska ordböckerna som det har blivit möjligt att göra frasem av olika slag rättvisa såväl i källspråket som i målspråket. Lemmatiseringen är inte längre ett problem och lexikografen måste inte välja mellan å ena sidan användarvänliga men platskrävande upprepningar och å andra sidan platsbesparande men för användaren frustrerande hänvisningar. Men medan många av de tekniska problemen har undanröjts återstår andra lexikografiska problem och nya tillkommer. Exempelvis råder olika meningar bland lexikografer och metalexikografer om i vilken utsträckning, hur och var frasem av olika slag ska behandlas i en tvåspråkig ordbok. Framför allt är det fasta ordkombinationer med lexikaliserad betydelse och/eller konventionaliserad funktion som vållar målspråkslexikografen problem (se Hannesdóttir, Jónsson & Tingsell 2010). Några av dessa diskuteras nedan.

När det gäller kollokationer finns det inom lexikografen en etablerad ståndpunkt att de är lätta att förstå, men kan vara svåra att producera. Däremot är idiom, såväl pragmatiska som metaforiska, erkänt svåra att avkoda. Att det inte råder någon enighet om hur dessa ska behandlas i en tvåspråkig ordbok framgår tydligt av den diskussion som refereras och utvecklas vidare av Farø (2004). Idiom, menar han, utgör distinkta enheter i ett språks ordförråd och bör behandlas som sådana i den tvåspråkiga ordboken. Däremot hävdar han att ”der er meget få situationer, hvor man strengt taget har brug for at producere idiommer på et fremmedsprog” (Farø 2004:100). Denna ståndpunkt kan behöva nyanseras. Det är kanhända inte *nödvändigt* att producera idiom på ett andraspråk eller främmande språk – men man kanske vill.

4.2. Inläraraspekter

Inom andraspråksforskningen har det de senaste åren kommit många studier som rör L2-talare och deras användning av det som vi här kallar frasem. När det gäller kollokationer tycks det råda enighet om att de lärs in sent i inlärningsprocessen (Arnaud & Savignon 1997) och att inläringen av dem går långsamt (Groom 2009). Om hur svåra de är och vari svårigheten består finns det emellertid skiftande meningar. Exempelvis visar en studie av danska studenters förståelse av engelska partikelverb att det ”för langt de flestes vedkommende [er] vanskeligt at gennemskue deres betydning, da den ikke kan udledes af betydningerne af de enkelte dele af flerordsverbet” (Andersen 2010:56). Henriksen däremot utgår ifrån att frasem är lätta att förstå. Eftersom de inte nödvändigtvis innebär problem vid avkodning för en L2-talare uppfattas de heller inte som problematiska och uppmärksammas därför inte särskilt av denne (Henriksen 2010).

Behärskningen av sådana enheter har på flera sätt stor betydelse för L2-talaren. Att kunna producera frasem är viktigt för idiomaticiteten: ”Att tala ’som en infödd’ dvs. att tala idiomatiskt, är uppenbarligen något annat och något mer än att producera grammatiskt korrekta yttranden” (Ekberg 2004:269). Det är inte bara i produktion som frasemen är viktiga. Det anses också vara en kognitiv fördel att ha tillgång till prefabricerade enheter, ”chunks”, när man talar ett språk som inte är ens L1.

De svårigheter som konventionaliserade uttryck orsakar L2-talare kan vara av olika slag. De kan dels vara av kvantitativ natur, d.v.s. bero på att antalet kombinationer de behäskar är litet, dels av kvalitativ natur, d.v.s. bero på hur idiomatiska de uttryck de behäskar är (Gyllstad 2007). Typiskt för L2-talare av ett språk är att de använder frasem i mindre omfattning än modersmåls-talare. Ekberg visar att barn som har svenska som L2 använder färre konventionaliserade strukturer än jämnåriga L1-talare av

svenska. Detta, menar Ekberg, leder till att L2-talarna har färre ”färdiga val” (Ekberg 2004:272) att tillgå än L1-talarna, vilket i sin tur också gör dem mer beroende av produktiva, grammatiska regler. L1-talarna blir på det sättet mindre förutsägbara än L2-talarna och Ekberg drar därför den tillsynes paradoxala slutsatsen att det kanske just är ”stereotypin som i vissa fall utmärker ett idiomatiskt språkbruk!” (2004:274).

Fraseologisk kännedom är inte bara viktig för idiomaticiteten. Den ger också vissa kognitiva fördelar. Ur ett L2-perspektiv är alltså förekomsten av frasem i ordboken ett stöd för tillägnandet av en mer modersmålslik behärskning av språket. Det är därför motiverat med en generös redovisning av frasem i lexikografiska arbeten som vänder sig till L2-talare.

Trots denna pedagogiska insikt ägnas kollokationer inte så stor uppmärksamhet i andraspråksundervisningen. Formell undervisning tenderar att fokusera på inläring av nya ord; syntagmatiska förbindelser med redan inlärd ord får mindre utrymme (Henriksen 2010). Det ligger i sakens natur att kollokationer och andra typer av konventionaliserade uttryck heller inte förekommer lika frekvent i språket som de enskilda ord de består av. Inläraren exponeras därför inte för uttrycken i någon större omfattning. Medan L1-talarna utifrån sin modersmålskompetens kan välja mellan flera möjliga kollokationer till en given bas är L2-talaren hänvisad till en uppsättning inlärd kollokationer utan några större variationer. Dessa s.k. *collocational teddy bears* (Hasselgren 1994) kan bidra till att talarens språk uppfattas som mindre idiomatiskt än det i andra avseenden faktiskt är (jfr Ekberg 2004).

En god behärskning av många kollokationer bidrar alltså till att talaren uppfattas som mer idiomatisk och inföddlik i sin produktion. Att ha tillgång till sådana uttryck innebär också andra fördelar. Pawley & Syder (1983:218) visar att tillgången till färdigpaketerade enheter gör det lättare att processa språket i realtid. Som Boers m.fl. påpekar erbjuder detta möjligheter till ”säkra

kort”, d.v.s. uttryck som talaren vet med sig att han eller hon behärskar (Boers, Eyckmans & Stengers 2006).

5. Isländsk fraseologi och svenska motsvarigheter i ISLEX

Som redan påpekats utgörs ordförrådet i ISLEX-databasen av den moderna isländskan ”med särskild uppmärksamhet på att presentera ett rikt urval av kollokationer, idiom och språkprov med åtföljande översättningar till målspråken” (<http://islex.se/se&um=1>). I en multifunktionell ordbok av det här aktuella slaget är det naturligt att fråga sig vem de isländska frasemen är till för. Rent praktiskt är det den isländska redaktionen som väljer och lemmanätter frasemen.

Ur strikt lexikografiskt perspektiv skiljer sig de olika målgruppernas, islänningarnas respektive skandinavernas, behov åt när det gäller information om olika typer av frasem. I ISLEX presenteras kollokationer i vid bemärkelse, dvs. ”regelbundet återkommande ordförbindelser” (Svensén 2004:10) ofta (men inte alltid) både i den artikel som har basen som ingång och i den artikel som har kollokatorn som ingång. Exempelvis redovisas kollokationen *seðja hungrið* (stilla hungern) både under *seðja* och *hungur* medan *taka ákvörðun* (fatta beslut) endast redovisas under basen, *ákvörðun*, och inte i den omfångsrika artikeln *taka*.

Idiomen i ISLEX förses med en svensk motsvarighet av något slag. Finns det ett svenskt idiom som förefaller vara tillräckligt ekvivalent såväl semantiskt som funktionellt ges det som *en* möjlig ekvivalent. Bedömningen av idiomequivivalensen är problematisk på flera sätt. Ett inom lexikografin mindre vanligt problem som beror på språkens nära släktskap uppstår när det i svenskan finns ett idiom där komponenterna i de båda språken är morfologiska kognater. Trots att det inte är idiomets bokstavliga eller komposi-

tionella betydelse som motiverar dess förekomst i ordboken kan den uttrycksmissiga likheten skymma idiomets metaforik. De ingående orden är uppenbarligen ekvivalenta. Ett idiom som det isländska *brenna allar brýr að baki* (och den längre varianten *brenna allar brýr að baki sér*) illustrerar problemet väl:

brenna allar brýr að baki (sér)

bränna alla broar (bakom sig)

Det isländska idiomet är såväl funktionellt som semantiskt ekvivalent med det svenska idiomet *bränna alla broar* (respektive *bränna alla broar bakom sig*). Som framgår är de enskilda komponenterna i idiomerna också kognater eller interlinguala allomorfer och teckensträngarna blir därmed också nominellt ekvivalenta. Idiomets kompositionella betydelse, där fysiska broar sticks i brand, kan därigenom vilseleda användaren så att den metaforiska ekvivalensen inte uppfattas. Likheten mellan de båda språken ställer sålunda, i fall som detta, speciella krav på explicit och entydig information för att den funktionella ekvivalensrelationen ska framgå. En strategi kan, där så är möjligt, vara att prioritera ett funktionellt sett ekvivalent idiom på målspråket, ett idiom som inte enbart innehåller kognater till källspråksidiomet. Ur det perspektivet kan varianten *bränna sina skepp* därför tydligare förmedla den metaforiska och funktionella ekvivalensen än *bränna alla broar (bakom sig)*. Det är dock sällan som det finns en uppsättning med alternativa, likvärdiga idiom att välja bland.

Även om frasem av olika slag uppmärksammas särskilt i IS-LEX-ordböckerna är de inte renodlade idiomordböcker. Det finns därför ingen uttalad strävan efter att i första hand förse ett isländskt idiom med ett svenskt idiom med en i sammanhanget central metaforisk betydelse. Det som prioriteras är snarare den funktionella ekvivalensen. Ett idiom som *leika á als oddi* kan därför behandlas på olika sätt. Det tillstånd som detta idiom beskriver

kan uttryckas med en rad semantiskt och pragmatiskt ekvivalenta uttryck: *vara på ett strålande humör, vara överlycklig, vara stormförtjust, vara alldeles till sig* m.fl. Här ställs därför målspråkslexikografen inför ett urvalsproblem: är *vara på ett strålande humör* en lämpligare ekvivalent än, låt säga, *vara överlycklig*, eller helt enkelt *vara jätteglad*?

Problemet blir tydligt också vid vissa kollokationer. Utan att gå in på Mel'čuks lexikala funktioner (se exempelvis Svensén 2004:213 ff.) kan man konstatera att när vädret är bra på Island talar islänningen om *gott veður* eller till och med om *fallet veður*. I svenskan är *gott väder* semantiskt ekvivalent med *gott veður* men det är däremot lågfrekvent. Vädret är istället *fint, bra* eller *vackert*.

Det elektroniska formatets i det närmaste obegränsade utrymme tillåter i princip lexikografen att ge alla möjliga ekvivalentalternativ, försedda med all den kompletterande information som krävs för att användaren ska kunna avgöra vilket av alternativen som är det korrekta i det aktuella fallet. Denna lösning ställer emellertid stora krav såväl på användaren som på lexikografen. Istället för att eftersträva denna heltäckande redovisning kommer de svenska motsvarigheter som frasemen förses med i ISLEX-databasen i förekommande fall att länkas till en onomasiologiskt ordnad, svensk fraskomponent i en egen, separat databas som nu är under uppbyggnad: Frasbanken.se.

6. Frasbanken

I Frasbanken redovisas svenska frasem av olika slag. Genom att tillhandahålla flera alternativa, närsynonyma uttryck för ett givet innehåll ska Frasbanken tjäna som stöd för i första hand isländska ISLEX-användares men även andra L2-talares produktion av texter på svenska. Som utgångspunkt för indelningen i begrepps-fält används den struktur som Jónsson etablerat i *Orðaheimur*

(Jónsson 2002, se vidare Hannesdóttir, Jónsson & Tingsell 2010). Vissa modifieringar av fältindelningen har dock varit nödvändiga, främst med tanke på de olika målgrupperna för Frasbanken respektive *Orðaheimur*. Medan perspektivet i *Orðaheimur* är enspråkigt isländskt och arbetet primärt avsett för islänningar ska frasemen i Frasbanken vara tillgängliga både för användare av den isländsk-svenska ordboken i ISLEX och som självständig nätresurs för andra svenska L2-talare och även för andra svensktalande.

För insamling av material till Frasbanken har flera källor använts. *Orðaheimur* har utnyttjats, dels direkt genom att motsvarande svenskt frasem till där presenterade isländska frasem i förekommande fall har inkluderats i Frasbanken, dels indirekt som källa till de isländska fraserna i ISLEX. Bland de svenskspråkiga källorna kan nämnas den onomasiologiskt ordnade *Svenskt ordförråd* (Bring 1930), den semasiologiskt ordnade konstruktionsordboken *Svenskt språkbruk. Ordbok över konstruktioner och fraser* (2003) samt den svensk-tyska *Wörterbuch der schwedischen Phraseologie in Sachgruppen* (Schottmann & Petersson 1989). Dessutom har en rad andra svenska ordböcker använts. Särskilt ska där nämnas *NEO* (1995–1996) och *Svensk ordbok utgiven av Svenska Akademien* (2009) vars redaktioner har gett oss tillgång till alla flerordsenheter som ingår i deras material, liksom den elektroniska resursen SALDO (se vidare Borin & Forsberg 2011).

Frasemkandidaterna har provats mot olika korpora. Främst har vi använt Språkbankens korpusar och Google. Eftersom Frasbanken inte är avsedd att fungera som avkodningsverktyg bedöms kandidater som uppträder sparsamt i dessa källor som mindre aktuella och de har inte införlivats i Frasbanken. De frasem som presenteras ska vara möjliga att använda i samtida svenska.

Förbindelsen mellan ISLEX och Frasbanken består av en begreppsmarkering som kollokationer och idiom i ISLEX förses med. Denna begreppsmarkering länkas till motsvarande fält-rubrik i Frasbanken.

6.1. Frasemen i Frasbanken

I Frasbanken har vi inkluderat frasem av de slag som Hannesdóttir & Jónsson (2001) kallat ordförbindelser och bruksmönster. Bland dessa återfinns så gott som samtliga typer av flerordsuttryck som diskuterats av Farø & Lorentzen (2009:77ff). Frasbanken innehåller således kollokationer (han är en *glad prick*) och idiom (hon *slog klackarna i taket*), liknelser (han var *glad som en gris*), ordpar (hon var *nöjd och glad*) och i viss mån ordspråk. Uppsättningen av flerordskonstruktioner är naturligtvis inte uttömmande. Syftet är att presentera ett urval av den uttryckspotential som konventionaliserats inom de aktuella begreppsfälten. Presentationen är avsedd att vara till nytta i produktion av svenska för L2-talare, snarare än i avkodning (se ovan).

Som framgår av Hannesdóttir, Jónsson & Tingsell (2010) får rutinformler och pragmatiska idiom stort utrymme i Frasbanken. Här redovisas mycket förutsägbara rutinformler såsom hälsnings- och välgångsfraser, men också vanliga uttryck vid exempelvis kontroll (*försök inte!*) och rädsla (*jag törs inte*). Pragmatiska idiom av den här typen kan uppfattas som fria när det gäller selektionsrestriktioner medan de i själva verket har genomgått en betydande konventionalisering beträffande form. Vissa flerordsenheter uppfattas därför som mer idiomatiska än andra utan att det är frågan om någon metaforisering. Det pragmatiska idiomet *Vad glad jag blir!* för att kommunicera reaktionen på en glad överraskning är således idiomatiskt på ett sätt som det fullt grammatiska och semantiskt ekvivalenta *Jag blir glad nu!* inte är. En särskild typ av pragmatiska idiom som bereds plats i Frasbanken är samtalsstrukturerande uttryck av olika slag, s.k. *gambits*. Gambits har definierats som ”expressions [...] involved in organizing turn-taking, indicating a speaker’s attitude to his interlocutor, and generally ensuring smooth interaction between speakers” (Cowie [1994] 2008:167, se även Keller 1979). Gambits kan bestå såväl av satser

(*men vänta lite nu!*) som av enklare fraser (*eller hur?*) och enskilda diskurspartiklar.

Frasbanken innehåller också två typer av ettordsenheter: kärnord och rematiska uttryck. Kärnorden är ettordsenheter som är centrala i ett begreppsält. I fältet GLÄDJE (se figur 1 nedan) utgör adjektivet *glad* ett kärnord. Centrala ettordsenheter av denna typ tas med i Frasbanken och presenteras där i kontext i form av enkla, standardiserade verbfraser. Fraser av typen *hon var glad* motiveras av att där redovisas ett kärnord snarare än av att ordförbindelsen är konventionaliserad.

De rematiska uttrycken kan som sagt utgöras av ettordsenheter. Dessa är inte nödvändigtvis lika centrala i begreppsältet som kärnorden. Typiskt för de rematiska uttrycken är att de vanligen förekommer predikativt och att de tenderar att förbises i lexikon (se Hannesdóttir och Jónsson 2001:68). I Frasbanken presenteras de, precis som flerordsenheter och kärnorden, alltid i verbfrasform. Ett ettordsuttryck som mycket sällan ingår i en utsagas tema utan så gott som uteslutande förekommer som rema är adjektivet/adverbet *stormförtjust*. Det presenteras under fältet GLÄDJE i frasen *hon var stormförtjust*. Också frasem kan vara av sådan natur att de endast förekommer predikativt och sådana uttryck prioriteras i Frasbanken. Där redovisas därför frasem av typen *till sig i trasorna* i frasen *hon var till sig i trasorna* (se figur 1 nedan).

6.2. Frasemen och andraspråksinläraren

De ettordsuttryck av olika slag som redovisas i Frasbanken ska bidra till en balans i det ordförråd som presenteras och till en mer idiomatisk språkbeskrivning. Få språkbrukare väljer alltid flerordsenheter för att kommunicera ett visst innehåll. En fras-samling som endast redovisar metaforiska uttryck och därmed besläktade flerordsenheter kan ge intryck av att språket är mer figurativt än det nödvändigtvis måste vara.

Vid urvalet av frastyper följs i stort sett definitionen av termen *construction* som den ser ut hos t.ex. Goldberg (2006). Hon menar, intressant nog, att även flerordsenheter som inte är konventionaliserade i vanlig mening också måste sägas vara ett slags konstruktioner i en talares vokabulär. Goldberg har en kognitiv utgångspunkt för sitt resonemang, men det är relevant även ur ett lexikografiskt perspektiv. Ett kognitivt anslag är också fruktbart om man utgår från att ordboksanvändarna behöver hjälp att bygga upp en vokabulär som liknar en L1-talares:

Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency. (Goldberg 2006:5)

Frasbankens utformning ställer inga krav på att användaren kan urskilja de olika typerna av fraser redovisade ovan (för en diskussion kring explicit redovisade kategorityper, se t.ex. Sköldbberg (2009:228 f.) med referenser). Däremot kan information som inte inkluderas, men som en användare skulle kunna förvänta sig, vara ett potentiellt problem för användaren. Sådan information, såsom exempelvis valensinformation, har vi valt att utesluta eftersom den redovisas utförligt i andra lättillgängliga större svenska ordböcker och i viss mån även i ISLEX. Frånvaron av redovisade valensmöjligheter i Frasbanken ska därför inte tas till intäkt för att dessa inte finns i språket.

6.3. Strukturen i Frasbanken

Frasemen i Frasbanken är onomasiologiskt strukturerade på så sätt att de är ordnade efter sitt innehåll och inte efter de enskilda

ord som de består av. De är indelade i ca 700 semantiskt motiverade begreppsfält. Begreppsfälten förses med en (vanligen) substan-

GLÄDJE

vara glad

<hon> var glad
 <hon> var uppåt [vard.]
 <hon> var på gott humör
 <hon> blev varm om hjärtat
 <hon> gladdes över <segern>
 <hon> var uppsluppen

se glad ut

<hon> strålade (<av glädje>)
 <hon> var idel leenden

bli glad – synlig förändring

<hon> blev glad över <beskedet>
 <hon> sken upp
 ansiktet lystes upp av ett leende

glädje blandad med stolthet

<hon> var belåten
 <hon> var tillfreds

glädjas med någon annan

<hon> delade <hans> glädje
 <hon> gladdes med <pristagarna>
 <hon> var glad för <hans> skull

stor glädje

<hon> var i sjunde himlen
 <hon> var lycklig
 <hon> var överlycklig
 <hon> visste inte till sig av glädje
 <hon> var stormförtjust
 <hon> var själglad
 <hon> var lycksalig

<hon> var salig
 <hon> var upprymd
 <hon> var euforisk
 <hon> var till sig i trasorna [vard.]

allmän glädje

glädjen stod högt i tak
 det var inga sura miner
 det var muntert
 <de> slog klackarna i taket
 stämningen var hög

väcka glädje

<han> gladde <henne> (<med en present>)
 <han> gjorde <henne> glad
 <han> muntrade upp henne
 (<med en utlandsresa>)

något väcker glädje

<nyheten> togs emot med glädje

uttryck när något väcker glädje

det var goda nyheter
 det var (verkligen) roligt att höra
 vad glad jag blir

glädje vid återförening med någon

det var ett kärt återseende

uttryck för glädje vid möte med någon

vad roligt att se dig

Figur 1: Begreppsfältet GLÄDJE i Frاسبanken.

tivisk rubrik och indelas vidare i flera underavsnitt. Till skillnad från traditionella tesaurusar förgrenas begreppsfälten inte vidare i semantiska hierarkier. Nivåerna begränsas i princip till bara en nivå under varje begreppsfält. Begreppsfältet GLÄDJE (figur 1) får tjäna som exempel på artiklarnas struktur. Artikelns ovan är förkortad och visar ett urval av frasemen.

Lexikografisk notation används sparsamt i artiklarna. Parenteser används för led som kan uteslutas, medan vinkelparenteser används för led som måste realiseras men som får varieras. Användaren behöver inte förstå denna notation för att kunna använda resursen, men den användare som gör det får viktiga ledtrådar om frasemens struktur (och faktiskt även bruklighet – se nedan).

Strukturen inom varje artikel är beroende av begreppsfältets semantik. Några drag är emellertid återkommande. Artiklarna inleds alltid med så generella uttryck som möjligt och sedan presenteras, i görligaste mån, situationer som är mer specifika. Vissa fält låter sig organiseras kronologiskt. Fältet LIVSTID är ett sådant. Det inleds med översiktliga fraser som *han levde sitt liv i Stockholm* och avslutas med underrubriker som **livet är över**. Varje fält för sig kräver en semantisk analys som läggs till grund för fältets vidare struktur i underavdelningar. Underrubrikerna varierar därför mellan de olika artiklarna. Vissa strukturer är emellertid återkommande; rutinformlerna placeras under rubriken ”uttryck” närmast efter den underrubrik till vilken de hör. I figur 1 ovan följer rubriken **uttryck när något väcker glädje** omedelbart efter underrubriken **något väcker glädje**. Semantiska roller återkommer också som indelningsgrund inom betydelsefälten. Frasem och uttryck som har en AGENT placeras vanligen före frasem som har MOTTAGARE och liknande roller som subjekt. På samma sätt brukar animata subjekt placeras före inanimata. I fältet i figur 1 är underrubriken **väcka glädje** med fraser som *han glädde henne (med en present)* placerad före **något väcker glädje**.

I figur 2 nedan illustreras en del av begreppsältet KONTROLL. Först presenteras fraser där den som utövar kontrollen utgör frasemets (eller uttryckets) subjekt och omedelbart därefter rutinformler som agenten skulle kunna yttra vid kontrollutövning. Sedan presenteras frasem som hör till underrubriken *vara kontrollerad* d.v.s. där subjektet är den kontrollerade.

KONTROLL

ha kontroll över någon/något

<hon> höll efter <honom>

uttryck när man kontrollerar någon/något

Försök inte!

Jag ser dig allt!

vara kontrollerad

<han> hade ögonen på sig

Figur 2: Utdrag ur begreppsältet KONTROLL i Frasbanken.

Strukturen såsom den beskrivits ovan är en produkt av flera överväganden. Utgångspunkterna är framförallt två: relationen till ISLEX och det faktum att den primära målgruppen är isländska L2-talare av svenska. Projektet som sådant motiveras bland annat av att frasem av olika slag är ett stort problem för inlärare. Uppläggningsen har därför i stort och smått präglats av ett L2-perspektiv – från initiativet och den övergripande strukturen till ordvalet i de enskilda fraserna.

Som framgått ovan (se 4.2.) hör frasem till det som man lär sig sent i inläringen av ett andraspråk. De isländska användarna av den isländsk-svenska ordboken är just avancerade inlärare, inte minst genom att de har flera års studier av ett skandinaviskt språk bakom sig i den isländska skolan. För just denna grupp är konven-

tionaliserade uttryck något som ofta återstår att lära sig behärska. I det arbetet är Frasbanken avsedd att vara en resurs.

7. Den onomasiologiska ordboken ur ett andraspråksperspektiv

Som framgått ovan är onomasiologiska ordböcker vanligen enspråkiga och avsedda för användare som har den modersmåls-talandes förutsättningar att använda den information som ordböckerna ger. L2-talarnas behov skiljer sig från L1-talarnas och rör flera aspekter av ordbokens innehåll och utformning.

Användarna antas inte – till skillnad mot vad som förväntas av L1-talare – självklart ha kännedom om frasemens och uttryckens hemvist i olika stilar, register och kontexter och de antas heller inte kunna avgöra enheternas bruklighet. De är därför inte hjälpta med enbart okommenterade ”thesaurus blocks of equivalents” av den typ som förekommer i Kuzmin’s rysk-engelska fraseologiska ordbok (Lubensky & McShane 2007). För att hjälpa L2-talarna utan att ge direkta översättningar av uttrycken har vi arbetat med en liten uppsättning av bruksetiketter. Dessa används främst för att göra användaren uppmärksam på att vissa fraser är vardagliga och varna för att andra kan väcka anstöt. I figur 1 ovan har frasemet *hon var uppåt* markerats som vardagligt.

För att förse användarna med ytterligare information om ett frasems bruklighet och register utan att tvingas kommentera varje fras explicit, har vi utnyttjat de led i frasemet som måste realiseras men som får varieras. Sådana led markeras alltid med vinkelparenteser (se figur 1 och 2 ovan) som hjälper användaren att se vilka delar av ett frasem som är obligatoriska. Led som är fakultativa men mycket vanliga i eller kring frasemet står inom vanliga parenteser. I de fall där ett led är vanligt förekommande men kan varieras står det inom såväl vinkelparenteser som vanliga parenteser. De fakultativa leden har valts för att i görligaste

mån illustrera ett frasems bruklighet, register och kontexten för användningen. I frasem med brett användningsområde har vi valt personliga pronomen som platshållare i de led som måste realiseras. I vartannat begreppsält har vi, slumpmässigt, valt det maskulina *han*, i vartannat *hon*. Denna praxis har frångåtts bl.a. vid rollbyte (se figur 1) och för fraser som endast kan användas om ettdera könet. För frasem med mer restriktiv distribution har vi valt platshållare som illustrerar kontext och/eller register. Under begreppsältet ANSEENDE har t.ex. frasemet *átnjuta gott anseende* konstruerats med subjektet *biskopen*, eftersom det har hög stil och används i ett begränsat register. Frasen *vara uppskattad* har däremot konstruerats med det personliga pronomenet *hon*, eftersom det har ett vidare användningsområde.

Det faktum att de pragmatiska idiomerna har fått stort utrymme är också en anpassning till målgruppens behov. Medan en L1-talare knappast behöver hjälp med uttryck som *vad glad jag blir* under fältet GLÄDJE är just sådana uttryck viktiga för att L2-talaren skall uppfattas som idiomatisk och dessutom ha en pålitlig fras att ta till när en viss kontext påkallar det (se Pawley & Syder 1983 och Boers, Eyckmans & Stengers 2006).

Den platta hierarkiska strukturen är också en anpassning till L2-talarens behov. En mer hierarkisk struktur hade möjligen bidragit till en falsk föreställning om de semantiska hierarkiernas betydelse för ordförrådets struktur: "semantic space is multi-dimensional and [...] not a hierarchical structure" (Bukharov 2005:125). Istället upprepas delar av rubrikerna flera gånger (som i fallet rutinformler i figur 1 ovan).

8. Avslutning

En islänning som vill producera svensk text är den ena målgruppen för den isländsk-svenska ordboken i ISLEX. Dessa ISLEX-användare utgör också Frasbankens primära målgrupp. Frasbanken

har utvecklats för att fylla de behov dessa antas ha beträffande produktion av svenska konventionaliserade uttryck.

De isländska ISLEX-användarna har utomordentliga förutsättningar att snabbt uppnå en god behärskning av svenska. Språken är nära besläktade och danska ingår som obligatoriskt ämne under flera år i den isländska grundskolan. En inlärare som redan har kunskaper på såpass avancerad nivå har givetvis delvis andra behov än en nybörjare. Med utgångspunkt i sin skoldanska och med stöd av sin isländska modersmålskompetens har den isländske användaren goda möjligheter att få till en produktion på svenska som fungerar väl för vanliga kommunikativa syften – även utan att känna till konventionaliserade frasem. Han eller hon uppfattas bara som mindre idiomatisk. Frasbanken är avsedd att bistå dessa användare att uttrycka inte bara det de faktiskt *kan* utan också det de faktiskt *önskar* uttrycka – kanske till och med så nyanserat som de faktiskt *vill*.

Litteratur

Ordböcker

Bring, Sven Casper 1930: *Svenskt ordförråd ordnat i begreppsklasser*. Stockholm.

Jónsson, Jón Hilmar 1994: *Orðastaður. Orðabók um íslenska málnotkun*. Reykjavík: Mál og menning.

Jónsson, Jón Hilmar 2002: *Orðaheimur. Íslensk hugtakaorðabók með orða- og orðasambandaskrá*. Reykjavík: JPV útgáfa.

Jónsson, Jón Hilmar 2005: *Stóra orðabókin um íslenska málnotkun*. Reykjavík: JPV útgáfa.

NEO 1995–96 = *Nationalencyklopedins ordbok 1–3*. Höganäs: Bra böcker.

Norstedts nederländsk-svenska ordbok. Stockholm 2008.

- Roget, Peter Mark 1852: *Thesaurus of English Words and Phrases*. London.
- Schottmann, Hans & Rikke Petersson 1989: *Wörterbuch der schwedischen Phraseologie in Sachgruppen*. 1 upplagan. Münster: Kleinheinrich Verlag.
- Svensk ordbok utgiven av Svenska Akademien*. Stockholm: Norstedts 2009.
- Svenskt språkbruk. Ordbok över konstruktioner och fraser*. Stockholm: Norstedts Akademiska Förlag 2003.
- Variarum rerum vocabula* [1538] 1579. Utgiven i faksimil av Isak Collijn 1925. Stockholm.
- Wörter-Büchlein. A German-Swedish-Polish-Latvian Dictionary Published in Riga in 1705*. L. Larsson (ed.). 2011. Slavica Suecana Series A – Publications, Vol. 2. Stockholm: Kungl. Vitterhets Historie och Antikvitets Akademien.

Annan litteratur

- Aitchison, Jean, Alan Gilchrist & David Bawden 1997: *Thesaurus construction and use: a practical manual*. 3rd ed. London: Aslib.
- Andersen, Birger 2010: Etablering og strukturering af database til generering af tre elektroniske monofunktionelle ”phrasal verb” ordbøger. I: *Nordiska studier i lexikografi* 10, 56–71.
- Arnaud, Pierre J. L. & Sandra J. Savignon 1997: Rare words, complex lexical units and the advanced learner. I: J. Coady & T. Huckin (eds.): *Second language vocabulary acquisition*. Cambridge: Cambridge University Press, 157–173.
- Boers, Frank, June Eyckmans och Hélène Stengers 2006: Motivating multiword units. Rationale, mnemonic benefits and cognitive style variables. I: *EUROSLA* 6. Amsterdam: John Benjamins, 169–190.
- Borin, Lars & Markus Forsberg 2011: Swesaurus – ett svenskt ordnät med fria tyglar. I: *LexicoNordica* 18 (i denna volym).

- Burkhanov, Igor 2005: Requirements for a Learner-Oriented Onomasiological Dictionary of the General Vocabulary. I: *Lexicographica* 21, 122–130.
- Cowie, Anthony P. [1994] 2008: Phraseology. I: T. Fontenelle (ed): *Practical lexicography*. Oxford: Oxford university press, 163–167.
- Ekberg, Lena 2004: Grammatik och lexikon i svenska som andraspråk på nästan infödd nivå. I: K. Hyltenstam & I. Lindberg (red.): *Svenska som andraspråk – i forskning, undervisning och samhälle*. Lund: Studentlitteratur, 259–276.
- Farø, Ken 2004: Hvornår går man over åen efter vand? Idiomatiske ækvivalensproblemer i leksikografi og leksikologi. Med dansk og tysk som eksempel. I: *LexicoNordica* 11, 85–108.
- Farø, Ken & Henrik Lorentzen 2009: De oversete og mishandlede ordforbindelser – hvilke, hvor og hvorfor? I: *LexicoNordica* 16, 75–102.
- Goldberg, Adele E. 2006: *Constructions at work. The Nature of Generalization in Language*. Oxford: Oxford university press.
- Groom, Nicholas 2009: Effects of Second Language Immersion on Second Language Collocational development. I: A. Barfield & H. Gyllstad (eds.): *Researching Collocations in Another Language. Multiple Interpretations*. Basingstoke: Palgrave Macmillan, 21–33.
- Gyllstad, Henrik 2007: *Testing English Collocations, Developing Receptive Tests for Use with Advanced Swedish Learners*. Lund: Lund university.
- Hannedóttir, Anna Helga & Jón Hilmar Jónsson 2001: Að hafa í sig og á. Ísländsk fraseologi i ett ísländskt-svenskt perspektiv. I: *LexicoNordica* 8, 67–91.
- Hannedóttir, Anna Helga, Jón Hilmar Jónsson & Sofia Tingsell 2010: Mot en begreppsbasead ísländsk och svensk fraseologisk ordbok. Reflektioner kring pragmatiska idiom. I: *Nordiska studier i lexicografi* 10, 140–149.

- Hasselgren, Angela 1994: Lexical teddy bears and advanced learners: a study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics* 4, 237–258.
- Henriksen, Birgit 2010: Studying collocational competence and development – a special challenge for SLA researchers? Paper presented at the Approaches to the Lexicon conference, Copenhagen, 8th–10th of December 2010.
- Jónsdóttir, Halldóra & Þórdís Úlfarsdóttir 2008: ISLEX: Islandsk-skandinavisk webordbog. I: *Nordiske studier i leksikografi* 9, 197–208.
- Keller, Eric 1979: Gambits: Conversational Strategy Signals. I: *Journal of Pragmatics* 3, 219–238.
- Lindén, Krister & Lauri Carlson 2010: FinnWordNet – WordNet på finska via översättning. I: *LexicoNordica* 17, 119–140.
- Lubensky, Sophia & Marjorie McShane 2007: Bilingual Phraseological Dictionaries. I: H. Burger et al. (eds): *Phraseologie/Phraseology. An International Handbook of Contemporary Research*. Berlin/New York: de Gruyter, 919–928.
- Pawley, Andrew & Frances Syder 1983: Two puzzles for linguistic theory: nativelike selection and nativelike fluency. I: J. Richards & R. Schmidt (eds.): *Language and Communication*. London: Longman, 191–226.
- Sköldberg, Emma 2009: *Talesätt och talemaader: om ordförbindelser hos Dalin och Molbech*. I: *LexicoNordica* 16, 219–239.
- Svensén, Bo 2004: *Handbok i lexicografi. Ordböcker och ordboksarbete i teori och praktik*. 2 upplagan. Stockholm: Norstedts Akademiska förlag.
- Tarp, Sven 1995: Wörterbuchfunktionen: Utopische und realistische Vorschläge für die bilinguale Lexikographie. I: *Germanistische Linguistik* 127–128, 17–61.

Vossen, Piek & Christiane Fellbaum 2009: Universals and idiosyncrasies in multilingual WordNets. I: H. Boas (ed.): *Multilingual FrameNets in Computational Lexicography. Methods and Applications*. Berlin/New York: de Gruyter, 319–345.

Internethenvisningar

ISLEX = www.islex.se (april 2011)

LEXIN = <http://lexin2.nada.kth.se/lexin/> (april 2011)

SALDO = spraakbanken.gu.se/saldo/ (juni 2011)

Språkbanken = Spraakbanken.gu.se (juni 2011)

Anna Helga Hannesdóttir
docent, universitetslektor
Institutionen för svenska språket
Göteborgs universitet
Box 200
SE-405 30 Göteborg
Anna.Hannesdottir@svenska.gu.se

Sofia Tingsell
fil. dr, universitetslektor
Institutionen för svenska språket
Göteborgs universitet
Box 200
SE-405 30 Göteborg
Sofia.Tingsell@svenska.gu.se

Íslenskt orðanet: Et skritt mot en allmennspråklig onomasiologisk ordbok

Jón Hilmar Jónsson & Þórdís Úlfarsdóttir

Íslenskt orðanet (Icelandic wordnet) is an ongoing project which aims to analyse and describe the Icelandic vocabulary and its internal relations from the onomasiological point of view. The source material is mainly derived from a comprehensive lexicographical description of Icelandic collocations and phrasemes. The semantic relations between words and other independent lexical units are traced and identified by revealing their syntactic relations, thus building up a lexicographical database with approx. 250,000 single-word and multi-word lemmas. The wordnet combines the roles of a synonym dictionary and a thesaurus, placing special emphasis on verbal constructions and other multi-word lexical units. Within the groups of semantically closely related words, the characteristic lexical and syntactic patterns of the individual words are compared, which gives clues to their degree of semantic relatedness.

1. Innledning

I vår daglige omgang med ordbøker vil vi av og til oppleve at den ordbokstradisjon og ordbokskultur som omgir oss og som vi har mange gode grunner til å respektere, har satt normer og regler vi ikke lenger synes er i takt med tiden, eller som i hvert fall fortjener å bli vurdert på nytt. I mange tilfeller henger dette sammen med at vi blir kjent med nye muligheter til formidling av leksikografisk informasjon, og innser at tidligere begrensninger for ordboksarbeid og barrierer i leksikografens kommunikasjon med brukerne ikke lenger er til stede.

Dette forholdet gjelder ikke minst for vår oppfatning av de ty-

pologiske forskjeller mellom ordbøker og hvilke funksjoner som tildeles de enkelte ordbokstyper. Her tilsier tradisjonen at det er problematisk å tildele en ordbok mange og ulike funksjoner, og at de enkelte funksjoner (i hvert fall mange av dem) best kan oppfylles gjennom spesialordbøker av ulike slag.

Denne innstillingen er preget av lang erfaring med utarbeidelse av ordbøker i trykt form. Med framveksten av elektronisk databehandling og den elektroniske ordboken er dette bildet helt endret, og leksikografien står overfor den utfordring å vurdere verdien av inngrodd teorier og å lansere og teste nye ideer og initiativer.

I denne artikkelen blir denne problematikken drøftet med spesielt hensyn til den mest sentrale og prototypiske ordbokstypen, den allmennspråklige ordboken. Det grunnleggende spørsmålet er hvordan den best kan tilpasses de forutsetninger som ligger i den elektroniske ordboken, slik at den på en dynamisk måte kan formidle mest mulig allsidig informasjon.

Fra en brukersynsvinkel er det ønskelig å kunne forene ulike bruksfunksjoner i ett og samme ordboksverk, slik at brukerne ikke behøver å ha vesentlig forhåndskunnskap om de faktorer som utgjør forskjellen mellom de ulike ordbokstyper. Dette blir enda mer aktuelt når ordboken ikke lenger eksisterer i fysisk form og dens styrke og karakter primært gjenspeiles i hvordan den gir svar på de ulike spørsmål brukerne kommer på.

2. Den semasiologiske ordboken som prototype

Forskjellen mellom den semasiologiske og den onomasiologiske ordboken har lenge stått sentralt innenfor ordbokstypologien. Vi kan trygt si at her har den semasiologiske ordboken og det semasiologiske prinsippet hatt overtaket, og de mest allmenne og sentrale ordbøkene er av den typen. De fleste håndbøker om leksiko-

grafi bærer også preg av dette, f.eks. NLO (1997), Svensén (2004) og Fjeld & Vikør (2008), for å nevne tre sentrale verk innenfor nordisk leksikografi. I alle disse verkene får det onomasiologiske aspektet en forholdsvis liten vekt.

Den allmennspråklige ordboken er tradisjonelt den vanligste semasiologiske ordbokstypen. Her gir den alfabetiske makrostruktur brukerne klar og umiddelbar tilgang til de enkelte artikler, som hver for seg innledes av et lemma i form av et konkret ord. Artiklene inneholder en strukturert beskrivelse av de ulike egenskaper som karakteriserer det aktuelle lemmaet. Hver artikkel er dermed en avgrenset enhet, og de relasjoner den har til andre artikler, er i stor grad usynlige.

Den allmennspråklige ordbokens karakteristiske egenskaper består bl.a. i følgende faktorer:

- Lemmaseleksjonen bestemmes av ulike og ofte uklare kriterier
- Antall lemnaer blir sterkt begrenset og «lemmaverdien» varierer
- Lemmalisten består av ettordslemnaer
- Samtlige (entydige og flertydige) lemnaer får betydningsbeskrivelse (og betydningsdifferensiering)
- Grammatisk informasjon er for det meste begrenset til ordklasse- og bøyningsangivelser til lemmaene
- Fraseologiske enheter er underordnet lemmaene, og deres andel og omfang styres (delvis) av forholdene i de enkelte artikler
- Skillet mellom ordforbindelser (kollokasjoner) og bruks-eksempler er gjerne uklart
- Fast og ofte komplisert artikkelstruktur gjør det vanskelig å foreta revidering og oppdatering av ordboksteksten

Dette betyr at ordboksteksten har et statisk preg, og brukeren vil i mange tilfeller måtte ha god innsikt i strukturen for å komme fram til den informasjon han/hun leter etter. Hvis teksten foreligger i elektronisk form, kan brukeren ved fritekstsøking finne fram til indre tekstenheter og elementer, men det vil ikke gi noe helhetlig bilde av de sammenhenger og relasjoner som ligger skjult i de leksikografiske data teksten inneholder.

3. Tilpasning til det elektroniske mediet

Det elektroniske mediet byr på en mer fleksibel og dynamisk presentasjon og framstilling av leksikografiske data enn det er mulig å få til innenfor rammen av en trykt ordbok. Mange av de ovennevnte faktorer gjelder ikke elektroniske ordbøker, der leksikografen lettere kan definere og bygge opp en helhetlig ordbok som fungerer som en bred og overgripende beskrivelse av ordforrådet til et gitt språk, de leksikalske enheter det er sammensatt av og deres ulike relasjoner slik de gjenspeiles i den faktiske språkbruken. Mange fundamentale sider ved ordbøker og ordboksarbeid blir derved satt i et nytt lys. Det gjelder for lemmaseleksjon og lemmatisering, strukturer og strukturelle sammenhenger og for informasjonstyper og deres innbyrdes relasjoner. Men det gjelder ikke minst forholdet mellom lemmaenes formbaserte og innholdsbaserte egenskaper (dvs. forholdet mellom det semasiologiske og det onomasiologiske aspektet) og hvordan dette kommer til syne for brukerne og i ordboksteksten (se her også Jónsson 2009a).

På bakgrunn av dette kan en allmennspråklig ordbok i elektronisk form tildeles en rekke funksjoner som en tilsvarende ordbok i trykt form ikke kan oppfylle, og som da må knyttes sammen på en mest mulig hensiktsmessig og brukervennlig måte. Det kan gjelde funksjoner som primært har tilhørt andre og mer spesielle ordbokstyper, men også funksjoner som en allmenn ordbok i kraft av

sin datamengde og en fleksibel data- og informasjonsstruktur best er i stand til å oppfylle.

4. Fraseologien som grunnlag

Tanken om en allmennspråklig elektronisk ordbok av den typen som blir tematisert her, har vokst fram under et mangeårig arbeid med islandsk leksikografi. Dette arbeidet var i første omgang konsentrert om fraseologien, og resulterte bl.a. i tre nært beslektede ordbøker i perioden 1994 til 2005, *Orðastaður* (1994, 2. utvidede utgave 2001), *Orðaheimur* (2002) og *Stóra orðabókin um íslenska málnotkun* (2005). Den første har en alfabetisk makrostruktur og kan defineres som en kombinatorisk ordbok (med en fraseologisk og morfologisk komponent) med hovedvekt på kollokasjoner og konstruksjoner. I *Orðaheimur* er målet å dokumentere faste ordforbindelser og idiomer. Hovedlemmalisten består av begrepslemmaer (i alfabetisk orden) med krysshenvisninger til beslektede lemmer, men i tillegg kommer et alfabetisk register over innholdsord i samtlige fraser der de enkelte frasene kan bli lokalisert ut fra en fast sorteringsregel. Den sistnevnte ordboken er resultatet av en sammenslåing av de to første med en kombinasjon av ord- og begrepslemmaer der kollokasjoner og konstruksjoner fra *Orðastaður* er inkludert i fraseregisteret. Denne ordboken utkom både som papirordbok og elektronisk på cd, og den fins nå også som nettordbok (<http://snara.is/>).

5. Utformingen av Íslenskt orðanet

Gjennom arbeidet med de fraseologiske ordbøkene ble de semantiske relasjonene mellom fraser og ordforbindelser i stigende grad trukket inn, noe som også påvirket og endret ordboksstrukturen

på viktige punkter. Erfaringen med dette ordboksprosjektet banet veien for et pågående prosjekt som forfatterne av denne artikkelen nå arbeider med i fellesskap og som har fått overskriften *Íslenskt orðanet*.¹ Her spiller dataene og datatypene til de fraseologiske ordbøkene fortsatt en viktig rolle, men det semantiske og onomasiologiske aspektet er skjerpet ytterligere ved omfattende strukturendringer og ved å utvide den semantiske markeringen og klassifiseringen til også å omfatte enkeltord av alle sentrale ordklasser.

Før vi går inn på sammensetningen og oppbyggingen av ordnettet, vil vi framheve noen grunnleggende trekk i den ordboksmodellen det skal representere:

- Betydningsbærende flerordsenheter (fraser) blir omformet til flerordslemmaer og får samme status som ettordslemmaer
- Frasematerialet og dermed flerordslemmaene har en primær funksjon i oppbyggingen av databasen, noe som preger ordbokens karakter
- Enhetlig behandling av lemnaene krever (etter mønster fra flerordslemmaene) semantisk entydiggjøring av samtlige ettordslemmaer (innholdsord i ordforbindelser)
- De enkelte relasjonstyper og deres innbyrdes sammenhenger står i fokus
- Gjennom analysen av dataene bygges det opp en semantisk komponent som dekker beskrivelsen av begrepsammenhenger og semantiske relasjoner innenfor ordforrådet
- Grammatisk tagging av flerordslemmaene muliggjør grammatisk sortering i vekselvirkning med den semantiske klassifiseringen
- Tilleggsdata i form av teksteksempler, ordforbindelser og nye lemnaer sikrer kontinuerlig oppdatering av ordboksbasen i takt med fortsatt analyse av dataene

1 Prosjektet er blitt støttet av det islandske forskningsrådet RANNIS.

5.1. Materiale og omfang

Materialet til ordnettet blir lagret i to adskilte databaser. Bearbeidelsen og analysen foregår i en redigeringsbase, men ved siden av den fins det en åpen nettbasert database der produktet av analysen blir tilgjengeliggjort i form av en leksikografisk beskrivelse og løpende blir supplert ettersom analysen skrider fram. Materialet er i hovedsak hentet fra følgende kilder:

- *Stóra orðabókin um íslenska málnotkun*
- En omfattende samling fraser basert på sitater i sitatarkivet til Orðabók Háskólans (www.arnastofnun.is/page/arnastofnun_gagnasafn_ordasambond)
- Korpus med tekster fra islandske aviser og tidsskrifter (<http://timarit.is>)
- Sideordnede islandske ekvivalenter til fremmedspråklige kildeord fra tre tospråklige ordbøker

Alle innholdsord som forekommer i frasene er inkludert i lemmalisten, sammen med alle sammensetninger (ca. 100 000) som opprinnelig tilhørte den morfologiske komponenten i *Orðastaður*. Gjennom analysen av materialet er lemmabestanden utvidet og supplert, for en stor del ved lemmatisering av betydningsbærende fraser som flerordslemmaer. Lemmalisten inneholder nå over 250 000 lemnaer fordelt på følgende måte:

- 125 000 substantiver
- 35 000 adjektiver
- 80 000 verb (og verbale fraser)
- 11 000 adverb og adverbialer

Denne store mengden lemnaer er et resultat av at ordnettet er ment å skulle dekke ordforrådet i dets mangfold og bredde og

gjengi de innbyrdes relasjonar som blir avdekket gjennom analysen av materialet. Det betyr at lemmalisten er åpen og uavgrenset, og beskrivelsen av de enkelte lemmaene blir komplettert og utvidet i takt med framdriften av analysen.

5.2. Lemmatisering og bearbeiding av materialet

Grunntanken bak ordnettet er at de syntagmatiske relasjonene innanfor kollokasjoner og andre typer fraser gjenspeiler semantisk slektskap mellom ord og andre leksikalske enheter. På samme måte vil paradigmatisk relasjonar mellom ordleidd i sammensetningar i mange tilfeller avdekke semantisk nærhet.

Forutsetninga for at denne kartlegginga av lemmaenes innbyrdes relasjonar kan realiseres, er at lemmaene blir entydiggjort. Dette kjemmar klarest til uttrykk blant substantivene. I lemmalisten blir flertydige substantiver framstilt som adskilte enheter i form av en homonymindeks (etterfulgt av en identifiserende forklaring):

- 1 *ráðning* no kvk (hirting)
- 2 *ráðning* no kvk (skipan í stöðu)
- 3 *ráðning* no kvk (lausn)

Eksemplet ovenfor viser hvordan det flertydige ordet *ráðning* med delbetydningene ‘straff’, ‘ansettelse’ og ‘løsning’ blir framstilt som tre adskilte lemmaer (no kvk står for substantiv (nafn-ord) hunkjonn (kvenkyn)). Sammen med lemmatiseringa av betydningsbærende ordforbindelser som flerordslemmaer utgjør entydiggjøringa et grunnleggjende tiltak for å få til en konsistent og standardisert lemmaliste som dekkjer hele datamengden og kan utvides etter behov.

Bearbeidinga av materialet består ellers i hovudsak i å gjennomgå fraser og sammensetningar og på bakgrunn av det å mar-

kere og gruppere semantisk nærstående ord. Det foregår delvis ved mekanisk prosessering og samgruppering, men for en stor del vil prosessen styres av leksikografens vurdering og innsikt i materialet. Figur 1 viser noen glimt fra artikkelen *þjóð* ‘nasjon’, der kollokasjoner ordet opptrer i, blir gruppert for å forene synonyme kollokatorer.

45:	
OSAMB: efnahagsástand þjóðarinnar	efnahagsástand no hvk
OSAMB: efnahagur þjóðarinnar	efnahagur no kk
OSAMB: fjárhagur þjóðarinnar	fjárhagur no kk
46:	
OSAMB: aldursdreifing þjóðarinnar	aldursdreifing no kvk
OSAMB: aldurssamsetning þjóðarinnar	aldurssamsetning no kvk
47:	
OSAMB: athafnalíf þjóðarinnar	athafnalíf no hvk
OSAMB: atvinnuhættir þjóðarinnar	atvinnuhættir no kk flt
OSAMB: atvinnulíf þjóðarinnar	atvinnulíf no hvk
48:	
OSAMB: bókhneigð þjóð	bókhneigður lo
OSAMB: bókmenntahungruð þjóð	bókmenntahungraður lo

Figur 1: Ordforbindelser (kollokasjoner) i artikkelen *þjóð* ‘nasjon’. De tre første gruppene inneholder genitivkonstruksjoner, den siste forener attributive adjektiver.

Synonym- og antonymrelasjoner står i fokus for dette arbeidet, slik at det bygges opp en synonym- og antonymkomponent innenfor ordnettet som både dekker ettords- og flerordslemmaer. Vurdering og markering av synonymrelasjoner er problematisk, bl.a. fordi det i praksis dreier seg om ulik grad av synonymi. Av hensyn til dette blir det ved analysen skilt mellom to typer, synonymer og nærsynonymer.

Ordnettet dekker hermed funksjonen til en synonymordbok som bygges opp kontinuerlig under bearbeidingen av materialet.

Men ordnettet skal også fungere som begrepsordbok ved å knytte sammen lemmaer som tilhører samme begrepsfelt. Oppbyggingen av denne komponenten foregår også etappevis uavhengig av synonymgrupperingen, men vil i mange tilfeller dra nytte av den samtidig som den fører sammen beslektede grupper synonymer.

Markeringen av antonymer bygger for det meste på leksikografens skjønn og vurdering og henger nøye sammen med markeringen av synonymer.

Mekanisk prosessering av de leksikografiske data underletter redaksjonsarbeidet i ulike sammenhenger og har dessuten en selvstendig verdi. Resultatet av slike operasjoner kan i visse tilfeller vises fram direkte til orientering for brukerne, selv om det dreier seg om grovsortering av materialet. Men det tjener først og fremst som grunnlag for å kunne definere og presisere det semantiske forholdet mellom beslektede ord og lemmaer, enten det gjelder synonymer eller mer overgripende semantiske relasjoner, og gjøre dette forholdet synlig for brukerne av ordboken.

Dette kan illustreres nærmere. En kollokasjon av typen adjektiv + substantiv kan gjennom en konstant substantivisk basis forene en rekke semantisk beslektede adjektiviske kollokatorer, f.eks. sammen med ordet *sjúkdómur* 'sykdom': *hættulegur* / *mannskæður* / *banvænn sjúkdómur* 'farlig / som koster mange livet / dødelig sykdom'. De enkelte adjektivene kan gi tilknytning til andre substantiver av samme karakter: *hættulegur faraldur* 'farlig epidemi', *mannskæð influensa* 'influenza som koster mange livet', *banvæn drepsótt* 'dødelig pest'. Disse substantivene kan trekke til seg flere ord gjennom andre syntagmatiske og paradigmatiske relasjoner, slik at det samles opp en lang rekke relaterte lemmaer. En god del av dette vokabularet vil høre sammen betydningsmessig. Iblant dreier det seg om ord fra andre semantiske felt, jo mer desto videre rammen er om relasjonene.

5.3. Flerordslemmaenes sentrale posisjon

Som allerede nevnt utgjør flerordslemmaene en betydelig del av lemmalisten og er spesielt karakteristiske for verbene og adverbialene (se også en nærmere beskrivelse i Jónsson 2009b:180f.). For å gi dem selvstendig lemmastatus og sorteringsgrunnlag på linje med ettordslemmaene får de grammatisk markering i form av tagger. Framstillingen er standardisert etter mønster av frasene i de ordbøkene de er hentet fra. Lemmalisten i figur 2 viser en liten del av de verblemmaer som innledes med verbet *leggja* ‘legge’, der den grammatiske taggen (som primært henviser til ordklasse og kasus) står bakerst i hakeparenteser.

395	<i>leggja niður vinnu sína so</i>	[so ao no-a fn-a]
396	<i>leggja niður vopn so</i>	[so ao no-a]
397	<i>leggja niður vopnin so</i>	[so ao no-ag]
398	<i>leggja niður völd so</i>	[so ao no-a]
399	<i>leggja niður <byrðina> so</i>	[so ao <no-ag>]
400	<i>leggja niður <embætti sitt> so</i>	[so ao <no-a fn-a>]
401	<i>leggja niður <fornan> óvanda so</i>	[so ao <lo-a> no-a]
402	<i>leggja nótt með degi so</i>	[so no-a fs no-d]
403	<i>leggja nótt við dag so</i>	[so no-a fs no-a]
404	<i>leggja ofuráherslu á <þetta> so</i>	[so no-a fs <fn-a>]

Figur 2: En liten del av de verblemmaer som innledes med verbet *leggja* ‘legge’; lemmaene er grammatisk tagget. Taggen so står for verb (sagnorð), no for substantiv (nafnorð), lo for adjektiv (lýsingarorð), fn for pronomen (fornafn), ao for adverb (atviksorð), fs for preposisjon (forsetning). Sammensatte tagger viser kasusmarkering, a står for akkusativ, d for dativ, g for genitiv; bestemt form angis ved g etter kasusmarkeringen.

Denne taggingen av verbene gir interessante muligheter til sortering og søking, og når enkelte verb kommer til syne i en lang liste med flerordslemmaer, gir det en oversikt over karakteristiske mønstre og dermed grunnlag for sammenlikning mellom ulike verb. Men man kan også fokusere på enkelte konstruksjonsmønstre og deres utbredelse innenfor lemmaene og dermed knytte sammen formmessig sammenhengende lemmaer, som vist i figur 3.

249	<i>brunatryggja</i> <húsið, innbúíð> <i>so</i> [so <no-ag>]
250	<i>brúa</i> <lækinn, ána> <i>so</i> [so <no-ag>] <i>cc</i>
251	<i>brúka</i> <hestinn> <i>so</i> [so <no-ag>]
252	<i>brúleggja</i> <ána, stéttina> <i>so</i> [so <no-ag>] <i>cc</i>
253	<i>brúna</i> <steikina; kartöflurnar> <i>so</i> [so <no-ag>]
254	<i>brúnelda</i> <steikina> <i>so</i> [so <no-ag>]
255	<i>brydda</i> <skóna> <i>so</i> [so <no-ag>] <i>cc</i>
256	<i>brytja</i> <kjötið, mörinn> <i>so</i> [so <no-ag>] <i>cc</i>
257	<i>brýna</i> <
258	<i>bræða</i> <
	56 <i>festa</i> <plankanum> <i>so</i> [so <no-dg>]
	57 <i>flarstýra</i> <tækinu> <i>so</i> [so <no-dg>]
	58 <i>fletta</i> <blaðinu> <i>so</i> [so <no-dg>] <i>cc</i>
	59 <i>fletta</i> <trénu; borðunum> <i>so</i> [so <no-dg>]
	60 <i>fleygja</i> <ruslinu> <i>so</i> [so <no-dg>] <i>cc</i>
	61 <i>flota</i> <bátnum> <i>so</i> [so <no-dg>] <i>cc</i>
	62 <i>flugríða</i> <hestinum> <i>so</i> [so <no-dg>]
	63 <i>flýta</i> <suðunni> <i>so</i> [so <no-dg>]

Figur 3: Den grammatiske taggingen forener verb som viser samme konstruksjonsmønster. I spalten til venstre består konstruksjonen av et verb, etterfulgt av et varierende objekt i akkusativ (bestemt form). Spalten til høyre viser en tilsvarende dativkonstruksjon.

Flerordslemmaene muliggjør også i mange tilfeller leksikalsk betingede søk, f.eks. på bestemte substantiver som opptrer som nominaler i de sammensatte lemmaene. Substantivet *eldur* ‘ild, brann’ knytter f.eks. sammen omtrent 50 lemmaer der det står i ulike funksjoner, som subjekt (<*eldurinn*> *gýs upp* ‘brannen bryter ut’, <*eldurinn*> *logar* ‘ilden brenner’, <*eldurinn*> *breiðist út* ‘brannen sprer seg’) eller objekt (*glæða* <*eldinn*> ‘friske på ilden’, *slökkva* <*eldinn*> ‘slukke ilden’). Funksjonelle enheter innenfor lemmaene kan også være av interesse, som kombinasjonen *eins og* ‘som’ i sammenlikningsledd, som forekommer bortimot 900 ganger i ulike lemmatyper, bl.a. i lemmaene *drekka eins og svín* ‘drikke som et svin’, *lifa eins og kóngur* ‘leve som en konge’, *reykja eins og strompur* ‘røyke som en skorstein’.

Den grammatiske taggingen av flerordslemmaene er ikke minst til stor nytte ved den semantiske klassifiseringen av lemmaene, enten det gjelder synonym- eller begrepsrelasjoner, for formell likhet henger i ganske stor grad sammen med semantisk slektskap.

Denne beskrivelsen av flerordslemmaene har vært konsentrert om verbene, men flerordsforbindelser blir også lemmatisert innenfor andre ordklasser: substantiver (nominaler), adjektiver og ikke minst adverbialer. Blant adjektivene blir entydiggjøringen synliggjort i sammensatte lemmaformer av typen *beittur* [*hnífur*] ‘skarp kniv’, der nominalet i hakeparenteser har en identifiserende og differensierende funksjon overfor andre betydninger av adjektivet og samtidig utgjør et forenende element for adjektiver som karakteriserer vedkommende nominalbegrep: *bitur* [*hnífur*] ‘skarp kniv’, *bitgóður* [*hnífur*] ‘skarp kniv’, *egghvass* [*hnífur*] ‘kniv med skarp egg’ osv.

For adverbialenes vedkommende gir den grammatiske markeringen et verdifullt grunnlag for semantisk klassifisering av leksikalske enheter som tradisjonelt ikke har kommet til syne som selvstendige lemmaer, men som her inngår i semantiske og syntagmatiske relasjoner på linje med andre lemmaer. Figur 4 gir inn-sikt i omfanget av og variasjonen i de adverbiale lemmaene. Den første listen har et felles innledende element (*þegar* ‘når’) og den andre listen forener lemmaer med samme grammatiske tagg (konjunksjon + substantiv i nominativ + verb).

76	<i>þegar til lengdar lætur</i> <small>ao [st fs no-g so]</small>
77	<i>þegar til reyndanna kemur</i> <small>ao [st fs no-gg so]</small>
78	<i>þegar til stykkisins kemur</i> <small>ao [st fs no-gg so]</small>
79	<i>þegar tímar líða</i> <small>ao [st no-n so]</small>
80	<i>þegar tímar líða fram</i> <small>ao [st no-n so ao]</small>
81	<i>þegar um hægist</i> <small>ao [st fs so]</small>
82	<i>þegar upp er stadið</i> <small>ao [st ao so so]</small>
83	<i>þegar út á líð</i> <small>ao [st no-n so]</small>
84	<i>þegar úthalla</i> <small>ao [st no-n so]</small>
6	<i>ef þörf krefur</i> <small>ao [st no-n so]</small>
7	<i>meðan aldri renna</i> <small>ao [st no-n so]</small>
8	<i>meðan kraftar endast</i> <small>ao [st no-n so]</small>
9	<i>meðan land byggist</i> <small>ao [st no-n so]</small>
10	<i>meðan öld lifir</i> <small>ao [st no-n so]</small>
11	<i>sem fætur toga</i> <small>ao [st no-n so]</small>
12	<i>sem lög tilstanda</i> <small>ao [st no-n so]</small>
13	<i>þegar ferð fellur</i> <small>ao [st no-n so]</small>
14	<i>þegar færi gefst</i> <small>ao [st no-n so]</small>

Figur 4:
Adverbiale
flerordslemmaer.

5.4. Henvisninger og indikasjoner på semantisk slektskapsgrad

En sentral målsetting for arbeidet med ordnettene er å få innsikt i slektskapsforholdet mellom semantisk beslektede ord og lemmaer. Prosessering av leksikografiske data byr på interessante muligheter i denne sammenheng, og det har vært utviklet ulike metoder for å vurdere og måle slektskapsgraden (se bl.a. Nikulásdóttir og Whelpton 2010). I ordnettene er arbeidet med dette fortsatt under utvikling, men hittil består kartleggingen først og fremst i to observasjoner: a) sammenlikning av konkrete lemmaer ut fra felles tilknytning i relasjoner til andre lemmaer og b) sammenlikning ut fra antall felles partnere i parataktiske forbindelser (ordpar).

5.4.1. Felles lemmatilknytning

Artiklene i ordnettene blir strukturert slik at samtlige lemmaer som opptrer som partnere i ordforbindelser eller sammensetninger, blir oppført som lenker. Figur 5 viser to bruddstykker fra artikkelen *sigur* ‘seier’, som i alt dekker omtrent 70 ordforbindelser og fraser og ca. 100 sammensetninger, og som har lenker til bortimot 300 ettords- og flerordslemmaer. Spalten i midten inneholder de fraser og sammensetninger lemmaet opptrer i (sortert primært etter type og sekundært alfabetisk), spalten til høyre viser hvilke lemmaer som også er lenket til den aktuelle frasen eller sammensetningen, og spalten til venstre inneholder samtlige lenker (nøkkelord), sortert primært etter ordklasse og sekundært alfabetisk. I det første utsnittet er adjektivene i fokus, men i det andre får vi se (sammensatte) verblemmaer i den venstre spalten og sammensetninger i midten.

Word	Relation	Target
1 hjarntanlega so		
2 innilega so		
3 3d-fold		
4 sigðður lo		
5 auðunninn lo		
6 auðveldur lo		
7 djartur lo		
8 djúgur lo		
9 djúkþéttur lo		
10 djúður lo		
11 endanlegur lo		
12 endilegur lo		
13 flægur lo		
14 frækilegur lo		
15 frækinn		
16 fátur lo		
17 glæður		
256 bera sigur af hólmi so		
256 bera sigur úr bytum so		
257 berjaat til sigurs so		
258 eiga sigri að hróa so		
259 eiga sigurinn vísan so		
260 fagna sigri so		
261 fara með sigur af hólmi so		
262 fá sigur so		
	MERRHT: B 15, C 20, U, H 9, O 13, Rlo_m_no, Rlo_m_no,	
	GVID: viðtögn sigur (þr)	
	ANDHEITI: sigur - mistök	mistök no hvk flt
	ANDHEITI: sigur - ósigur	ósigur no kk
	PAR: hefðir og sigur	hefðir no kvk
	PAR: sigur og framfarir	framfarir no kvk flt
	OSAMB: berjaat til sigurs	berjaat til sigurs so
	OGAMH: dýrlegur sigur	dýrlegur lo
	OGAMH: eiga sigur/sigurinn vísan	vís lo
		eiga sigurinn vísan so
	ORDHLUTAR: sigur-þjóð	veitnigur so
		sigurþjóð no kvk
		þjóð no kvk
	ORDHLUTAR: sigur-þoræti	sigurþoræti no kk
		þoræti no kk
	ORDHLUTAR: sigur-ökur	sigurökur no hvk
		ökur no hvk
	ORDHLUTAR: stór-sigur	stórsigur no kk
		stór lo


Figur 5: To bruddstykker fra artikkelen *sigur* 'seier'.

Det framgår av dette at lemmatilknytningen til de enkelte lemmaene gir et godt grunnlag for sammenlikning mellom lemmaer, ut fra det resonnement at jo større likheter det dreier seg om, desto mer vil lemmaene være i slekt med hverandre. Figur 6 viser et avsnitt av en slik sammenlikning mellom lemmaene *krá* 'kro' og *knæpa* 'kneipe'. Spalten i midten inneholder de lemmaer som er lenket til begge to (47), mens spaltene til venstre og høyre inneholder de lemmaer som utelukkende er knyttet til det ene av de to. Antallet felles nøkkelord er her ganske høyt og det vil, ut fra sammenlikning med flere lemmaer i vedkommende felt, gi et verdifullt vitnesbyrd om den semantiske nærheten mellom disse to ordene.

255 bera sigur af hólmi so		
256 bera sigur úr bytum so		
257 berjaat til sigurs so		
258 eiga sigri að hróa so		
259 eiga sigurinn vísan so		
260 fagna sigri so		
261 fara með sigur af hólmi so		
262 fá sigur so		
	ORDHLUTAR: sigur-þjóð	veitnigur so
		sigurþjóð no kvk
		þjóð no kvk
	ORDHLUTAR: sigur-þoræti	sigurþoræti no kk
		þoræti no kk
	ORDHLUTAR: sigur-ökur	sigurökur no hvk
		ökur no hvk
	ORDHLUTAR: stór-sigur	stórsigur no kk
		stór lo

Figur 6: Sammenlikning av lemmaene *krá* 'kro' og *knæpa* 'kneipe' med hensyn til (felles) lemmatilknytning. Ikonene foran de relaterte lemmaene står for de ulike typer relasjoner, O står for ordforbindelse (kollokator), P for ordpar, H for sammensetningsledd, R for sidestilt ekvivalent.

5.4.2. Antall felles partnere i parataktiske forbindelser

I sammenlikningskartet i figur 6, som i selve artiklene, får de enkelte data- og relasjonstyper blant nøkkelordene en særskilt markering i form av et ikon. Blant disse har ikonet , som står for parataktisk relasjon (ordpar, som *diskótek og krár* ‘diskotek og kroer’, *kaffistofur og knæpur* ‘kaféer og kneiper’), høy frekvens. Dette gjenspeiler den erfaringen at slike relasjoner i svært mange tilfeller belyser det semantiske forholdet mellom nærstående ord, og en nærmere kartlegging av dette forholdet kan gi et godt bilde av hvor nært et gitt lemma er beslektet med andre lemmaer.

I ordnettet blir dette kartlagt på den måten at det fra databasen blir hentet tall om i hvor mange tilfeller et gitt lemma er relatert til andre lemmaer gjennom en felles partner i parataktiske relasjoner. Troverdigheten av denne sammenlikningen vil være avhengig av datamengden (antall ordpar) til de enkelte lemmaene. En sammenlikning av lemmaene *krá* og *knæpa* vitner klart om et nært semantisk slektskap mellom de to ordene, som det framgår av figur 7. Her blir det for hvert nøkkelord angitt i den nest siste kolonnen hvor mange ordpar vedkommende artikkel inneholder. Kolonnen foran angir antall ordpar med lemmaet i overskriften, og den siste kolonnen viser antall felles partnere. For eksempel forekommer *krá* i 90 ordpar med ulike partnere, mens *veitingahús* ‘restaurant’ er knyttet til 138 partnere i slike forbindelser, og 49 av disse partnerene er felles for begge to.

Figuren viser at 11 av de 12 lemmaene som har flest felles partnere med henholdsvis *krá* og *knæpa*, er de samme for begge lemmaene. En slik sammenlikning vil kunne være aktuell for ulike grader av semantisk slektskap, også for ord der det gjelder å bekreftede at et slikt forhold ikke er til stede.

krá:	knæpa:
1 veitingahús no hvk 90 138 49	1 veitingahús no hvk 61 138 35
2 veitingastaður no kk 90 115 44	2 krá no kvk 61 90 33
3 kaffihús no hvk 90 144 43	3 veitingastaður no kk 61 115 27
4 sjoppa no kvk 90 106 34	4 kaffihús no hvk 61 144 26
5 knæpa no kvk 90 61 33	5 bjórstofa no kvk 61 41 19
6 bjórstofa no kvk 90 41 24	6 sjoppa no kvk 61 106 19
7 öldurhús no hvk 90 33 20	7 öldurhús no hvk 61 33 16
8 vínstofa no kvk 90 32 16	8 gildaskáli no kk 61 38 16
9 gistihús no hvk 90 37 16	9 vínstofa no kvk 61 32 14
10 gildaskáli no kk 90 38 14	10 ölkrá no kvk 61 22 13
11 skemmtistaður no kk 90 14 10	11 gistihús no hvk 61 37 13
12 kvikmyndahús no hvk 90 12 9	12 skemmtistaður no kk 61 14 11

Figur 7: Antall felles partnere i parataktiske forbindelser:
Sammenlikning av lemmaene *krá* og *knæpa*.

Lemmaenes slektskapsforhold innenfor ordnettet er ennå ikke blitt kartlagt på denne måten i større målestokk. Men erfaringen hittil og de eksempler vi har vist, gir forventninger om at denne metoden kan være av stor verdi for en vesentlig del av lemmaene, spesielt substantivene og adjektivene. Ordpar som datatype er dessuten svært nyttig med hensyn til behovet for å utvide og komplettere ordforrådet på de enkelte semantiske felt.

6. Elektronisk ordbok på nettsiden www.ordanet.is

Ordnettet er hittil beskrevet ut fra dets oppbygning i den sentrale ordboksbasen. Den fungerer som redaksjonsbase, og den vil ikke være åpen for vanlige brukere. Men datastrukturen byr på ulike muligheter til å bygge opp et sammenhengende leksikografisk produkt i form av en elektronisk ordbok. En særskilt database er blitt utformet for dette formålet, og oppbyggingen av en slik ordbok har foregått i takt med framdriften i redaksjonsarbeidet. Produktet blir synliggjort på nettsiden *ordanet.is*, der det blir oppdatert regelmessig.

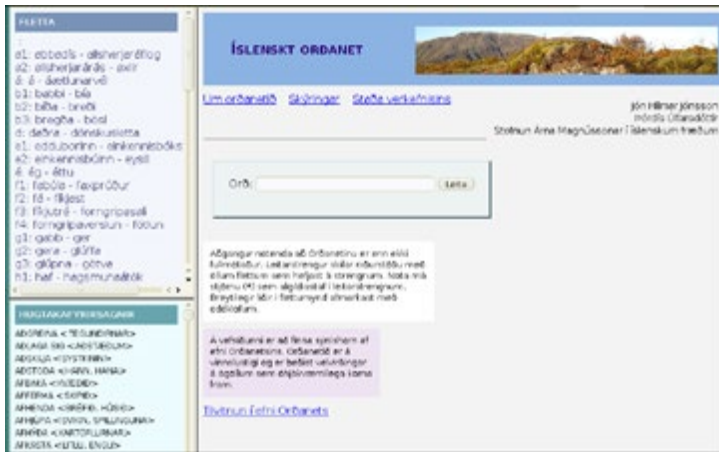
Her blir det særlig fokusert på semantiske og begrepsmessige

relasjoner, slik at ordboken kan sies å ha en kombinert funksjon som synonym- og begrepsordbok. Men samtidig bevarer den sterke elementer fra de kollokasjons- og konstruksjonsordbøker den opprinnelig bygger på.

Nettordboken omfatter per juni 2011 omtrent 150 000 lemmer, der hvert enkelt lemma står i en relasjon (én eller flere) til andre lemmer, som da fungerer som lenke(r) innenfor vedkommende artikkel. Disse lemmaene har svært ulik status. Noen er knyttet til et større begrepsfelt som dekker en hel rekke lemmer, andre er (foreløpig) bare relatert til ett bestemt lemma. Dette bildet endrer seg stadig ettersom redaksjonsarbeidet og analysen skrider fram, tidligere artikler blir utvidet, nye lemmer kommer inn og analysen blir oppdatert på ulike punkter.

En ordboksmodell av denne typen er svært ulik ordbøker der strukturen er innviklet og fastlåst og der brukerne står overfor en ferdig og urokelig ordbokstekst. Her gjelder det å gjøre brukerne bevisste på at ordboken er under kontinuerlig komplettering og oppdatering. Men den største forskjellen består i at de enkelte lemmaene ikke framstår som avgrensede og isolerte enheter, men står i et aktivt samspill med andre lemmer som kan utvide brukerens synsfelt i ulike retninger.

Som det framgår av figur 8, kan søk i ordboken rettes mot to adskilte lemmalister. Hovedlisten (FLETTA) er en formbasert lemmaliste som dekker samtlige lemmer i basen, og som man kan søke i via søkefeltet til høyre. Denne listen står i konsentrert form i den øverste rammen til venstre, der man kan klikke på en bestemt alfabetisk del. Nedenfor fins det en annen begrepsbasert alfabetisert lemmaliste (HUGTAKAFYRIRSAGNIR) med overskrifter til begrepsklassene i form av et sentralt og belysende lemma i vedkommende klasse.



Figur 8: Elektronisk ordbok på nettsiden *ordnanet.is*. En formbasert lemmanliste gir adgang til enkelte lemnaer og de begrepsfelt de tilhører. En særskilt liste med begrepsoverskrifter gir direkte tilgang til de enkelte feltene.

Brukeren kan fritt definere sitt søkeobjekt i søkefeltet ved hjelp av jokertegn og dermed søke på ulike elementer i lemmanaene. Med hensyn til standardiseringen av lemnaformene betyr det at strukturelle elementer kan trekkes fram, på lik linje med enkeltord og andre leksikalske enheter.

Figur 9 viser et enkelt eksempel der søkeordet har vært *sofa* 'sove'. Det gir en liste med bortimot 60 lemnaer der *sofa* står øverst. Ved å klikke på *sofa* åpnes det en artikkel som består av tre felt. Det øverste feltet viser antonymet *vaka* 'være våken', i midten er det et større felt med lemnaer som tilhører den samme begrepsklassen, og det nederste feltet rommer lenker til beslektede klasser. Midtfeltet inneholder 143 lemnaer som står i alfabetisk rekkefølge. Bare den første delen er synlig, men listen kan utvides ved å klikke på *meira*. I øverste høyre hjørne er det to andre lenker. Ved å klikke på *sýna setningargerð* ('vise konstruksjonsmønstre') får vi en liste med grammatisk sortering av lemnaene (jf. bildet til

høyre i figur 9). Ikonet ved siden av fører til en side med forklaringer på de grammatiske taggene.

Figur 9: Artikkelen *sofa* 'sove' på nettsiden www.ordanet.is.

7. Sluttord

Man kan sikkert se det slik at den ordboksmodellen vi har presentert her, ikke kan måle seg med en tradisjonell allmennspråklig ordbok, ikke minst fordi betydningsbeskrivelsen tilsynelatende mangler. På andre punkter går den langt ut over rammen for en allmennspråklig ordbok i tradisjonell forstand: med hensyn til avgrensning og omfang, den vekt som legges på fraser og flerordsenheter, og ikke minst den sentrale stilling de semantiske relasjonene får i modellen. Grunnen til at betydningsbeskrivelse og definisjoner ikke blir prioritert, er at ordet (i kraft av sine formbaserte egenskaper) ikke er den sentrale leksikografiske enheten. Men betydningsbeskrivelsen har sin plass i modellen, selv om fokuset er

blitt flyttet fra det semasiologiske til det onomasiologiske aspektet. I første omgang har den en identifiserende funksjon i forbindelse med entydiggjøring av lemmaene ved å skille mellom homonymer (lemmatiserte betydningsvarianter). Men den vil også komme inn på et senere stadium for å utdype og komplettere den informasjon som formidles gjennom andre elementer i ordboksbasen. Hovedsaken er at betydningsbeskrivelse og definisjoner ikke er obligatoriske, og at lemmaenes betydningsmessige karakter også gjenspeiles i de relasjoner de inngår i og – især for flerordslemmaenes vedkommende – i selve lemmaformene.

Vi har i denne artikkelen lagt vekt på at det onomasiologiske aspektet bør få en mer framtrædende plass i allmenn leksikografisk beskrivelse. Om noen får en ny oppfatning av hva begrepet *allmennspråklig ordbok* innebærer, spiller i og for seg en mindre rolle. Motivet er å finne fram til en hensiktsmessig og adekvat formidling og presentasjon av leksikografiske data uten å være bundet av eldre tradisjoner og et fastlåst system. Det innebærer bl.a. at ordboken kan kombinere egenskaper og funksjoner som tradisjonelt tilhører ulike typer. Den store utfordringen i den forbindelse er nettopp å få til et aktivt og balansert samspill mellom det semasiologiske og det onomasiologiske aspektet.

Litteratur

Ordbøker

NLO 1997 = Henning Bergenholtz/Ilse Cantell/Ruth Vatvedt Fjeld/Dag Gundersen/Jón Hilmar Jónsson/Bo Svensén: *Nordisk leksikografisk ordbok*. Oslo: Universitetsforlaget.

Orðastaður 2001 = Jón Hilmar Jónsson: *Orðastaður. Orðabók um íslenska málnotkun*. 2. utgave. Reykjavík: JPV-útgáfa. [1. utgave 1994.]

Orðaheimur 2002 = Jón Hilmar Jónsson: *Orðaheimur. Íslensk hugtakaorðabók með orða- og orðasambandaskrá*. Reykjavík: JPV-útgáfa.

Stóra orðabókin um íslenska málnotkun 2005 = Jón Hilmar Jónsson: *Stóra orðabókin um íslenska málnotkun*. Reykjavík: JPV-útgáfa. <http://snara.is>.

Annen litteratur

Fjeld, Ruth Vatvedt & Lars Vikør 2008: *Ord og ordbøker*. Kristiansand: Høyskoleforlaget.

Jónsson, Jón Hilmar 2009a: Lexical description. An onomasiological approach on the basis of phraseology. I: Sandro Nielsen & Sven Tarp (red.): *Lexicography in the 21st Century. In honour of Henning Bergenholtz*. Amsterdam / Philadelphia: Benjamins, 257–280.

Jónsson, Jón Hilmar 2009b: Lemmatisation of Multi-word Lexical Units: Motivation and Benefits. I: Henning Bergenholtz, Sandro Nielsen & Sven Tarp (red.): *Lexicography at a Crossroads*. Bern: Peter Lang, 165–194.

Nikulásdóttir, Anna Björk & Matthew Whelpton 2010: Lexicon Acquisition through Noun Clustering. I: *LexicoNordica 17*, 141–161.

Svensén, Bo 2004: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. Stockholm: Norstedts Akademiska Förlag.

Nettsteder

Fraseologisk register til *Orðabók Háskólans*:

www.lexis.hi.is/osamb/osamb.pl

Íslenskt orðanet: www.ordanet.is

Nettside med islandske ordbøker: <http://snara.is>

Jón Hilmar Jónsson
forskningsprofessor
Árni Magnússon-instituttet for
islandske studier
Íslands universitet
Neshagi 16
IS-107 Reykjavík
jhj@hi.is

Þórdís Úlfarsdóttir
orðboksredaktør
Árni Magnússon-instituttet for
islandske studier
Íslands universitet
Neshagi 16
IS-107 Reykjavík
disa@hi.is

FinnWordNet och kulturbundna ord

Nina Martola

In this paper I analyze the beta version of the Finnish WordNet from a societal and cultural point of view. FinnWordNet has been created through translation of the Princeton WordNet (Lindén & Carlson 2010). Certain lexicographical problems are to be expected due to the method.

Essential terms referring to Finnish society structure and Finnish culture cannot be found in the present FinnWordNet. The English definitions for words that can be found are not adequate from a Finnish viewpoint and the concept hierarchies are only partly relevant.

Constructing a WordNet is a laborious process and starting by translating is not a bad alternative. The linking to the English WordNet comes 'for free'. If the work had started by building up an independent WordNet for Finnish the linking process would have been very time-consuming. But pure word translation from L1 to L2 without any supplementing work does, of course, not result in an adequate description of the target language L2.

1. Inledning

Den första versionen av WordNet vid Princetons universitet kan sägas ha förelegat 1985, och under årens lopp har ordnätet bearbetats och utvecklats på många olika vis (Miller 1998). För närvarande (augusti 2011) är det version 3.1 som är aktuell och som ligger fritt tillgänglig på webben (<http://wordnetweb.princeton.edu>). Ett ordnät är i korthet enligt Miller (1995:39):

an online lexical database designed for use under program control. English nouns, verbs, adjectives, and adverbs are

organized into sets of synonyms, each representing a lexicalized concept. Semantic relations link the synonym sets

Den amerikanska WordNet-databasen har varit incitamentet till ett stort antal ordnät på andra språk. Ett sjukspråkigt ordnät, EuroWordNet, skapades i Europa under åren 1996–1999. Sedermera har WordNet-versioner kommit till för ett flertal språk. (<http://www.illc.uva.nl>)

Under slutet av 2009 och början av 2010 skapades ett ordnät för finska språket. Metoden man valde för detta var översättning av det amerikanska ordnätet (se närmare Lindén & Carlson 2010) och översättningsarbetet utfördes manuellt på mycket kort tid. En av tankarna med att skapa ett finskt ordnät genom översättning var att nätet då genast skulle fungera som en tvåspråkig källa. Betaversionen av FinnWordNet (i fortsättningen FWN) ligger i skrivande stund fritt tillgänglig på webben. I sina konklusioner över processen konstaterar Lindén & Carlson (2010:139) att behovet av att förbättra FWN kommer att utvärderas. De räknar med att åtminstone undersöka vilka (högfrekventa) finska ord som saknas och vilka betydelser som saknas för existerande ord. Vidare nämner de att en separat studie kommer att göras för att utvärdera om det finns delar av FWN som inte har adekvat ontologisk struktur för finska.

Denna undersökning är en helt fristående analys av vissa aspekter av den första versionen av det finska ordnätet, och den baserar sig i huvudsak på den version som var tillgänglig våren 2011. Ändringar och förbättringar som gjorts efter det är inte systematiskt beaktade. Jag gör analysen utifrån min bakgrund som tvåspråkslexikograf, inte med språkteknologiska ögon. Jag kommer att koncentrera mig på en typ av ord som mig veterligen inte har diskuterats särskilt mycket i WordNet-sammanhang, nämligen ord som är nära förknippade med det samhälle språket används i.

FWN ska alltså kunna fungera som en tvåspråkig ordbok, men

jag antar att ett lika viktigt (om inte ännu viktigare) syfte är att det finska ordnätet ska fungera som en beskrivning av det finska språket och det finska samhället. Här ser jag som tvåspråkslexikograf stora problem. Utan ett stort anpassningsarbete kan en vänd ordbok – vilket FWN i detta skede måste betraktas som – knappast tjäna som en adekvat beskrivning av det nya källspråket. Detta är trivialt i lexikografiska sammanhang, och det kan verka som att slå in vidöppna dörrar att ägna frågan en studie. Men FWN ska utvecklas vidare, och jag hoppas att jag kan bidra till arbetet genom att peka på konkreta problem som en lexikograf lätt hittar. Finska språket förtjänar ett gediget ordnät.

I det följande kommer jag först att se på definitionerna (i WordNet-termer *glosser*) för ett antal finska ord som på olika vis ansluter sig nära till det finska samhället. Definitionerna är inte översatta till finska utan föreligger på engelska. Det är bara själva orden, eller närmare bestämt ordgrupperna, *synseten*, som översatts till finska i det finska ordnätet.¹ Ekvivalensproblematik tar jag däremot inte systematiskt upp, eftersom jag är fel person att bedöma finsk-engelska motsvarigheter.

FWN är ju ett onomasiologiskt verk, och därmed ägnar jag en rätt stor del av studien åt att diskutera begreppshierarkierna, som i likhet med definitionerna är direkt övertagna från Princetons WordNet.

FWN kan laddas ner till den egna datorn. Det har jag inte gjort, utan jag har utnyttjat ordbasen via det sökgränssnitt som ligger fritt tillgängligt på webben. Ord nätet går att använda som en tvåspråkig ordbok. Om man klickar för *Show translations* visas de engelska ursprungsorden till varje översatt ord både i betydelseöversikterna och i begreppshierarkierna.

1 Noga taget borde det egentligen, benämningen *WordNet* ~ *ordnät* till trots, handla om begrepp som ska översättas (jfr Ruppel 2011), men översättarna har utfört uppdraget som ordöversättning.

2. Betydelsebeskrivningarnas relevans

Ett språk är nära förbundet med det samhälle det används i, och samhälls- och kulturtypiska ord brukar vålla tvåspråkslexikografer problem. I det följande kommer jag därför att undersöka hur begreppen ”president” och ”län” samt begrepp som hänför sig till skolsystem och folkslag är behandlade i FWN.

2.1. Begreppet ”presidentti” (’president’)

Betydelsebeskrivningen för ordet *presidentti* ’president’ i FWN visas i exempel 1.

1. Yhdysvaltain presidentti, presidentti, ylipäällikkö, ** -- (the person who holds the office of head of state of the United States government; ”the President likes to jog every morning”)
2. presidentti, ** -- (the chief executive of a republic)
3. presidentti, ** -- (the judge who presides over a supreme court)
4. Yhdysvaltain presidentti, presidentti, ylipäällikkö, ** -- (the office of the United States head of state; ”a President is elected every four years”) ***

Exempel 1.²

Betydelse 1 och 4 hänför sig alltså till USA:s president, vilket är bakvänt ur ett finskt användarperspektiv. Betydelse 2 är den allmänna betydelsen och kan i själva verket snarast ses som ett överbegrepp för begreppet ”president” i olika länder (Finland,

2 Svensk övers. ’Förenta staternas president, president, överbefälhavare’. Asteriskparen är länkar till sidor där användaren kan lägga till egna synonymförslag. Funktionen har kommit till efter januari 2011.

USA, Frankrike, Tyskland, Chile osv.). Att USA:s president ges prominens i det amerikanska ordnätet kan accepteras om hierarkisystemet är upplagt så, men i ett finskt ordnät borde i så fall den finländska presidenten på motsvarande sätt lyftas fram, inte den amerikanska.

2.2. Begreppet ”lääni” (’län’)

Ordet *lääni* ’län’ anges ha bara en betydelse: ”the domain controlled by a feudal lord”. Förklaringen hänför sig till ordet *fiefdom* som kommer upp om man ber om den engelska översättningen. Det är visserligen en adekvat översättning av ordet *lääni* men inte i dess moderna betydelse, och ska ett ordnät beskriva finska språket måste självfallet den moderna användningen beaktas. Ord som hör nära samman med *lääni* är t.ex. *lääninhallitus* ’länsstyrelse’ och *maaherra* ’landshövding’. De hittas inte i FWN. Men också länen är numera historia i Finland. År 2010 avskaffades län och länsstyrelser och i stället indelades landet i regioner förvaltade av regionförvaltningsverk, *aluehallintovirasto*. Ordet saknas i FWN. I ett helt uppdaterat finskt ordnät borde det alltså anges tre olika betydelser eller användningar för ordet *lääni*: 1) den feodala enheten, 2) en tidigare förvaltningsenhet i Finland och 3) en förvaltningsenhet i Sverige. Ordningsföljden kan diskuteras.

2.3. Begrepp inom skolsystemet

Det följande beskrivningsområde jag ska analysera är det finländska skolsystemet. Jag börjar med en kortfattad redogörelse för hur systemet ser ut i Finland och anger de finska benämningarna på de olika begreppen och markerar med [Ø] om orden inte hittas i FWN. I Finland har vi allmän läroplikt, *oppivelvollisuus* [Ø] (alltså inte skolplikt). De elever som genomför sin läroplikt går i grundskola (årskurserna 1–9) *peruskoulu*. I den finska skollagen

och skolförordningen avser *peruskoulu* numera bara de enskilda skolorna medan själva utbildningen för årskurserna 1–9 benämns *perusopetus* [Ø] 'grundläggande utbildning'. Grundskolan delas in i *alakoulu* (årskurs 1–6) och *yläkoulu* [Ø] (årskurs 7–9). I allmänspråket förekommer fortfarande i någon mån också de tidigare benämningarna *ala-aste* 'lågstadium' och *yläaste* 'högstadium'. Efter genomgången läroplikt går eleverna vidare till andra stadiets utbildning *toisen asteen koulutus* [Ø] och söker då antingen till gymnasium, *lukio*, eller till en yrkesläroanstalt *ammattioppilaitos* [Ø], i dagligt tal *ammattikoulu* 'yrkesskola'.

Definitionerna för de skolord som återfinns i FWN är inte adekvata ur finsk synvinkel eftersom de hänför sig till det amerikanska skolsystemet. För *peruskoulu* anges det att det avser *ala-aste*, *alakoulu*, *oppikoulu* och definitionen lyder "a school for young children; usually the first 6 or 8 grades" (definitionen gäller am. *grammar school*). Ordet *oppikoulu* 'läroverk' hänför sig till ett gammalt finländskt skolsystem som avvecklades på 70-talet. Ordet *yläaste* anges ha två ungefärliga betydelser: 1) "a secondary school usually including 7th and 8th grades" och 2) "a school for students intermediate between elementary school and college; usually grades 9 to 12". De finska förklaringarna (synonymerna) är *yläaste ja lukio*, *lyseo*, *lukio*, *yläaste* ('högstadium och gymnasium', 'lyceum', 'högstadium', vilket antagligen fungerar förhållande väl som närmaste motsvarigheter för att beskriva det amerikanska systemet men som blir ytterst rörigt i fråga om finska skolor.

I det moderna finska skolsystemet utgör gymnasier separata utbildningsenheter. I det gamla läroverkssystemet, dit också ordet *lyseo* 'lyceum' hänför sig, hade man mellanskola och gymnasium inom en och samma skola. Mellanskolan var fem- eller sexårig, och eleverna sökte till läroverk efter fyra eller fem år i folkskola. Ordet *lyseo* lever visserligen kvar i namnet på en del skolor, men som skolterm är beteckningen försvunnen utom med avseende på tiden fram till 70-talet.

Skolsystemen i olika länder avviker från varandra och ekvivalenterna kan aldrig bli mer än ungefärliga, men i ett finskt ordnät borde definitionerna beskriva det finska systemet, och system från olika tider borde hållas i sär.

Därmed aktualiseras också den principiella frågan om i vilken grad ett ordnät ska beakta även tidigare förhållanden. Det är en komplicerad fråga och det finns inget självklart svar, utan svaret är beroende av vad ordnätet är avsett att användas till.

2.4. Begreppen ”suomalainen” (’finne; finsk’) och ”karjalainen” (’karelare; karelsk’)

Den följande typen av ord jag ska studera är folkslagsbeteckningar. Ordet *suomalainen* anges som substantiv ha betydelsen ”a native or inhabitant of Finland” medan betydelseangivelse för adjektivet är ”of or relating to or characteristic of Finland or the people of Finland”. Det framgår inte att *suomalainen* också kan ha den snävare innebörden ’finskspråkig’, en betydelse som anges i den moderna, enspråkiga ordboken *Kielitoimiston sanakirja* (KT).

Ordet *karjalainen* ’karelare; karelsk’ har definitionen ”a member of the Finnish people living in Karelia in northwestern European Russia”. Här saknas uppgift om att ordet liksom *suomalainen* också kan vara adjektiv. Betydelseangivelsen är inte riktigt adekvat. Karelarna i Ryssland är inte ett finskt folk utan ett folk besläktat med finnarna. Men dessutom kan *karjalainen* syfta på invånarna i det finländska landskapet Karelen, som i sin tur hör till det finska folket. Karelen införlivades till största delen med Sovjetunionen vid andra världskriget och befolkningen evakuerades i stor utsträckning till Finland. Delar av landskapet hör fortfarande till Finland, och namnen *Pohjois-Karjala* (Norra Karelen) och *Etelä-Karjala* (Södra Karelen) finns kvar som benämningar på områden i Finland. I områdena talas så kallade *karjalaismurteet* ’karelska dialekter’. En adekvatare benämning är *sydostfinska dialekter*. (Torikka m.fl. 2004).

2.5. Betydelsebeskrivningen för allmännare ord i FWN

De begrepp jag sett på i det föregående har gett en rätt negativ bild av FWN, men min angreppspunkt – att undersöka begrepp som är typiska för det finska samhället – har förstås också varit orättvis. Mer allmänna uppslagsord kan förväntas ge ett bättre utslag, och i rättvisans namn har jag därför kastat en blick på betydelsebeskrivningarna för tio slumpmässigt valda ord på *a-*, *h-* och *p-* och jämfört FWN med två andra ordböcker, den enspråkiga finska ordboken Kielitoimiston sanakirja (KT) och den finsk-engelska ordboken Uusi suomi-englanti suursanakirja (USE).

KT och USE (som redigerats oberoende av varandra) visar tämligen stor samstämmighet i fråga om antalet angivna betydelser i de undersökta artiklarna, medan siffrorna för FWN överlag är högre. Vissa (kortare) artiklar stämmer förhållandevis väl överens. En del av betydelseangivelserna i FWN blir upprepningar, eftersom de härrör från olika engelska ord. Ordet *puhdas* 'ren' utgör motsvarighet till de engelska orden *pure*, *clean* och *clear*, och likartade betydelseangivelser återkommer därför flera gånger. En del av angivelserna ligger också långt borta från uppslagsorden betydelsemässigt. Det gäller bland annat nyansen *clear* för *puhdas* (klar luft är inte riktigt detsamma som ren luft). En av motsvarigheterna till verbet *ajaa* är *course* 'hunt with hounds'. Verbet *ajaa* kan visserligen användas med avseende på drivande hundar men inte om jägare som jagar hare eller räv med stövare. Verb som *haiskah-taa* 'lukta' (intransitivt) och *vaikuttaa* 'verka, förefalla' borde inte uppträda som synonymer till *havaita* 'observera' (transitivt).

Överlag verkar betydelsebeskrivningarna för mer allmänna ord mer adekvata i FWN än de för de kulturbundna orden. Men också de artiklarna skulle kräva en genomgång, eftersom FWN bär tydliga spår av att vara en vänd ordbok.

3. Begreppshierarkierna

FWN är ett onomasiologiskt verk och begreppshierarkierna utgör en viktig del av innehållet. Jag ska därför ägna de följande avsnitten åt dem. Det finska ordnätet har kommit till genom ren översättning. Ingen onomasiologisk bearbetning har tills vidare gjorts (våren 2011). Hierarkierna är övertagna från det amerikanska ordnätet med de styrkor och svagheter det uppvisar. I och med att betydelsebeskrivningarna för de genomgångna kulturella och samhällseliga orden inte till alla delar är adekvata i FWN kan man vänta sig att också hierarkierna är problematiska.

Det finns flera olika typer av begreppshierarkier i FWN. De typer jag kommer att diskutera är hypernymer, hyponymer och ko-hyponymer (systrar). Exemplet jag visar är hämtade direkt ur FWN, och hierarkierna går från det mest specifika högst upp till det mest generella längst ner. I de fall då inte hela kedjan varit relevant för diskussionen har jag kortat av den. För icke-finskkunniga läsare har jag låtit de engelska motsvarigheterna ur FWN ingå i exemplen.

3.1. Begreppshierarkidjup och särskiljningsnivå

Mycket påfallande när man slår upp kultur- och samhällsbundna ord i FWN är de långa hypernymkedjorna. Begreppshierarkierna är med andra ord djupa.

Ber man FWN om hypernymerna för begreppet ”presidenti” får man fyra hierarkier för de fyra olika betydelserna. För betydelse 1 (USA:s president) ser hierarkin ut som i exempel 2.

- Yhdysvaltain presidentti [President of the United States, United States President], presidentti [President], ylipäällikkö [Chief Executive]
 ⇒ valtionpäämies [head of state], valtionjohtaja [chief of state]
 ⇒ edustaja [representative]
 ⇒ välittäjä [negotiator], neuvottelija [negotiant], sovittelija [treator]
 ⇒ viestijä [communicator]
 ⇒ henkilö [person, individual, someone, somebody], ihminen [mortal, soul]
 ⇒ eliö [organism], olento [being]
 ⇒ elävä olio [living thing, animate thing]
 ⇒ kokonaisuus [whole, unit]
 ⇒ esine [object], fyysinen esine [physical object]
 ⇒ fyysinen entiteetti [physical entity]
 ⇒ entiteetti [entity], kokonaisuus [entity]
 ⇒ aiheuttaja [causal agent], syy [cause], kausaalinen agentti [causal agency]
 ⇒ fyysinen entiteetti [physical entity]
 ⇒ entiteetti [entity], kokonaisuus [entity]

Exempel 2.

Hierarkin för ”presidentti” påminner starkt om de taxonomiska hierarkierna för t.ex. växter och djur, men jag ställer mig frågande till de många nivåerna. Behövs exempelvis hela serien *viestijä–välittäjä–edustaja*, och på vilka grunder har de tre ordens inbördes hierarkiska position definierats?

Men hierarkin i exempel 2 är dessutom svårtolkad. Det förefaller som om det handlade om två alternativa hierarkier, dels en serie med tolv nivåer (sökordsnivån inberäknad), dels en serie med fyra nivåer. Å andra sidan ligger *aiheuttaja* och *eliö* på samma avstånd från vänster och det kan möjligen betyda att bara de tre översta hierarkinivåerna är alternativa. Något omnämnande om alternativa hierarkier har jag inte hittat på FWN:s instruktions-sida. Eftersom *entitet* och *fysisk entitet* återfinns i båda serierna ska trädet ändå knappast tolkas som ett enda hierarkiträd med femton nivåer.

För betydelse 4 ”presidentti” med avseende på presidentinstitutionen i USA ser hierarkin ut som i exempel 3.

- Yhdysvaltain presidentti [President of the United States], presidentti [President], ylipäällikkö [Chief Executive]
- ⇒ presidentintoimi [presidency], presidenttikausi [presidency], presidentinvirka [presidentship]
 - ⇒ asema [position], toimi [post, billet], paikka [berth, spot, place], virka [office], työpaikka [situation]
 - ⇒ elinkeino [occupation], liiketoiminta [business], työ [job], toimiala [line of work], ala [line], kiinnostuksen kohde [line]
 - ⇒ aktiviteetti [activity], toiminta [activity]
 - ⇒ teko [act, deed], työ [act], toimenpide [act], inhimillinen toiminta [human action, human activity]
 - ⇒ tapahtuma [event]
 - ⇒ psykologinen ominaisuus [psychological feature]
 - ⇒ abstrahointi [abstraction], käsitteellistäminen [abstraction], abstrakti entiteetti [abstract entity]
 - ⇒ entiteetti [entity], kokonaisuus [entity]

Exempel 3.

I detta exempel är det påfallande att både statiska (”psykologisk egenskap”, ”position”) och dynamiska begrepp (”händelse”, ”aktion”) ingår i en och samma hierarki som överbegrepp för varandra.

Med avseende på USA:s president får ordet *presidentti* alltså två olika hierarkier, en för den personsyftande användningen och en för den institutionssyftande. Men om man väljer att skilja på person och institution borde samma tudelning i konsekvensens namn göras också för betydelserna ’överhuvud i republik’ och ’ledare för högre rättsinstans’, vilket alltså inte är fallet.

Strängt taget kan skillnaden mellan tjänstposition och innehavare anses gälla de flesta arbeten, inte bara högre tjänstemän. Om en arbetsgivare i en platsannons söker en lärare, en sekreterare, en städare, en lastbilschaufför, en rörmokare eller liknande så kan benämningarna i annonsen knappast anses syfta på en regel-

rätt person utan snarast på uppgiften personen behövs till.

Frågan är om det inte vore ändamålsenligare att lägga in tudelningen person–position på en högre nivå i begreppshierarkin. Ska skillnaden kodas in borde det i varje fall göras konsekvent för alla ord som den kan aktualiseras för.

I det danska ordnätet DanNet (<http://andreord.dk>) görs ingen skillnad person–institution. Ordet *præsident* anges i betydelseöversikten ha tre betydelser ('domare', 'ledare' och 'statsöverhuvud') och för betydelsen 'statsöverhuvud' är de överordnade begreppen sex till antalet (exempel 4).

statsoverhoved → leder → person → væsen → dyr → organisme

Exempel 4.

Man kan förstås ställa frågan om *president* i satser av typen ”i Finland väljs det president vart sjätte år” passar in i begreppsserien ”person”–”organism”, men å andra sidan är innehavande av olika positioner och tjänsteställningar i samhället så intimt sammankopplat med begreppet ”person” att man kan fråga sig om inte ett ’potentiell samhällsfunktion’ utgör en viktig betydelsekomponent i ordet.

De FWN-egenskaper som här diskuterats är ärvda från den amerikanska förlagan och det finns kanske goda argument för att begreppsserierna ser ut som de gör. För en utomstående betraktare verkar de dock inte särskilt stringenta. Med tanke på en fortsatt bearbetning av FWN borde en diskussion föras om de direkt övertagna hierarkierna är optimala eller om justeringar kunde genomföras. Problemet är förstås att kopplingen till det amerikanska ordnätet försvagas om ändringar görs. Å andra sidan antar jag att kopplingen kan försvagas också vid revideringar av Princeton WordNet.

3.2. Perspektiv och precisionsnivå

Det som är att vänta när ett ordnät skapat för ett språk översätts till ett annat språk är att perspektivet blir fel ur det nya språkets synvinkel.

Avsaknad av finsk infallsvinkel i FWN är påfallande om man ser på hierarkin för begreppet ”karjalainen” ’karelare’, som ju är en folkslagsbeteckning nära förknippad med den finska kulturfären. Hypernymserien visas i exempel 5. Endast de nedersta nivåerna är medtagna.

Sense 1

karjalainen [Karelian, Carelian]

⇒ venäläinen [Russian]

⇒ alkuasukas [native, indigen, aborigine, aboriginal], maakalainen
[indigene]

⇒ henkilö [person, individual, someone, somebody], ihminen
[mortal, soul]

[...]

Exempel 5.

En karelare betecknas alltså rätt och slätt som ryss, vilket i och för sig är korrekt i fråga om betydelsen ’medborgare i Ryssland’. Men i ett finskt ordnät saknar man en alternativ hierarki för karelarnas tillhörighet till de finsk-ugriska folken. Karelarna och andra finsk-ugriska grupper innehar en särställning i finskt perspektiv jämfört med andra folk i Ryssland.

Anmärkningsvärt i hierarkin i exempel 5 är också att ryssar definieras som ”urinvånare”, medan hypernymkedjan för t.ex. finnar, danskar och tyskar är ”europé–invånare–person” och för t.ex. turkar, indier och israeler ”asiat–invånare–person”. Bakgrunden är möjligen att inte alla ryssar är européer och att ”ryss” å ena sidan borde ingå i två hypernymkedjor, å andra sidan är övergripande för ”europé” och ”asiat”.

Systemangivelserna för ”karjalainen” utgår förstås också från ”ryss” som överordnat begrepp, och i raden av ko-hyponymer återfinns därmed beteckningar för olika folkslag i Ryssland, både finsk-ugriska och andra (exempel 6). Att ”karelare” också är systerbegrepp till ”finne” i egenskap av beteckning på finsk-ugriska folk framgår ingenstans.

- venäläinen [Russian]
 - ⇒ tataari [Tatar]
 - ⇒ udmurtti [Udmurt], votjakki [Votyak]
 - ⇒ komi [Komi]
 - ⇒ tseremissi [Cheremis, Cheremiss], Mari [Mari]
 - ⇒ inkeriläinen [Ingrian, Inger, Ingerman]
 - ⇒ karjalainen [Karelian, Carelian]
 - ⇒ osjakki [Ostyak], hanti [Khanty]
 - ⇒ mordvalainen [Mordva, Mordvin, Mordvinian]
 - ⇒ nganasani [Nganasan]
 - ⇒ selkupi [Selkup], ostjakkisamojedi [Ostyak-Samoyed]
 - ⇒ samojedi [Samoyed]
 - ⇒ vepsäläinen [Veps, Vepse, Vepsian]
 - ⇒ voguli [Vogul], mansi [Mansi]
 - ⇒ jeniseiläinen [Yeniseian]
 - ⇒ isovenäläinen [Great Russian]
 - ⇒ moskovalainen [Muscovite]
 - ⇒ georgialainen [Georgian]
 - ⇒ siperialainen [Siberian]
 - ⇒ tsetseeni [Chechen]

Exempel 6.

Systrarna till ”suomalainen” ’finne’ är över sextio till antalet och de representerar en brokig skara med ”eurooppalainen” ’europé’ som överbegrepp. En stor, entydig grupp systrar är benämningar på invånare i nu existerande stater som t.ex. *tanskalainen* ’dansk’, *italialainen* ’italienare’ och *valkovenäläinen* ’vitryss’. I tillägg till dessa finns det invånarbetesbeteckningar på nyligen upplösta stater

som *jugoslavialainen* 'jugoslav' och *tsekkoslovakialainen* 'tjeckoslovak'. Däremot är det tveksamt om angler, saxare, franker och kelter för att inte tala om teutoner och langobarder hör hemma under begreppet "europé". Jämförbara med beteckningar för moderna statsmedborgare är de i varje fall inte.

I ko-hyponymlistan för "suomalainen" ingår vidare beteckningar på invånare i områden som inte utgör stater t.ex. *kreetalainen* 'invånare på Kreta', *balkanilainen* 'balkanbo', *saamelainen* 'same', *pohjoismaalainen* 'nordbo' och *tirolilainen* 'tyrolare'. De avser alla i och för sig också moderna européer, men ska den typen av beteckningar anses som systrar till *suomalainen* borde också ett mycket stort antal andra invånarbeteckningar komma med.

Den märkligaste ko-hyponymen till "suomalainen" är beteckningen *sahib*. Visserligen har benämningen förekommit med avseende på européer i Indien, men ordet fungerade som tilltalsord med innebörden 'herre' (SO *sahib*) och inte som en folkslagsbeteckning.

Avsaknaden av underbegrepp för "suomalainen" 'finne; finländare' är också flagrant med tanke på att ett betydande antal amerikanska invånarbeteckningar återfinns i FWN. För det första har vi beteckningar för invånare i de olika landskapen, exempelvis *hämäläinen* 'tavastlänning', *savolainen* 'savolaxare', *pohjalainen* 'österbottning' med flera. En lite annan typ representerar *suomenruotsalainen* 'finlandssvensk'. Hyponymer är också alla inbyggarbeteckningar till stadsnamn: *helsinkiläinen* 'helsingforsare', *tamperealainen* 'tammerforsare', *vaasalainen* 'vasabo'. I och för sig är det knappast påkallat att ta med invånarbeteckningar för alla städer i Finland i ett finskt ordnät, men ser man till frekvens i moderna finska texter är *helsinkiläinen* och *tamperealainen* vida vanligare än exempelvis *permiläinen* 'permjak' som finns med i FWN.

Men också i detta fall finns det en historisk dimension. Benämningen *suomalaiset* avser tillika en av de stammar som har befolkat Finland. I det avseendet är alltså "finne" ett systerbegrepp till t.ex.

”tavast” och ”karelare”, inte hypernym som i fråga om moderna förhållanden. I ett finskt ordnät borde med andra ord en historisk betydelse med tillhörande hierarkier anges separat.

Det sista exemplet på perspektivproblem som jag ska ta upp är ko-hyponymerna för ”presidentti” i betydelsen ’statsöverhuvud’ (exempel 7)

- valtionpäämies [head of state], valtionjohtaja [chief of state]
 - ⇒ pääministeri [chancellor, premier, prime minister]
 - ⇒ pääministeri [Prime Minister, premier], PM [PM]
 - ⇒ presidentti [president]
 - ⇒ Yhdysvaltain presidentti [President of the United States, United States President], presidentti [President], ylipäällikkö [Chief Executive]
 - ⇒ hallitsija [sovereign], kruunupää [crowned head], monarkki [monarch]

Exempel 7.

Ett land har normalt ett och endast ett statsöverhuvud (*valtion päämies, valtionjohtaja*), antingen en president eller en monark beroende på statsskick. Regeringschefen (statsminister, premiärminister, förbundskansler) är inte statshuvud (utom möjligen ställföreträdande) och ”pääministeri” är inte en ko-hyponym till ”presidentti”. I USA är presidenten tillika regeringschef, och premiärminister saknas. I den meningen kan ”president” med syftning på USA:s president alltså faktiskt betraktas som ko-hyponym både till ”monark” och till ”stats- och premiärminister”. Men i exempel 7 handlar det inte explicit om USA:s president.

De anförda exemplen visar inte enbart på ett avlägset perspektiv på finsk kultur och det finländska samhället utan de aktualiserar också ett problem som måste vara generellt för ordnät, nämligen hur noga man ska definiera begrepp. Att exempelvis definiera langobarderna som européer utan någon vidare kategorisering blir ytterst oprecist. Går man in för underkategorisering kan det

dock ha betydande återverkningar på ett stort antal ord. Samtidigt inställer sig frågan vad ett ordnät ska beskriva: ska det i första hand beskriva ett visst språk i ett visst samhälle eller ska det vara en allmän beskrivning som täcker så mycket som möjligt av ”världen” utan att gå in i detaljer.

3.3. Översättningsrelaterade problem

Som jag nämnde i inledningen hör engelsk-finsk översättning inte till mitt kompetensområde, så kvaliteten på själva översättningsarbetet undandrar sig min bedömning. Vissa missar har jag i alla fall stött på i FWN under undersökningens gång, och felaktigheterna kan ha återverkningar på begreppshierarkierna. I exempel 8 är *Baltic-Finnic* översatt med *itämerensuomalainen* och *Fennic, Finnic* med *itämerensuomalainen kieliryhmä* (m.fl.).

Sense 2

suomi [Finnish, Suomi]

⇒ *itämerensuomalainen* [Baltic-Finnic]

⇒ *itämerensuomalainen kieliryhmä* [Fennic, Finnic], *itämerensuomalaiset kielet* [Fennic, Finnic], *ei-ugrilainen kieliryhmä* [Non-Ugric], *ei-ugrilaiset kielet* [Non-Ugric]

⇒ *suomalais-ugrilainen kieli* [Finno-Ugric], *suomalais-ugrilainen kielikunta* [Finno-Ugric]

⇒ *uralinen kielikunta* [Uralic], *uralinen kieli* [Uralic language]

⇒ *ural-altalaiset kielet* [Ural-Altic]

⇒ *kieli (yli)* [natural language], *kieli* [tongue]

[...]

Exempel 8.

Ordet *itämerensuomalainen* kan inte ensamt syfta på språk, utan ordet är adjektiv och ingår som bestämning i språkgruppsbenämningen *itämerensuomalaiset kielet* ’översjöfinska språk’, vilket är den finska motsvarigheten till *Baltic-Finnic*. I plural kan adjektivet

användas substantiviskt, men då avser det folkgrupp, inte språk. *Itämerensuomalaiset kielet* är synonymt med *itämerensuomalainen kieliryhmä* 'den östersjöfinska språkgruppen'. Samma begrepp återkommer alltså på två olika nivåer i exemplet. De engelska benämningarna *Fennic*, *Finnic* hänför sig till den finska grenen av de finsk-ugriska språken, där östersjöfinska språk utgör en av flera grupper.

Också i detta sammanhang kan en inkonsekvens nämnas: i begreppshierarkin för "saame" 'samiska' står överbegreppet "saamelaiskieli" 'samiskt språk' direkt under "uraliska språk". Nivån finsk-ugriska saknas alltså.

3.4. Sammanfattande om begreppshierarkierna

Genomgången av begreppshierarkierna har visat på flera problem med FWN. En del har direkt att göra med hur det finska ordnätet har kommit till. Synseten i det finska ordnätet har skapats genom direkt översättning av de amerikanska synseten utan lexikografisk eller ontologisk bearbetning, och FWN har därmed övertagit de problem som finns i förlagan. En del direkta fel har kommit till på grund av felöversättningar, men den stora svagheten med FWN – i en lexikografs ögon – är att beskrivningen av finska förhållanden framstår som mycket distanserad och rentav oinitierad. Möjligen är ordnäten tänkta att vara så allmänna att de är inbördes kompatibla och att de alltså inte ens ska röra sig på språkspecifik nivå. Men i så fall borde FWN gallra bort många säramerikanska drag.

Det som trätt fram under analysens gång som en viss överraskning är inkonsekvenser i behandlingen av olika begrepp av samma typ. Jag hade förväntat mig att en datorbaserad systematik skulle vara mycket konsekventare. Man får intrycket av att materialet som det amerikanska ordnätet bygger på varit ojämnt. En begreppsordbok förutsätter synnerligen hög grad av systematik i beskrivningarna och en allsidig täckning av hela språkssystemet eller åtminstone av de områden man väljer att ha med.

Antalet hierarkier för ett begrepp är beroende av antalet betydelser man anför. Därför borde betydelsenanserna bedömas mycket noga, och man måste också ta principiell ställning till på vilken nivå man ska lägga beskrivningen för olika typer av begrepp.

4. Utvecklingsarbete

I Linden & Carlson (2010) påpekas det att den första versionen av FWN ska utvecklas och bearbetas. Ett led i detta är s.k. *crowd-sourcing* (jfr Kann 2010:110) där användarna ombeds utvärdera synonymer. I januari 2011 var utvärderingen frivillig och de ord man sökte i FWN kom upp för utvärdering. Användaren ombads också föreslå egna synonymer.

Ur språklig synvinkel är det problematiskt att det är finska ord som ska utvärderas medan den kontext som ges är engelskspråkig. Visserligen kan det försvaras med att utvärderingen principiellt gäller begrepp, inte ord, och att bedömningen därför är språkoberoende. Frågan är dock om alla de användare som råkar komma in på FWN på webben har detta klart för sig. I vissa fall kan kontexten dessutom vara svår att förstå för den som inte har mycket goda kunskaper i engelska. Jämför också Ruppels (2011) ifrågasättande av språkoberoende begrepp.

Någon gång våren 2011 gjordes synonymbedömningen mer obligatorisk, så att man inte kommer in på sökfunktionen i FWN förrän man tagit ställning till ett slumpmässigt synonympar (man kan svara *I'm not sure*). När man sänt sin utvärdering kommer dialogrutan för sökning fram och man får ett nytt ordpar vars synonymigrad man kan bedöma om man vill. Några egna synonymförslags bes man inte längre om³.

I och med att man ska värdera synonymin hos slumpmässigt

3 Den funktionen har integrerats i själva artiklarna: via en länk (dubbelasterisk) kan man lägga till synonymer. Se avsnitt 2.

givna par blir bedömningen ibland svår. Orden man ställs inför kan representera mycket främmande områden. En principiell fråga är också hur väl crowdsourcing lämpar sig som metod för bedömning av termer och termartade ord där graden av synonymi inte kan avgöras via en gallupundersökning utan där det handlar om precisa benämningar som antingen är synonymer eller inte är det. Men regelrätta fel kanske upptäckts genom metoden.

För att bedöma ord som är mer vaga till betydelseinnehållet fungerar metoden uppenbarligen rätt väl (åtminstone enligt t.ex. Kann 2010), och många sådana synonympar kommer naturligtvis också upp i FWN.

Att be användarna utvärdera synonymer är ett led i förbättringsarbetet. Men jämsides med det borde systematiska genomgångar av olika slag utföras av en redaktion.

5. Avslutning

WordNet och dess avläggare på olika språk är främst skapade för språkteknologiska ändamål. Om DanNet sägs det att det ska vara en resurs för IT-system som arbetar med intelligent informationshantering (<http://www.wordnet.dk>). För det amerikanska WordNet anger man en mängd olika tillämpningar (Miller 1998, <http://wordnet.princeton.edu/>), men Miller (1998:xxi) nämner explicit också mänskliga användare. En version som tillhandahålls i ett webbgränssnitt bör väl vara tänkt för mänskliga webbanvändare.

Vilka egenskaper som krävs av resurser för intelligent informationshantering är jag inte klar över, men jag kan inte tro annat än att samma krav bör ställas på dem som på resurser där mänskliga språkanterare hämtar sin information: de ska vara adekvata beskrivningar av språket i fråga. Förmodligen måste kraven ställas ännu högre, för en dator bearbetar ju i ett nafs sådana mängder av material som en människa inte ens kan drömma om att gå igenom

under en hel livstid. Alla eventuella fel återverkar oändligt mycket mer. Och en dator kan inte tänka själv.

En gedigen och adekvat resurs för finska språket är FWN tills vidare inte. Som jag ser det fungerar ordnätet för närvarande knappast tillfredsställande för något annat än som källa för översättning från (amerikansk) engelska till finska. Men inte ens vid översättning vill jag slå upp i ett finskt ordnät och få begreppet ”grundskola” beskrivet för mig i amerikanska termer, och jag känner mig lite kränkt som finländare om min finska källa anger att den första betydelsen för ordet *presidentti* är USA:s president.

Problemet med att ord som är specifika för ett visst samhälle kan vålla översättningsproblem tas upp av Lindén & Carlson själva (2010:133f.). De avfärdar det dock med att de udda orden utgör bara en liten del och att 95 % av alla ord låter sig översättas. Det motiverar enligt deras åsikt översättning som metod att skapa ett ordnät för ett nytt språk. Det är till detta jag som tvåspråkslexikograf har mina betänkligheter. Jag kan i och för sig gärna tro att 95 % av orden låter sig översättas men jag ställer mig frågan vad översättningsresultatet ska användas till om produkten inte förädlas avsevärt.

Min undersökning har visat att en gammal lexikografisk sanning verkar hålla streck också för ordnät: man gör inte en ordbok från språk 1 till språk 2 genom att vända på en ordbok från språk 2 till språk 1. Det ska mycket mer redigeringsarbete till.

Därmed inte sagt att översättning inte kan vara det första steget på vägen. Som Lindén & Carlson (2010:123) slår fast: att skapa ett finskt ordnät från början hade varit att upprepa årtionden av lexikografiskt arbete som utförts vid Princetons universitet. Det är helt sant. Men varje källa har sina svagheter och översättningsarbetet borde utföras parallellt med eller åtminstone följas av en kritisk analys och framför allt av målgruppsanpassning.

Jag hoppas det ska finnas resurser att arbeta vidare med FWN så att det småningom kan bli det gedigna ordnät för finska som

det är tänkt att vara. Att bygga upp ett ordnät för ett språk är en intrikat och mödosam process oavsett i vilken ända man börjar (jfr Pedersen 2010 om DanNet). Ett ordnät som kommit till genom översättning är en god råvara för förädling. Och förädlingen av FWN kunde förslagsvis ske med hjälp av resultaten av många års mödosamt *finskt* lexikografiskt arbete. På så vis kunde finsk lexikografi och finsk språkteknologi berika varandra.

Litteratur

Ordböcker och uppslagsverk

KT = *Kielitoimiston sanakirja* (Elektronisk version) ('Språkbyråns ordbok'). Eija-Riitta Grönros (huvudredaktör), Minna Haapanen (redaktionssekreterare) m.fl. Helsinki: Kotimaisten kielten tutkimuskeskus ja Kielikone Oy, 2008.

SO = *Svensk ordbok utgiven av Svenska Akademien*. Utarbetad vid redaktionen för Svenska Akademiens samtidsordböcker, Lexikaliska institutet, Institutionen för svenska språket, Göteborgs universitet. Svenska Akademien & Norstedts 2009.

USE = Raija Hurme, Riitta-Leena Malin & Olli Syväoja 1984/1989: *Uusi suomi-englanti suursanakirja* ('Ny finsk-engelsk storordbok'). Helsinki: WSOY.

Annan litteratur

Kann, Viggo 2010: KTHs morfologiska och lexikografiska verktyg och resurser. I: *LexicoNordica* 17, 99–118.

Lindén, Krister & Carlson, Lauri 2010: FinnWordNet – WordNet på finska via översättning. I: *LexicoNordica* 17, 119–140.

Miller, George A. (utg.): 1998: *WordNet. An electronic database*. Cambridge (Mass.)/London: The MIT Press.

- Pedersen, Bolette S. 2010: Semantiske sproressourcer – mellem sprogteknologi og leksikografi. I: *LexicoNordica* 17, 163–180.
- Ruppel, Klaas 2011: Ontologi, begreppssystem och WordNet. I: *LexicoNordica* 18 (i denna volym).
- Torikka, Marja, Raija Miiikkulainen & Anneli Hänninen 2004: *Karjala*. Kotimaisten kielten tutkimuskeskuksen julkaisuja 129. Helsinki. <http://kaino.kotus.fi/www/verkkojulkaisut/julk129/> (mars 2011)

Internethänvisningar

- DanNet*. <http://wordnet.dk/dannet/>, sökningar: <http://andreord.dk/> (mars 2011)
- EuroWordNet*. <http://www.illc.uva.nl/EuroWordNet/> (mars 2011)
- FWN = *FinnWordNet*. <http://www.ling.helsinki.fi> (dec. 2010–aug. 2011)
- Miller, George A. 1995: WordNet: A Lexical Database for English. I: *Communications of the ACM* Vol. 38, 39–41. citeseerx.ist.psu.edu/viewdoc/download (mars 2011)
- WordNet*. A lexical database for English. <http://wordnet.princeton.edu/> (mars 2011)

Nina Martola
 forskare, fil. dr
 Forskningscentralen för de inhemska språken
 Berggatan 24
 FI-00100 Helsingfors
 nina.martola@focis.fi

Semantiske relationer i en ny dansk begrebsordbog: genbrug på tværs af ordbøger

Sanni Nimb

This article discusses the synergies between thesauri, wordnets and semasiological dictionaries, describing how the lexical data from a wordnet (DanNet), as well as the data from a monolingual dictionary (The Danish Dictionary), are being used to establish a Danish thesaurus. In opposition to other thesauri, this one's main organizing principle is the grouping of words by semantic criteria. It thus combines words from different parts of speech in different types of groups. Furthermore, a number of formalized data on the words' semantic relations leave us with a large number of possibilities when it comes to using thesaurus information in the online presentation of The Danish Dictionary. It also means that the thesaurus can be used to extend the number of semantic relations in DanNet.

1. Indledning

En ny dansk begrebsordbog er under udarbejdelse ved Det Danske Sprog- og Litteraturselskab (DSL, 2010-13). Projektet er finansieret af Carlsbergfondet og skal munde ud i en trykt ordbog, men ordbogens data vil også blive tilgængelige på nettet. I selve den ontologiske inddeling af betydningsområder i begrebsordbogen tages der udgangspunkt i 906 kapiteloverskrifter oversat fra *Der deutsche Wortschatz nach Sachgruppen* (Dornseiff 2004). Trap-Jensen (2011) diskuterer begrebsordbøger generelt samt de overvejelser DSL har gjort sig mht. det metodiske grundlag for en ny dansk begrebsordbog. Den diskuterer også behovet for begrebsordbogen i forhold til online-udgaven af Den Danske Ordbog

(DDO). Ordbogen benytter i dag DanNet-data til visninger af beslægtede ord, hvilket ikke i alle tilfælde fungerer optimalt.

Indeværende artikel beskriver mere detaljeret hvordan begrebsordbogen i første omgang udnytter data fra det danske wordnet DanNet og fra DDO og efterfølgende beriger disse data med yderligere formaliseret semantisk information der kan genanvendes i de selvsamme ordbøger. Begrebsordbogens mikrostruktur bliver beskrevet i detaljer, bl.a. hvad angår de semantiske metadata der tildeles de forskellige ordgrupperinger.

2. DDO, DanNet og begrebsordbogen – og synergien imellem dem

Begrebsordbogen bygger som nævnt direkte på data i DDO og DanNet, som vi derfor beskriver indledningsvis. DSL udarbejdede i perioden 1992 til 2005 DDO og arbejder fortsat på at udvide ordforråd og betydningsantal i ordbogen, der i dag indeholder ca. 91.500 lemmaer og 116.000 betydningsbeskrivelser. Ordbogens betydningsinddelinger er grundstammen i det danske wordnet DanNet, som DSL udvikler sammen med Københavns Universitet (Pedersen 2010; Lorentzen & Nimb 2010; Pedersen et al. 2009).

DanNet indeholder pr. 1. juli 2011 65.000 af DDO's betydninger, struktureret i over-underbegrebs-hierarkier via relationen 'has_hyperonym' samt forbundet med ca. 75.000 andre indbyrdes semantiske relationer, bl.a. via nedarvning – og endnu flere hvis relationerne beregnes som transitive, jf. hjemmesiden *andreord.dk* der visualiserer wordnettet (Johannsen & Pedersen 2011) – her angives tallet for relationer til ca. 321.000. Synonyme betydninger er i DanNet knyttet sammen i det som er et wordnets grundenhed, et synset (fx {ben, bentøj}, pusselanke, skank, stang, underekstremitet}; {} bruges i denne artikel til at markere begreber, modsat udtryk eller lemmaer). Hvert synset er tildelt en ontologisk type

(fx for {kvinde} [Human], for {kasse} [Container] og for {ben, bentøj, pusselanke...} [Bodypart]).

DanNet udmærker sig i forhold til de fleste andre wordnets ved at tage udgangspunkt i betydningsinddelinger for sproget selv – i dette tilfælde dansk – frem for at anvende oversatte betydninger og dermed også automatisk arvede relationer fra det engelske Princeton WordNet (Fellbaum 1998), en metode som fx det finske wordnet har anvendt (Lindén & Carlson 2010). Desuden er DanNet suppleret med nogle i vores øjne meget vigtige semantiske relationer for menneskeskabte ting der ikke anvendes i andre wordnets, inspireret af SIMPLE-projektet (Lenci et al. 2000) og Pustejovskys qualia-struktur (Pustejovsky 1995). Nær-synonymi er også beskrevet i DanNet, både mellem ord fra samme ordklasse ('near_synonym') og på tværs af ordklasser. Derudover er der en lang række relationer der beskriver hvem eller hvad der er involveret i en handling (fx for {hamre}: involved_agent {person}), ligesom der er en række relationer der bruges til at beskrive delhelhed. Også en mere bred relation 'concerns' er tilknyttet en lang række begreber.

DanNet indeholder endnu ikke tilbunds gående beskrivelser af samtlige begreber, og ligesom vi ønsker at udnytte begrebsordbogens data i DDO (jf. Trap-Jensen 2011), ser vi også en stor mulighed i på sigt at kunne udnytte dem til en udvidelse af DanNet. Genbrug af data på tværs af DSL's ordbøger er mulig fordi id-numre fra hver af DDO's betydninger (fx *ben*, bet. 1, ID 21006106) går igen i DanNet, idet de er knyttet til hvert af de udtryk der indgår i et synset (ID 21006106 er tilknyttet udtrykket *ben* og indgår i synset 2109 sammen med 6 andre lemmaers id-numre (lemmaerne {ben, bentøj, skank, stang, underekstremitet}). *Bens* andre betydninger (knogle; stoleben m.fl.) indgår tilsvarende i andre synsets. I opbygningen af begrebsordbogen anvendes de samme id-numre. *Ben* indgår her i kapitel 2.24 *Legemsdele*. De fælles id-numre sikrer at den berigelse med semantisk information der nedfældes i be-

grebsordbogen, vil kunne udnyttes til at supplere både DanNet og DDO med oplysninger. Men inden vi vender tilbage til dette, vil vi se på hvordan DanNet og DDO i første omgang udnyttes i selve arbejdet med at opbygge begrebsordbogen.

3. Data fra DanNet og DDO til begrebsordbogen

Den grundlæggende over-underbegrebs-struktur i et wordnet som DanNet har et klart fællestræk med begrebsordbøgers traditionelle måde at gruppere begreber på. Fx indeholder både Harry Andersen (1945) og Dornseiff (2004) mange ordgrupper hvis fællestræk udgøres af overbegrebet (se figur 1, der viser borde og skuffemøbler; andre eksempler er tøj, skydevåben, værktøj og beholdere).

Bord – skuffemøbel

Substantiver: anretterbord, boghylde, bogreol, bogskab, bord, buffet, chatol, chaffioniere, dagligstuebord, disk, dragkiste, etagere, hjørneskab, hylde, kateder, kiste, klapbord, klædeskab, kommode, konsol, pengeskab, pult, reol, rygebord, sekretær, servante, skab, skrivebord, skrivepult, skuffemøbel, skænk, spisestuebord, toilet møbel.

Figur 1: Borde og skuffemøbler i Harry Andersen (1945). Overbegreber er fællestræk for ordene i gruppen.

Det er derfor meget oplagt at udtrække grupper af underbegreber fra DanNet til brug for en dansk begrebsordbog. Et udtrækningsprogram udviklet til formålet¹ gør redaktøren i stand til at vælge et udtryks underbegreber i DanNet og indsætte dem direkte i begrebsordbogen. Figur 2 illustrerer hvordan man ud fra DanNet, og i øvrigt i dette tilfælde uden særlig meget efterredigering, kan lave et lignende, moderne afsnit *bord – skuffemøbel*. Det skal nævnes at figurerne i denne artikel er taget fra redigeringsværktøjet og derfor

¹ Udarbejdet af Nicolai H. Sørensen, DSL.

ikke illustrerer hvordan den færdige begrebsordbog kommer til at se ud. Fx kan metadata ses på nogle af figureerne.

◉ **bord** ▶ træbord, fyrretræsbord, mahognibord, teaktræsbord, plankebord-4; ▶ kakkeltbord, glasbord, marmorbord-4; ▶ klappbord, udtræksbord-4; sofabord, spisebord, langbord, spisestuebord, køkkenbord; ▶ aftensbord, morgenbord, middagsbord, frokostbord, kaffebord, tebord, fødselsdagsbord, julebord, taffel-4; ▶ campingbord, klappbord-4; ▶ gavebord, fødselsdagsbord-4; ▶ bakkeltbord, indskudsbord, anretterbord-4; sidebord, hjørnebord, rygebord; ▶ rullebord, trådvogn-; telefonbord, sybord
 ◉ **arbejdsbord**, bænk; ▶ filebænk, høvlebænk-4; slagtebænk, vaskebænk; ▶ omstillingsbord, telefonbord-4; operationsbord, stinkskab, linnedskab, medicinskab; ▶ det grønne bord, eksamensbord-4; kateder; ▶ **syn**: skranke, retsskranke-4; vidneskranke, tallerstol
 ◉ **opbevaringsmøbel** ▶ **kommode**, skuffemøbel, skuffedarium-4; chiffonniere, rokokokommode; ▶ **skænk**, købmændsdis, dækketøjsskab, sølvtøjsskab, glasskab, vitrine, buffet-4; ▶ **chatol**, sekretær-4; ▶ **kiste**, skåbskiste, dragkiste, kistebænk, brudekiste-4; ▶ kommodeskuffe, skuffe-4;
 ◉ **skab**, jalousiskab, glasskab, vægskab; ▶ **syn**: garderobeskab, klædeskab-4; garderobe, karlekammerskab, hjørneskab, hjørneskab; ▶ svaleskab, isskab, varmeskab-4; ▶ kosteskab, værktøjsskab-4; skabsэлемент, skabsvæg
 ◉ **reol**, bogreol; ▶ byggereol, stigereol, stålereol-4; ▶ rack, chassis, kabinet, tv-tårn-4; hjemmebar; ▶ barskab, bar, hjemmebar; vinreol; ▶ hylde, boghylde-4; bogkasse; ▶ amagerhylde, kaminhylde, konsol, etager-4; ▶ glashylde, hattehylde-4;

Figur 2: Borde og skuffemøbler (opbevaringsmøbler) i den nye danske begrebsordbog – opstået som udtræk fra DanNets møbel-hierarki og endnu ikke forsynet med metadata (se også *møbel*, www.andreord.dk).

Typisk forestår der dog et efterfølgende finredigeringsarbejde. DanNet dækker kun 65.000 af DDO's 116.000 betydninger, og derudover kan den samme betydning være relevant at nævne i flere kapitler i begrebsordbogen; derfor bliver underbegrebshierarkierne suppleret med yderligere data fra DDO. Til dette arbejde har vi oprettet et særligt xml-dokument² der indeholder data fra DDO vedrørende ordenes semantik: definitioner, konnotative betydninger og encyklopædiske oplysninger, genus proximum, fag- og domæneoplysninger samt oplysninger om ordforbindelser, synonym-

² Udarbejdet af Thomas Troelsgård, DSL.

mer, antonymer og nærsynonymer. Xml-dokumentet er desuden for hver betydning suppleret med digitaliserede oplysninger om synonymer taget fra en dansk synonymordbog (Schultz 1997) til yderligere inspiration når der skal findes udtryk til et kapitel i begrebsordbogen. Via disse supplerende søgninger i DDO opdager redaktøren ofte endnu flere udtryk eller kommer selv i tanke om ord som endnu ikke er medtaget i DDO. Også en tidligere dansk begrebsordbog, Harry Andersen (1945), bliver konsulteret i arbejdet med at finde udtryk. Udtryk der ikke står i DDO i forvejen, medtages hvis de er udbredte på internettet og markeres samtidig som lemmakandidater til DDO. I arbejdet med foreløbig fem kapitler i begrebsordbogen er der tilføjet 437 lemmakandidater samt i øvrigt også knap 400 ordforbindelser der ikke er registreret i DDO. Begrebsordbogen kan derfor 'føde tilbage' til DDO med typiske ordforbindelser og ord af nyere dato i forhold til DDO's korpus (fx *simremad*, *øhop* og *running sushi* – første belæg i Infomedia henholdsvis 1997, 1998 og 2009) samt med huller i DDO's ordforråd. Dette gælder fx ord som *krydderkage*, *kanelbrød* og *dåseskjul*, der kan dateres tilbage til DDO's periode eller før.

3.1. Ikke-anvendelige hierarkier fra DanNet

Begrebsordbogens kapitloverskrifter lægger op til en beskrivelse af domænespecifikke situationer og deres tilhørende handlinger, egenskaber, aktører og rekvisitter – og dette går i mange tilfælde på tværs af de sprogligt set korrekte overbegreber man finder i DanNet. Derfor skal ko-hyponymer ofte fordeles på forskellige kapitler i begrebsordbogen. Fx har {maske} i DanNet følgende underbegreber: {ansigtsmaske}, {muddermaske}, {iltmaske}, der i begrebsordbogen placeres under 2.23 *Sundhed; hygiejne*; {fægtemaske}, der placeres under 17.19 *Kampsport* i begrebsordbogen; {dykkermaske}, der placeres under 17.20 *Vandsport*, og {fastelavnsmaske}, der placeres under 17.27 *Selskabelighed*. Et andet eksempel

er {pind}, der i DanNet har underbegreberne {kniplepind}, {kødpind}, {pølsepinde}, {røgelsespinde} og {trommestik}. {Kødpind} og {pølsepinde} hører i begrebsordbogen under 16.10 *Madlavning*, {trommestik} under 14.15 *Musikinstrumenter* og {kniplepind} under 19.08 *Stoffer, tekstiler*. Især når der er tale om meget flade, ordrige hierarkier i DanNet, er dette et problem. Nogle eksempler er konkrete overbegreber som {person} og abstrakte begreber som {følelse} og {tidsperiode}, der alle har et meget stort antal direkte underbegreber. Fx er {fritid}, {fyringssæson} og {halvleg} alle direkte underbegreber til {tidsperiode}. Udtrækkene kan i stedet fungere som oversigtslister, idet især udtryk fra de abstrakte betydningsområder heller ikke er lette at udtrække vha. søgninger i DDO. Lister over bestemte semantiske områder kan i øvrigt også i visse tilfælde fremfindes i DanNet ud fra den ontologiske type; fx har alle tidsbetydninger typen [TIME].

Alt i alt kan man konkludere at DanNet er særdeles velegnet til hurtigt at indsætte forholdsvis store mængder data i begrebsordbogen og dermed skabe en grundstruktur i et kapitelafsnit. Særligt de tilfælde hvor der er relativt mange, men ikke uoverskueligt mange underbegreber, og hvor overbegrebet selv hører til under det pågældende afsnit i begrebsordbogen, er velegnet til genbrug. I det videre redigeringsarbejde suppleres disse grundlæggende data med udtryk fra DDO.

Når et overbegreb er meget generelt i forhold til den domæneorienterede begrebsordbog, forestår et større arbejde med at sortere underbegreberne på forskellige kapitler. Denne sortering i domæner kan senere anvendes til at forbedre DDO's visning af beslægtede ord. Oplysning om domæne-tilhørsforhold kan ligeledes indsættes direkte i DanNet som en form for ekstra 'concerns'-relation og supplere de nuværende relationer. Og selve suppleringen med flere udtryk fra DDO vil mere eller mindre automatisk kunne anvendes til at udvide dækningsgraden i DanNet, ligesom den vil sikre at beslægtede ord-funktionen i DDO tilbydes ved

endnu flere opslagsord. Der kræves dog en række forholdsvis nemt indsatte metadata i begrebsordbogen for at genbruget kan foregå automatisk; dette beskriver vi i næste afsnit.

4. Semantisk struktur i begrebsordbogen

I modsætning til andre begrebsordbøger der præsenterer ordene efter ordklasse (jf. Trap-Jensen 2011), arbejder vi i den nye danske begrebsordbog med semantisk definerede grupperinger på første opdelingsniveau inden for et kapitel. Også andre semantiske grupperinger end den ovenfor beskrevne (underbegreber til samme overbegreb) er mulige. Semantisk gruppering i skrivefasen letter arbejdet med at skrive ordbogen da redaktøren i arbejdsprocessen 'tænker' i betydninger og leder efter udtryk ud fra semantiske kriterier – ikke ud fra ordklasser. Samtidig giver den semantiske gruppering os mulighed for at specificere en lang række underliggende metadata i form af semantiske relationer der gælder mellem ordene i gruppen. Disse metadata er som nævnt nødvendige hvis ordbogens oplysninger skal genbruges i DDO og DanNet. Andre begrebsordbøger (fx Dornseiff 2004, Harry Andersen 1945, Roge's Thesaurus (2002) og Norsk Tesaurus 2001) indeholder ofte en implicit semantisk gruppering på 2. niveau, fx ved at nævne et overbegreb fulgt af dets underbegreber inden et nyt overbegreb nævnes, eller ved at præsentere en række adjektiver der beskriver samme type kerne, inden en ny række adjektiver nævnes. I den nye danske begrebsordbog ekspliciteres denne viden i form af metadata der tilknyttes de enkelte grupperinger.

I modsætning til de omtalte begrebsordbøger indeholder den desuden også grupper med ord fra flere ordklasser. Det har den klare fordel at ord der ellers ville mangle kontekst fordi det semantiske område er ordfattigt, kan hægtes på en gruppe ord fra en anden ordklasse med beslægtet betydning. De fire omtalte be-

grebsordbøger har i vores øjne også det problem at de har for vage markeringer af de centrale begreber. I den nye danske begrebsordbog ønsker vi en tydelig markering af nøgleord (fx med fed skrifttype) på et forholdsvis detaljeret niveau. Fx markerer vi ofte et overbegreb inden underbegreberne nævnes. Ligeledes fremhæves det første synonym ud af en række synonymer typisk.

Vi etablerer så store semantiske grupperinger at man opnår den ideelle gensidige informationseffekt fra naboordene. Men da vi samtidig gerne vil undgå at grupperingerne skal blive for store og uoverskuelige, har vi udarbejdet forskellige ret fleksible regler for tilladte semantiske grupper. Man kan som redaktør vælge hvilken gruppering man vil benytte, afhængigt af antallet af ord i det givne domæne. Ved meget ordrige semantiske områder benytter redaktøren typisk en gruppering udelukkende med ord fra samme ordklasse (ofte en over-underbegrebs-gruppe), mens man ved mindre ordrige områder grupperer på tværs af ordklasser for at opnå en tilpas stor og informativ gruppe af ord. På denne måde tilgodeser vi både ordrige og mere ordfattige semantiske områder af sproget.

De semantiske grupperinger der er tilladt, er opstillet på baggrund af de intuitive grupperinger der blev foretaget i to gennemarbejdede kapitler (ca. 10 % af begrebsordbogen). Alle grupperinger er gjort fleksible idet de indeholder mulighed for en 'hale', hvor nært beslægtede udtryk der ikke overholder gruppens overordnede systematik, kan nævnes. Og alle grupperinger indeholder oplysninger om hvilken semantisk relation der gælder mellem ordene i gruppen, fx præcise oplysninger om hvad der er det fælles overbegreb eller hvem/hvad der har den egenskab der beskrives i gruppen. Vi forsyner desuden de semantiske grupperinger med forklarende overskrifter til brugeren.

Metadata-inventaret der anvendes, består af nogle af de semantiske relationer der allerede bruges i DanNet, men der er også indført nye relationer, inspireret af EuroWordNets og SIMPLE's forslag til semantiske relationer (Vossen 1998, Vossen 1999; Len-

ci et al. 2000). Ud over at metadataene, som tidligere nævnt, vil kunne bruges til at indsætte supplerende relationer i DanNet, vil de også muliggøre særlige præsentationer af ord både i ordnet.dk/ddo og på tværs af kapitlerne i begrebsordbogen.

På trods af den semantiske tilgang ønsker vi at have mulighed for at præsentere ordene på samme måde som andre begrebsordbøger, dvs. ordklassesorteret. Den ordklassesorterede visning har traditionen bag sig (jf. Trap-Jensen 2011), men den semantiske visning har også fordele, idet det jo netop er de omgivende ords semantik i en begrebsordbog der gør at man overhovedet får information om det enkelte ord. Jo færre semantisk beslægtede ord i en gruppe, jo mindre information om de enkelte ord, kunne man sige. Ordklassemarkering på hvert udtryk sikrer mulighed for ren ordklassesortering hvis det foretrækkes til den trykte ordbog eller til en supplerende digital visning i den elektroniske udgave af ordbogen. Vi vender tilbage til de forskellige typer af grupperinger i næste afsnit idet vi først vil se på den semantiske mikrostruktur.

4.1. Semantisk mikrostruktur

En semantisk gruppering består af flere ‘bundter’ af særligt tæt beslægtede ord. Selve grupperingen kan indledes af et nøgleord som i figur 2 (fx *bord* og *arbejdsbord*), som samtidig udgør overbegrebet for alle ord i gruppen (dette præciseres i metadata). Selve bundterne består typisk af nær-synonymer (hvis andet ikke er markeret for gruppen) eller af synonymer (markeret med *syn* i redigeringsværktøjets visning af data, se fx figur 3, de sidste fire udtryk i halen).

delige udtryk (vurderet alene ud fra introspektion) (fx <*sejlsport, søsport, sejlad*> og <*ro, padle, sejle i kajak, ro i kajak*>. Hvis undergruppen består af et overbegreb og dets underbegreber, vil det være overbegrebet der er markeret, fx <*gangart, pasgang, trav, galop, galopade*>. Nogleord kan række hen over flere undergrupper hvis der ikke er stort betydningsskift. Dette gælder fx undergruppen <*pukkelryg, pukkel, udvækst, gevækst, vansiret ansigt, skæmmende ar, hæsligt/grimt fjæs, tud*> i figur 4. Det indledende og markerede ord i synonym-gruppen umiddelbart før, *misdannelse*, fungerer som en overskrift for alle disse udtryk.

Der bliver ikke foretaget nærmere sproglige undersøgelser for at udfylde markeringer af undergrupperne; de beror udelukkende på subjektive skøn eller på oplysninger om synonymi og ordfelter der allerede er anført i DDO.

4.2. Semantiske metadata på grupperne

En overordnet semantisk gruppes overbegreb angives som metadata i starten af gruppen (se figur 4) i form af et typenummer (*01_Overbegreb*), et udtryk (*rædsel*) med tilhørende id-nummer og et attribut der angiver relationstypen ('has_hyperonym'). Id-nummeret sikrer som tidligere nævnt at oplysningen kan genbruges i DanNet og DDO. Det kan også være holonymet (den helhed som noget er del af) der indsættes som metadata, eller det kan være oplysning om hvem egenskaber i gruppen gælder (fx *property_of {person}* i figur 3). Ved grupper uden systematisk relation mellem ordene anvendes den lidt løsere relation 'concerns'. Fx har gruppen <*fitness, gå til fitness, fitnessinstruktør, fitnesscenter*> kun oplysningen 'concerns {fitness}' som metadata. Også andre relationer kan indsættes, fx hvem der udfører handlingerne i gruppen ('involved_agent').

De fleste grupper indeholder et afsluttende felt, en hale, hvor de relationer der angives som metadata, ikke længere gælder, og

hvor semantisk beslægtede udtryk alligevel kan nævnes. Se fx figur 3, hvor adverbialerne *<med sindsro, uden at blinke, uden at fortrække en mine, gladelig>* nævnes til sidst.

4.3. Grupperinger på tværs af ordklasser

Grupperinger der består af flere ordklasser (også når halen ikke medregnes), bruges i to tilfælde. For det første når der inden for et bestemt betydningsområde er tale om forholdsvis få ord fra mindst en af ordklasserne – så få at de, hvis de står alene, udgør en meget lille og derfor nemt lidt intetsigende gruppe. Ved at hægte de få ord på en større ordgruppe af over- og underbegreber får man både fortalt noget om den semantiske relation mellem ordene fra de to ordklasser (idet hver type gruppering bærer et typenummer), og man får etableret en mere informationsrig gruppering af en vis størrelse. Fx er en lille gruppe adjektiver ofte placeret efter den gruppe af substantiver som de typisk er adled til. Eks. fra kapitel 16.06 *Mad og madretter*:

frugt, eksotiske frugter, sydfrugter, kernefrugt, stenfrugt, modne frugter, runde frugter, nedfaldsfrugt, nedfalden frugt, nedfaldsæble – moden, solmoden, plukkemoden, frisk, nedfalden

Andre tilfælde hvor det er tilladt at hægte forholdsvis små ordgrupper på en over-underbegrebs-gruppering, er:

- Verber og de udråb der bruges i den givne situation (*væmmes / bvaðr*).
- Substantiver og de verber der udføres af substantivet (*dommer / dømme; cyklist / cykle*).
- Substantiver og de verbaludtryk der har dem som objekt (*pibe / stoppe; håndbold / gå til håndbold*).
- Substantiver der angiver sted, og de handlinger der foregår det pågældende sted (*mål / komme i mål*).

For det andet laves der grupperinger på tværs af ordklasser når der er tale om manifestationer af det samme begreb. Nogle af disse grupper er rene afledningsgrupper, hvoraf især én er blevet hyppigt anvendt i det hidtidige arbejde. Den udgøres af verber og deres modsvarende verbalsubstantiver. Også her angives der metadata for gruppens ord, idet der gives information om hvad overbegrebet er til verberne henholdsvis verbalsubstantiverne. Desuden gives der information om hvem der udfører handlingen, udløser begivenheden etc. En sådan gruppe kan fx se ud på følgende måde:

reagere, gå fra koncepterne, gå fra snøvsen, flippe ud, knække totalt sammen, bryde sammen, bryde grædende/hulkende sammen, gøre noget overilet, forhaste sig, forivre sig, forglemme sig (selv), fare op, bruse op etc. reaktion, respons, anfald, angstreaktion, nervesammenbrud, den umiddelbare reaktion, negative reaktioner etc.

I dette tilfælde indeholder den bagvedliggende struktur metadata om at den involverede agent er {person}, og at overbegreberne er {reagere} (for verberne) og {reaktion} for substantiverne.

En anden gruppering på tværs af ordklasser hvor der er tale om nærtbeslægtede begreber, er en gruppering af verber (samt eventuelt også de tilhørende verbalsubstantiver) og de adjektiver der betegner en efterfølgende tilstand forårsaget af den behandling som verberne udtrykker. Ofte vil adjektivet være en participiumsform af verbet og være beskrevet under verballemmaet i DDO, men der kan også være tale om ikke-afledte adjektiver. Et eksempel fra kapitel 16.07 *Krydderi* er:

krydre, smage til, tilsmage, runde af, tilsætte, iblande, blande i, komme i, pebre, salte, krydring, tilsmagning, tilsætning, iblanding, krydret, pebret, pikant, hot, stærk

Til sidst vil vi nævne en gruppering på tværs af ordklasser der har vist sig meget anvendelig i det foreløbige arbejde med begrebsordbogen – en semantisk gruppering der samler egenskaber på tværs af ordklasser (fx *begejstret*, *begejstring*, *i højt humør*, *ikke være til at styre på en hel tønde land*). Egenskaber udgør en forholdsvis stor del af ordforrådet. De beskriver ofte enten en person (som i eksemplet ovenfor), en genstand eller en hændelse, og vi ønsker at inddele egenskaberne i tydelige grupper ud fra dette. Desuden ønsker vi at præcisere hvad egenskaberne lægger sig til, både i form af en eventuel overskrift til brugeren og i en formaliseret form som metadata (sammen med de sædvanlige overbegreber). Figur 3 ovenfor er et typisk eksempel. Ofte kræver inddelingen information om ordene som ikke er ekspliciteret i DDO, når man som redaktør skal tage stilling til hvem eller hvad egenskaben gælder. Fx skelnes der i et andet kapitel 10.16 *Beundring*, *skønhed* mellem udtryk der beskriver smukke personer, og udtryk der beskriver smukke genstande (se figur 5):

Om person {04_Egenskaber/has_hyperonym: smuk
property_of: person}
 ▶ **smuk**, skøn, flot◄; ▶fin, graciøs, køn, vakker,
 henrivende, yndig, fager◄; ▶syn: underskøn, billedskøn,
 gudeskøn, ualmindelig(t) smuk, ufattelig(t) smuk,
 usædvanlig(t) smuk◄; ▶smuk og dejlig, ung og smuk,
 smuk og tiltrækkende, dårende dejlig◄; velplejet; ▶syn:
 velbygget, velskabt, veldrejet◄; kvieøjet; ▶syn: lækker,
 laber◄; ▶syn: smart, fiks, chik, elegant◄; ▶hyp:
skønhed, smukt udseende, kvindelig skønhed, ydre
 skønhed◄; ▶se godt ud, virke tiltrækkende◄; ▶lyse op,
 stråle◄; blomstre→→→skønhedsideal, man må lide for
 skønheden

Om genstand {04_Egenskaber/has_hyperonym: flot
property_of: genstand}
 ▶ **flot**, skøn, smuk◄; ▶ vakker, yndig, fager◄; ▶ **syn**:
ualmindelig(t) smuk, underskøn, ufattelig(t) smuk,
 usædvanlig(t) smuk◄; ▶ **syn**: abeskøn, skæppeskøn,
 dødskøn, superskøn◄; smuk og harmonisk; ▶ vidunderlig,
 paradisisk, overjordisk, underbar, underfuld◄;
 ▶ **overdådig**, prægtig, prud◄; ▶ **syn**: **smart**, fiks, chik,
 elegant, raffineret◄; ▶ **syn**: lækker, fjong, laber◄; ▶ super,
 superb◄; ▶ **glansfuld**, glitrende, strålende◄;
 ▶ **farvestrålende**, farveprægtig◄; flatterende, æstetisk,
 ren, rigt udsmykket, smukt udsmykket; ▶ **skønhed**, pragt,
 pyd, pænhed◄; ▶ ynde, gratie◄; æstetiske kvaliteter,
 ydre skønhed; ▶ **glans**, stjernestøv, glamour,
 overdådighed◄; naturskønhed; ▶ **se pæn ud**, pryde,
 pynte, passe perfekt/fint◄; ▶ **stråle**, lyse op, funkle,
 funkle, tindre◄; ▶ funkle, opalisere◄; ▶ **klæde nogen
 godt**, klæde nogen◄; → **æstetik**, æstetik, æstetisk
 sans, skønhedssans; ▶ **æstet**, skønånd, æstetiker◄;
 ingen roser uden torne

Figur 5: Kapitel 10.16 i begrebsordbogen. Ord der udtrykker skønhed ved personer henholdsvis genstande.

5. Begrebsordbogens omfang

Ordbogen vil komme til at indeholde et forholdsvis højt antal udtryk, væsentligt flere end Norsk Tesaurus (der kun indeholder ca. 7.000 udtryk), men også flere end både Harry Andersen (en skønsmæssig optælling af den alfabetiske ordliste giver ca. 33.500 forskellige udtryk) og Dornseiff (ca. 56.000 udtryk, også skønsmæssigt optalt). Den vil snarere nærme sig Rogets størrelse (over 300.000 udtryk ifølge oplysninger på omslaget).

Vi kan også konstatere at den nye danske begrebsordbog bliver ca. tre gange så omfangsrig som Harry Andersen, alene ved at se på antallet af udtryk i de færdiggjorte kapitler. Fx indeholder kapitel 10.08 *Ufølsomhed; sindsro* (figur 3) 230 udtryk i sammenligning med Harry Andersen, der har 75 udtryk i det tilsvarende kapitel. Dette er et meget gennemsnitligt eksempel på forskellen og svarer også meget godt til at DDO som leverer udtryk til begrebsordbogen, indeholder ca. 102.000 lemmaer og faste udtryk.

Begrebsordbogen kommer desuden til at indeholde et stort antal ordforbindelser, enten taget fra DDO's i alt ca. 32.500 oplysninger om typiske adled, kerne og andre naboord eller tilføjet som helt nye data via introspektion og frekvenstjek på internettet. Sådanne typiske ordforbindelser vil give god vejledning i sproglig formulering til brugere der ønsker at bruge begrebsordbogen til skrivehjælp, men de vil også kunne finde anvendelse inden for det sprogteknologiske fagområde, hvor de traditionelt betegnes *bi-grams*, *trigrams*, *n-grams*, idet de i begrebsordbogen er semantisk entydiggjort. I figur 4 ses et udsnit af kapitel 10.08 *Ufølsomhed*; *sindsro*'s 230 udtryk, hvor man ud over enkeltord og faste forbindelser med egne betydningsdefinitioner i DDO kan se adskillige typiske ordforbindelser (*i balance med sig selv, hvilende i sig selv, let til mode, svær at ryste* etc.). Begrebsordbogen angiver i øvrigt også typiske bøjningsformer, fx *afbalanceret*, der ikke er selvstændigt opslagsord i DDO, men beskrives som participium under lemmet *afbalancere*.

6. Konklusion og videre perspektiver

Redaktionsreglerne for de semantiske grupperinger er blevet til i et forsøg på at skabe klare linjer i de mange måder man kan gruppere ord på inden for et betydningsområde. Reglerne sikrer at man griber sorteringen af ord an på en nogenlunde ensartet måde fra kapitel til kapitel samt redaktørerne imellem. Det at man er tvunget til at tænke over hvilken type gruppe man er ved at etablere, samt hvilke metadata der knytter sig til den, sikrer også at grupperne bliver stringente. Samtidig gør halerne og 'nødløsningen' i form af en tilladt gruppe uden fast struktur det muligt også at beskrive udtryk der er svære at indplacere i en systematisk gruppe.

Grupperne og de forskellige muligheder for at angive metadata om relationerne mellem ordene er lavet ud fra et ønske om

på et senere tidspunkt at kunne anvende de meget strukturerede data i andre sammenhænge. Begrebsordbogens struktur bygger på de semantiske relationer der allerede bruges i DanNet, og det er derfor muligt at udnytte oplysningerne i den til at udvide antallet af eksisterende relationer i DanNet. Man vil fx kunne anvende de mange oplysninger om overbegreber til at supplere, evt. korrigere hierarkier i wordnettet. Særligt tilføjes der mange funktionelle over-underbegrebs-grupper i begrebsordbogen i forhold til hvad man finder i DanNet (fx grupper af fodboldudstyr, haveting osv.). Og fra mange af grupperne der indeholder ord fra mere end én ordklasse, vil man automatisk kunne supplere DanNet med bestemte semantiske relationer der gælder for de enkelte ord i gruppen ved at anvende de formaliserede oplysninger om typenummer og relationer i metadata.

Der er også indført relationer i begrebsordbogen som ikke anvendes i DanNet, og som det er meget oplagt at udvide DanNet med. Den oplysning der tegner til at kunne bidrage med mest ny information i DanNet, er relationen 'property_of'. Egenskaber er ikke behandlet særligt detaljeret i DanNet, og det vil være en klar forbedring af wordnettet inden for dette betydningsområde hvis en relation der beskriver hvad egenskaben lægger sig til, kan ind-sættes automatisk ud fra begrebsordbogens metadata. Fx vil alle ord i figur 3 automatisk kunne tilskrives relationen 'property_of {person}'. Der eksperimenteres i øjeblikket med at integrere disse data automatisk fra begrebsordbogen til DanNet.

Hvad angår brug af data fra begrebsordbogen i DDO, har et af formålene med begrebsordbogen været at opnå leksikalske data der kunne forbedre beslægtede ord-funktionen på ordnet.dk/ddo, jf. Trap-Jensen (2011). Dette vil absolut være muligt pga. begrebsordbogens detaljeringsgrad: Man vil på det snævraste semantiske niveau kunne vise kun de øvrige udtryk i samme undergruppe. Man vil også kunne vælge at præsentere hele grupperingen med alle dens undergrupper og den efterfølgende hale, og man vil end-

da kunne vise hele kapitlet som ordbetydningen optræder i. For nogle ords vedkommende kan der være tale om flere tilhørsforhold og dermed flere visninger.

De forskellige metadata vil også muliggøre særlige visninger af DDO's data på tværs af kapitlerne i begrebsordbogen. Endvidere får vi en meget interessant mulighed for at forklare et ords betydning alene ved at præsentere dets 'omgivelser' i begrebsordbogen, en visning der vil svare til en synonymdefinition. Ca. 18.000 lemmaer i netudgaven af DDO har nemlig endnu ikke definitioner da de i den trykte første udgave af DDO blot var nævnt som gennemskuelige sammensætninger eller afledninger under andre lemmaer. I den digitale udgave er de ophøjet til selvstændige lemmaer med morfologiske oplysninger. Mange adjektiver mangler fx definitioner; i det øjeblik disse adjektiver er medtaget i begrebsordbogen, er vi i stand til at give en synonymdefinition af dem, en definitionstype der i forvejen er brugt for adjektiver i DDO. Fx står adjektivet *uforsagt* der ikke er defineret i DDO endnu, i følgende undergruppe i begrebsordbogen <*på højde med situationen, cool, all right, fit for fight, fuld af gåpåmod, vel til mode, uforfærdet, ufortrøden, uforsagt, kæk, frejdig*>. Visningen af disse omgivende ord vil give en udmærket forklaring på betydningen af ordet.

Som det kan ses, åbner arbejdet med begrebsordbogen op for mange muligheder inden for det digitale leksikografiske arbejde der udføres på DSL og præsenteres via ordnet.dk.

Litteratur

Ordbøger

Andersen, Harry 1945: *Dansk Begrebsordbog*. København: Munksgaard.

- DDO = Ebba Hjorth, Kjeld Kristensen m.fl. 2003-2005: *Den Danske Ordbog*. København: Gyldendal og Det Danske Sprog- og Litteraturselskab. Onlineversion: <http://ordnet.dk/ddo>.
- Dornseiff, Franz 2004: *Der deutsche Wortschatz nach Sachgruppen*, 8. Auflage, Berlin/New York: Walter de Gruyter.
- Roget, Peter Mark: *Roget's Thesaurus*, 150th anniversary edition edited by George Davidson 2002. London: Penguin.
- Schultz: Ulla Albeck, Mikal Rode og Erik Timmermann 1997: *Dansk Synonymordbog*. København: J.H. Schultz Information A/S.
- Norsk Tesaurus: Johan Hammond Rosbach 2001: *Ord og begreper Norsk tesaurus*. Oslo: Pax Forlag A/S.

Anden litteratur

- Fellbaum, Christiane 1998 (ed.): *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Johannsen, Anders & Bolette S. Pedersen 2011: "Andre ord" – a wordnet browser for the Danish wordnet, DanNet. I: *Proceedings from 18th Nordic Conference of Computational Linguistics, NODALIDA 2011*, Riga, Latvia. Northern Association for Language Technology, University of Tartu, Vol. 11, 295-298.
- Lenci, Alessandro, Nuria Bel, Federica Busa, Nicoletta Calzolari, Elisabetta Gola, Monica Monachini, Antoine Ogonowsky, Ivonne Peters, Wim Peters, Nilda Ruimy, Marta Villegas, Antonio Zampolli: 2000. SIMPLE: A General Framework for the Development of Multilingual Lexicons. I: *International Journal of Lexicography*, 13(4), 249-263.
- Lindén, Krister & Lauri Carlson 2010: FinnWordNet – WordNet på finska via översättning. I: *LexicoNordica* 17, 119-140.
- Lorentzen, Henrik & Sanni Nimb 2010: Fra ordbog til wordnet. Hvordan udmøntes en traditionel ordbogsdefinition i en formaliseret wordnetbeskrivelse? I: *Nordiska studier i lexicografi* 10, 329-344.

- Pedersen, Bolette S. 2010: Semantiske sproressourcer – mellem sprogteknologi og leksikografi. I: *LexicoNordica 17*, 163-180.
- Pedersen, Bolette S., S. Nimb, J. Asmussen, N. Sørensen, L. Trap-Jensen, H. Lorentzen 2009: The challenge of compiling a WordNet for Danish by reusing a monolingual dictionary. I: *Language Resources and Evaluation, Computational Linguistics Series*, Volume 43, Number 3: 269-299.
- Pustejovsky, James 1995: *The Generative Lexicon*. Cambridge: MIT Press.
- Trap-Jensen, Lars 2011: Dansk begrebsordbog: baggrund og metode. I: *LexicoNordica 18* (i dette bind).
- Vossen, P. 1998 (ed.): *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Kluwer Academic Publishers, Dordrecht.

Internethenvisninger

- andreord.dk ved Anders Johannsen. Viser data fra DanNet.
DanNets hjemmeside: wordnet.dk
- Vossen, P. 1999 (ed.): *EuroWordNet General Document, Version 3 Final* <http://www.illc.uva.nl/EuroWordNet/finalresults-ewn.html>

Sanni Nimb
seniorredaktør, ph.d.
Det Danske Sprog- og Litteraturselskab
Christians Brygge 1
DK-1219 København K
sn@dsl.dk

Ontologi, begreppssystem och WordNet

Klaas Ruppel

In this article I try to point out similarities and differences between ontologies, dictionaries of synonyms and word nets. The focus lies on ontologies in relation to the other two types. The universal applicability and language-independentsness claimed for top hierarchies of ontologies is also discussed. As examples I look at two ontologies (Suggested Upper Merged Ontology and EuroWordNet Top Ontology), a dictionary of synonyms (Lagercrantz 1939) and the Princeton WordNet and some of its offsprings (EuroWordNet, DanNet). I emphasize particularly the content and the weight of ontological work and do not discuss datatechnology.

I denna artikel diskuterar jag likheter och skillnader mellan ontologier, synonymordböcker och ordnät. Som exempel använder jag två ontologier (Suggested Upper Merged Ontology och EuroWordNet Top Ontology), en samisk synonymordbok samt Princeton WordNet och två andra ordnät som haft WordNet som incitament (EuroWordNet, DanNet).

1. Ontologin

Ontologi är ett begrepp inom teoretisk filosofi. Ordet är bildat av de grekiska orden för 'varande' och för 'lära'. Ontologi är alltså läran om varandet eller om tillvaron. Ordet uppträder första gången i början av 1600-talet (Lorhard 1606), trots att man hade funderat mycket över varandet redan under antiken, då främst inom metafysiken. Ända från början har kategorisering spelat en stor roll vid behandling av varandet.

Inom modern informationsteknik är begreppet ”ontologi” mycket mer praktiskt till sitt innehåll. Det är ”en explicit specificifikation av en delad konceptualisering.” (Gruber 1993:199). Med hjälp av en ontologi presenteras alltså begreppsapparaten inom ett ämnesområde på ett sådant vis att den kan utnyttjas för informations-tekniska syften.

I en ontologi är idealtillståndet att varje entitet¹ har en och endast en entitet som närmast övergripande entitet. I de två ontologierna som jag granskar här är detta dock inte helt och hållet genomfört. I den ena, Suggested Upper Merged Ontology (SUMO), har vissa av entiteterna fler än en övre entitet. Den andra, toppontologin för EuroWordNet, är å sin sida bara delvis hierarkisk, eftersom entiteterna definieras också via komponentanalys.

I denna artikel ska jag fokusera på den högsta delen av ontologierna, på *toppontologierna* (*Upper Ontology*, *Top Ontology*), dvs. på den del av ontologierna som är planerade att vara universella och språkoberoende.

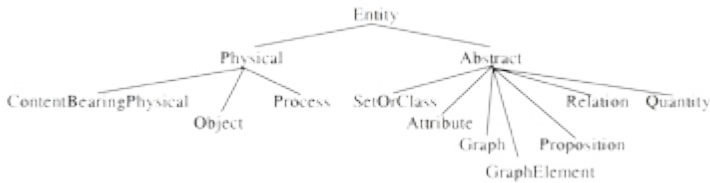
1.1. Suggested Upper Merged Ontology

SUMO är en ontologi som numera upprätthålls av Adam Pease, som arbetar på ett företag som producerar webbapplikationer. Anledningen till att jag i detta sammanhang granskar just den ontologin, är att den har tillämpats på WordNet (Niles & Pease 2003; Index of WordNetMappings). SUMO-systemet är uppdelat i tre delar. Den lägsta nivån utgörs av terminologier inom vissa fackområden. Den högsta nivån, själva SUMO, utgör en toppontologi baserad på logiska entiteter. Toppontologin samlar begreppen på de lägre nivåerna hierarkiskt och för till sist ihop dem till en enda topp. Mellan den översta och den nedersta nivån befinner sig det

1 För tydlighets skull använder jag termen *entitet* när jag talar om de hierarkiska noderna i ontologierna. Motsvarande term i begreppsordböcker och ordnät är *begrepp*.

hierarkiska mellanskiktet som alltså kopplar samman toppen och botten (se SUMO).

De terminologier som finns samlade på den nedersta nivån utgör ett mycket begränsat urval av verklighetens alla områden (t.ex. *Communications, Economy, Military, World Airports*). SUMO påstås emellertid vara oberoende av språk och kultur: ”SUMO is language independent.” Och vidare: ”... there is no deep-seated linguistic or cultural bias ...” (SUMOFAQ). Hur ser då den översta, allmängiltiga delen av SUMO ut? Den högsta noden i ontologin är ”Entity”, som förgrenar sig i de två noderna ”Physical” och ”Abstract”. Dessa delar i sin tur upp sig på många egna entiteter (figur 1):



Figur 1: SUMO:s högsta noder.

SUMO-systemets alla logiska entiteter och deras inbördes hierarkiska relationer finns på internet (SUMOclasses). När man granskar entiteterna ser man lätt att påståendet om universell giltighet för toppontologin inte håller streck. Argumentet verkar snarare gälla för kommersiell reklam än för vetenskap. Jag ska ge några exempel. Ontologin innehåller bl.a. följande hypernymkedjor (uppifrån och ner) (figur 2):

- 0 Entity
- 1 Abstract
- 2 Quantity
- 3 PhysicalQuantity
- 4 ConstantQuantity

5	TimeMeasure
6	TimePosition
7	TimeInterval
8	Month
9	Januar ... December

Figur 2: SUMO:s hypernymkedja för månader.

Entiteten ”månad” förgrenar sig alltså enligt det system som används i västvärlden med månaderna januari till december. Några andra månadssystem beaktas inte i ontologin. Likaså finns det utöver ”liter” bara anglosaxiska mått under noden ”VolumeMeasure”, och under ”CurrencyMeasure” förekommer bara dollar och euro. Exempelen av det här slaget är många. I praktiken är det på inget vis fatalt, för de saknade måtten, valutaenheterna osv. kan enkelt läggas till i samband med att nya terminologier läggs in i SUMO. Ur teoretisk synvinkel är det dock ytterst tveksamt om månaders namn och enskilda valutor ska ingå i en toppontologi som är avsedd att vara allmängiltig.

Vad gäller den påstådda allmängiltigheten är följande exempel mycket allvarligare på grund av att den ifrågasatta entiteten befinner sig betydligt högre upp i hierarkin, på kortare avstånd från toppnoden ”Entity” (figur 3):

0	Entity
1	Physical
2	Process
3	Motion
4	BodyMotion
5	Ambulating

Figur 3: SUMO:s hypernymkedja för ’förflytta sig till fots’.

Den nedersta entiteten ”Ambulating” ’det att förflytta sig till fots’ förgrenar sig vidare på entiteterna ”Running” och ”Walking”. Om springande och promenerande över huvud taget ska anses höra till de entiteter som ska ingå i en toppontologi så borde flera andra förflyttnings sätt som exempelvis krypande och galopperande läggas till för att täcka in benförsedda djurs olika sätt att ta sig fram. Begreppen kan förvisso ordna in sig just på det sätt som anges i SUMO på många språk, men exempelvis i samiska ser det annorlunda ut. Där återfinns skillnaden mellan tvåbentas och fyrbentas varelsers förflyttning på begrepps nivå. På nordsamiska heter det exempelvis *olmmoš vázzii* ’människan går’ men *boazu oahkuu* ’renen går’. Om man vill att ontologin ska lämpa sig som översta hierarki också för de samiska språken borde antingen entiteterna ”tvåbenta varelsers förflyttning” och ”fyrbenta varelsers förflyttning” läggas till under noden ”BodyMotion” eller så borde entiteten ”Ambulating” tas bort ur toppontologin.

Sammantaget verkar SUMO förgrena sig ner till alltför låg nivå för att utgöra en toppontologi. Många av de nedre noderna måste snarare anses representera språk- och kulturbundna begrepp än logiska entiteter. Det här är en återspeglning av de terminologier som är knutna till SUMO. De ovan nämnda valutabenämningarna ansluter sig till terminologierna *Economy* och *Finance*, och måttorden ingår i terminologin *North American Industrial Classification System*. Eftersom ontologin har byggts upp genom att vissa terminologier kombinerats är resultatet i sig inte förvånande, men däremot förvånar påståendet om att ontologin skulle vara universellt giltig.

1.2. EuroWordNet Top Ontology

EuroWordNet är ett projekt som syftade till att översätta det engelskspråkiga WordNet till andra språk. Det ursprungliga EuroWordNet-projektet, som omfattade engelska, spanska, italienska

och nederländska, avslutades för över tio år sedan. Basen i Euro-WordNet utgörs av en toppontologi, *Top Ontology* (illustration 1; Rodríguez 1998:139; Vossen 1998:22).

Top ⁰	
1stOrderEntity ¹	2ndOrderEntity ⁰
Origin ⁸ Natural ²¹ Living ³⁰ Plant ¹¹ Human ¹⁰⁸ Creature ² Anima ¹⁷³ Form ⁵ Artifact ¹⁴⁴ Substance ³² Solid ⁴³ Liquid ¹³ Gas ¹ Object ¹⁴² Composition ⁵ Part ³⁶ Group ⁴³ Function ⁵³ Vehicle ⁴ Representation ¹² MoneyRepresentation ¹⁰ LanguageRepresentation ³⁴ ImageRepresentation ⁹ Software ⁴ Place ⁴³ Occupation ²³ Instrument ¹¹ Garment ⁷ Furniture ⁶ Covering ⁸ Container ¹² Comestible ³² Building ¹³	SituationType ⁵ Dynamic ¹³⁴ BoundedEvent ¹⁸³ UnboundedEvent ⁴⁸ Static ²⁸ Property ⁴¹ Relation ³⁸ SituationComponent ⁰ Cause ⁶⁷ Agentive ¹⁷⁰ Phenomenal ¹⁷ Stimulating ²³ Communication ³⁵ Condition ⁶² Existence ²⁷ Experience ¹³ Location ³⁶ Manner ²¹ Mental ⁶⁰ Modal ¹⁰ Physical ¹⁴⁰ Possession ²³ Purpose ¹³⁷ Quantity ³⁹ Social ¹⁰² Time ²⁴ Usage ⁴
3rdOrderEntity ³³	

Illustration 1: Toppontologin i EuroWordNet.

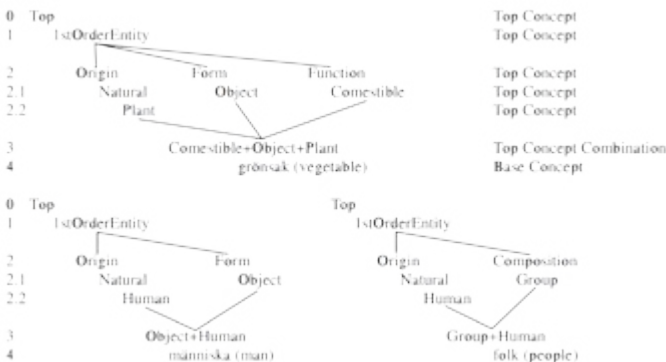
Ontologin består av tre typer av entiteter (Rodríguez 1998:136; Vossen 1998:21):

- 1stOrderEntity: en konkret entitet som kan uppfattas av sinnen och som har en plats i tid och rum
- 2ndOrderEntity: en statisk eller dynamisk situation som inte kan uppfattas som ett fysiskt objekt
- 3rdOrderEntity: ett påstående som är oberoende av tid och rum

Till entiteter av den första ordningen hör konkreta substantiv, till entiteter av andra ordningen substantiv, adjektiv och verb och till entiteter av den tredje ordningen abstrakta substantiv. (Rodríguez 1998:137; Vossen 1998:21) Tills vidare har andra ordklasser än substantiv och verb inte tagits med i ontologin. (Rodríguez 1998:148–149; Vossen 1998:31, Christiansson & Zimmerman 2006:35)

De 63 entiteterna i illustration 1 ingår i en mycket platt hierarki (jfr t.ex. Viberg 2002:136–138). De benämns *toppbegrepp* (*Top Concepts*, TC) och beskrivs som semantiska grundkomponenter ("fundamental semantic distinctions"). Nedanför denna hierarkiska nivå finns ett stort antal *basbegrepp* (*Base Concepts*, BC) som bestäms utifrån de semantiska grundkomponenter som visas i illustration 1. Indexsiffrorna anger hur många basbegrepp som definieras utifrån ett givet toppbegrepp. Basbegreppen kan definieras utifrån ett eller flera toppbegrepp. (En förteckning över alla toppbegrepp finns i Vossen (1998:34–44).) Den ontologiska strukturen är alltså en helt annan än i SUMO.

I figur 4 ges några exempel ur EuroWordNet. På raderna 0–2 i hierarkierna anges vilka semantiska grundkomponenter som använts vid definitionen och vilken deras plats är i hierarkin. På rad 3 syns komponentkombinationen och på rad 4 det basbegrepp som definieras genom kombinationen.



Figur 4: Definitionerna för basbegreppen "grönsak", "människa" och "folk" i EuroWordNet:s Top Ontology.

Toppontologin utgör EuroWordNets språkoberoende översta hierarkinivå, och med hjälp av den kan begrepp i de ingående språken kopplas till varandra. Ontologin är tudelad och består dels av entitetshierarkin, alltså de semantiska grundkomponenterna, dels av basbegreppen som erhålls genom kombination av grundkomponenterna. I toppontologin förenas alltså strukturalism och ontologi. (För mer information se Rodríguez (1998), Vossen (1998) och Viberg (2002).)

EuroWordNets toppontologi håller sig på ett mycket mer allmänt plan än SUMO, vilket troligen beror på att den från början är gjord för att vara språkoberoende och universell. Den är inte sammanställd utifrån ett smärre antal terminologier utan är avsedd att täcka språkets hela begreppssystem. Ändå påpekar författarna att man också i en ontologi på så här generell nivå konfronteras med språkliga skillnader. Som exempel nämns att det holländska språket inte har ett allmänt begrepp för ”container” (Rodríguez 1998:150; Vossen 1998:31).

1.3. En allmän finsk ontologi

I detta sammanhang bör det nämnas att det också i Finland bedrivs ontologisk forskning. YSO (*Yleinen suomalainen ontologia* ’allmän finsk ontologi’) baserar sig på Nationalbibliotekets ämnesordsindex. Det handlar sålunda om ett ontologiskt projekt av stor omfattning. Den gemensamma, allmänna toppontologin YSO utgör tak för 62 olika ämnesordsindex. Den innehåller för närvarande ca 23 000 begrepp och utgör en del av ett större projekt som syftar till att bygga upp ett nationellt semantiskt nätverk (FinnONTO). Det inbegriper också KulttuuriSampo, som tillhandahåller finländsk kultur via den semantiska webben.

1.4. Ontologier – ett informationstekniskt verktyg

De moderna ontologierna är inte utvecklade för människor utan för maskiner. De behöver därför inte följa intuitivt tänkande. Illustrationerna över SUMO och Top Ontology underlättar visualisering av ontologiernas struktur, men tekniskt består de av logiska fraser skrivna på programmeringsspråk.

Syftet med en ontologi är att göra det möjligt för datorn att ”förstå” betydelser. En dator kan skilja mellan olika teckensträngar. Den identifierar alltså till exempel ”tomat” och ”människa” som två olika saker och vet också att strängen ”potatis” är en tredje sak. Men utan ontologi kan den inte dra slutsatsen att tomat och potatis har mycket gemensamt (de är växter) och att en människa är något helt annat, även om människan har en relation till de här växterna (de odlas och de äts).

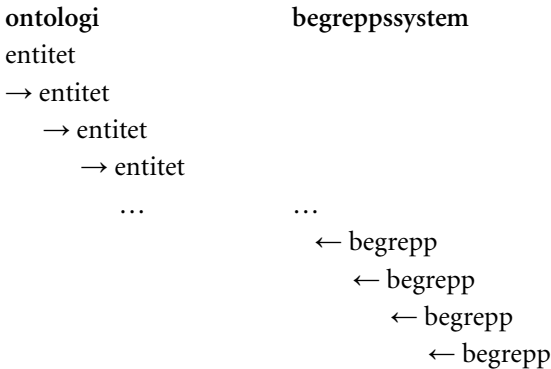
En ontologi försöker hitta likheter.

2. Begreppssystem för ett helt språk (tesaurus)

Begreppsordböcker är hierarkiskt strukturerade tesaurusar över hela språket. De skiljer sig från ontologier och datatekniskt skapade ordnät just genom det faktum att de presenterar hela lexikonet, inklusive slutna ordklasser som pronomen, konjunktioner, pre- och postpositioner etc. De grammatiska elementen har sällan en betydelse som kan fastställas utan kontext och de är därmed självfallet svåra att passa in i system där betydelsen eller begreppet (t.ex. ”huvudbonad”) har sin plats i systemet på basis av sin egen lexikala betydelse. Begrepp som ”när” eller ”att” är svåra att definiera lexikalt. I gränslandet mellan slutna och öppna ordklasser befinner sig pronomen, som å ena sidan utgör en sluten ordklass, å andra sidan har lexikal betydelse (t.ex. ”du”). Begreppsordböcker har en lång historia. Exempel som kan nämnas är Roget

(1856), Dornseiff (1933–1940), Wehrle (1942) och Nagy (1978).

Medan en ontologi byggs upp deduktivt skapas begreppssystemet i en begreppsordbok induktivt. Begreppen i ett språk samlas i synonymgrupper och dessa förses med lämpliga överbegrepp osv. (figur 5).



Figur 5: Ontologiska entiteter och språkliga begrepp.

På någon punkt möter den deduktiva kategoriseringen uppifrån och ner den induktiva kategoriseringen nerifrån upp. Den senare är kopplad till begrepp i språket, den grupperar språkliga begrepp och den är språk- och kulturbunden. Den deduktiva topontologin använder inte begrepp knutna till språk utan följer logikens lagar. Kategoriseringsenheten är entiteten. Skillnaden framgår direkt av hur begrepp och entiteter namnges. Typiska (språkberoende) begrepp är t.ex.:

byggnad, hus, stenhus; mat, gröt, morgongröt
Gebäude, Haus, Steinbau, Steinhaus; Speise, Brei, Morgenbrei
building, house, stone house; food, porridge

Typiska (logiska) entiteter i SUMO är t.ex.:

Process, IntentionalProcess, SocialInteraction, Contest

EuroWordNets toppontologiska kategorier (topp- och basbegreppen) används – beroende på ontologins strukturalistiska karaktär – som semantiska komponenter, men de är ontologiska element och därmed egentligen entiteter.

Entiteter kan översättas utan att referenten ändras. Däremot ändras referenten ofta i större eller mindre utsträckning om språkliga begrepp översätts från ett språk till ett annat. Organisationen av begrepp i över- och underbegrepp är språk- och kulturspecifik.

2.1. En begreppsordbok för de samiska språken

Jag ska nu undersöka begreppsordbokens natur utifrån en mindre känd språkgrupp, eftersom det på det sättet går bättre att få fram skillnaderna mellan begreppssystem å ena sidan och ontologier och ordnät å andra sidan. År 1939 publicerade Eliel Lagercrantz sitt storverk *Lappischer Wortschatz*, som innehåller ungefär 10 000 uppslagsord och omfattar 1036 sidor. Språken som beskrivs är i huvudsak umesamiska, pitesamiska, lulesamiska och nordsamiska eller de västsamiska språken. Ordboksartiklarna är etymologiskt ordnade så att de olika språkens motsvarigheter har kunnat redovisas tillsammans. Lemmat för en artikel är ett oavlett grundord. Avledningar och sammansättningar återfinns som sublemman, t.ex.:²

753. pites. *tjuoj'vat*, lules. *tjuojvvat*, nords. *čuoivjat* 'blå, grå'
753.4 pites. *tjuojvuk*, lules. *tjuojvuk*, nords. *čuoivvat* 'grå ren'

Figur 6: Utdrag ur artikel 753 (Lagercrantz 1939:105).

Lagercrantz begreppssystem *System der Bedeutungen* finns beskrivet på sidorna 1037 till 1182. Begreppssystemet togs fram in-

2 För tydlighetens skull presenterar jag bara en del av artikeln och jag använder också modern samisk ortografi.

duktivt. Lagercrantz tog ut alla betydelsangivelser för de samiska orden i sin ordbok och ordnade dem i begreppsgrupper. Det här innebär alltså att begreppssystemet omfattar hela den västsamiska språkgruppen, och inte bara ett språk.

På den översta nivån delas begreppen in i tre grupper (Lagercrantz 1939:1037). Till de metodologiska begreppen (I) hänför Lagercrantz kunskapsteoretiska, perceptionella och språkliga fenomen. Hit hör sinnesförmimmelser och högre varseblivningsområden såsom rum, rörelse, tid och form. Vidare ingår psykologiska begrepp och som fjärde grupp gester och språkstrukturella element såsom konjunktioner, räkneord och pronomen.

När de typerna av begrepp har sammanförts till en grupp är det naturligt att dela in resten av begreppsapparaten i två grupper, kulturella begrepp (II) och begrepp som hänför sig till naturen (III). De kulturella begreppen delas vidare in i samhällliga och ekonomiska. Naturbegreppen är indelade i undergrupperna organiska och icke-organiska.

Egentligen kan man redan av de tre nivåerna dra slutsatsen att de begrepp som är föremål för klassifikation tillhör ett kultursamhälle med traditionella näringar. Ända upp till den högsta hierarkiska nivån bygger systemet alltså på induktion, på de villkor som gäller för de beskrivna språken och dess kulturgemenskap, även om kategorin metodologiska begrepp som helhet naturligtvis är mer teoretisk till sin natur.

Det var redan ovan tal om en semantisk specialitet hos samiska, nämligen att tvåbenta och fyrbenta varelsers gående hålls isär. Det samiska ordförrådet har också andra särdrag, som återspeglar sig i begreppssystemet.

Många språk har ett ord för (djuret) ”ren” och möjligen därtill ord för renar av honkön och av hankön. I de samiska språken finns det över hundra olika renbenämningar. Renar kan särskiljas enligt en stor mängd olika kriterier, vilket visas i figur 7 (inom parentes

anges antalet begrepp som är inordnade under varje överordnat begrepp, Lagercrantz (1939:1141–1143)):

- ren, allmänt (24)
- rentjur, renoxe (16)
- renko (12)
- renkalv (9)
- renbeteckningar enligt ålder (25)
- renbeteckningar enligt färg (27)
- renbeteckningar enligt hornform (13)
- renbeteckningar enligt öronmärkning (6)
- renhjord, rajd (20)

Figur 7: Synonymgrupperna för 'ren' hos Lagercrantz 1939.

I den samiska kulturen har rennäringen varit en nyckelfaktor som hela storfamiljens välmåga varit beroende av.

Mycket rika är också begreppssystemen för vinter, is och snö (ca 200 begrepp) och för släktskapsförhållanden (ca 150). På de områden som samerna bebor är vintern lång, och ett överlevnadsvillkor är att man är förtrogen med och kan tolka vinternaturen.

Å andra sidan skiljer man inte på samiska mellan växande träd och trä som ämne (liksom inte heller på finska): det nordsamiska ordet *muorra* (och dess finska ekvivalent *puu*) motsvaras i svenska och tyska av två olika begrepp ”träd”, ”Baum” och ”trä”, ”Holz”.

En begreppsordbok beskriver en språkgemenskaps förhållande till omvärlden. Begreppssystemet är kopplat till språk- och kulturgemenskapen, och antalet begrepp är vanligen stort på just de områden som är särskilt viktiga för samhället ekonomiskt, psykologiskt, osv. Begreppssystem skiljer sig mest för olika språk just på de punkter där kulturerna skiljer sig.

Induktiva, språkbundna begreppssystem framhäver skillnader.

2.2. Gråzonen mellan ontologi och begreppssystem

Mellan ontologierna som opererar på hög hierarkisk nivå med logiska entiteter och de språkbundna hypernymkedjorna som byggs upp nerifrån finns det en gråzon. Inom SUMO kallas denna zon Mid-Level Ontology. Via den kopplas Upper Ontology, eller den egentliga ontologin samman med terminologier för olika fackområden. Med tanke på helhetsbeskrivningen är kanske just denna gråzon det mest intressanta området. Här möts entiteter och begrepp, logik och intuition, deduktion och induktion, det språkoberoende och det språkberoende. I WordNet (se avsnitt 3) kopplas den övre och den nedre delen av hierarkin, alltså ontologi och begreppssystem, samman så att begreppen längst nere ingår i hypernymkedjor som når ända upp till toppen av ontologin. Det varierar sannolikt från kedja till kedja och från språk till språk var ontologi och begreppssystem möts.

Inom de biologiska taxonomierna finns det däremot nästan inga gråzoner. *Taxonomi* är en biologisk term och syftar på en hierarkisk vetenskaplig klassificering av växter och djur. Också denna term har fått en annan betydelse inom informationsvetenskapen, där en taxonomi avser det hierarkiska systemet för vilken som helst fackområdesterminologi. Här håller jag dock isär termerna *taxonomi*, som avser ett hierarkiskt ordnat system, och *terminologi*, som inte nödvändigtvis avser ett hierarkiskt system. Taxonomier kan å ena sidan anses vara specialfall av begreppssystem, där objektet för systematiseringen inte är hela språket utan ett visst områdes termer. Å andra sidan påminner taxonomier och terminologier också om ontologiska system eftersom termer och taxoner³ – till skillnad från allmänspråkliga begrepp – är noga definierade.

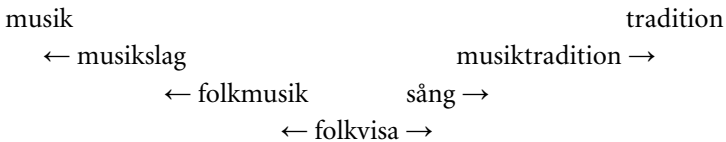
I fråga om både taxonomier och specialområdesterminologier finns det andra typer av problem. För det första kan nya forsk-

3 En taxon är en taxonomisk enhet (t.ex. ”änder”, ”andfåglar” och ”fåglar” är taxoner).

ningsresultat leda till ändringar i en taxonomi. En förändring sker inte över en natt utan är resultatet av vetenskapliga resonemang. Av detta följer att flera klassificeringar kan existera sida vid sida under en tid, och att den nya taxonomin först småningom vinner insteg. Det mest berömda exemplet på en taxonomi och dess variation torde vara det klassifikationssystem för växter som skapades av Linné (1753) och som sedermera undergått otaliga förändringar och preciseringar.

Terminologier och taxonomier är kopplade till det område där de tillämpas. Morfologin inom filosofi, biologi, medicin osv. är något helt annat än inom lingvistik, trots att det i samtliga fall är fråga om en formlära.

Ett närbesläktat problem är att ett och samma begrepp kan klassificeras i enlighet med olika principer inom olika områden. ”Folkvisa” kan t.ex. vara ett begrepp både inom musik och inom traditionsforskning (hypernymkedjorna i figur 8 är fiktiva):



Figur 8: ”Folkvisa” med två olika överbegrepp.

Om bägge terminologierna samlas under samma ontologiska tak måste man antingen räkna med två olika begrepp ”folkvisa” eller så måste man tillåta att ett begrepp har mer än ett överbegrepp, här ovan ”folkmusik” och ”sång”. I det danska ordnätet DanNet har denna induktiva dikotomi behandlats mycket elegant. Fastän nätet sannolikt innehåller två olika begrepp ”purjolök” ser det för användaren ut som ett enda begrepp med två olika överbegrepp (Pedersen 2010:170–). Jfr DanNet ”porre”.

3. Ordnät

3.1. WordNet

WordNet är ett begreppssystem sammanställt med hjälp av informationstekniska metoder. I likhet med begreppsordböckerna sammanför WordNet synonymerna i synonymgrupper (i WordNet-termer i *synset*) och kopplar dem till överordnade begrepp i hierarkin.

WordNet koncentrerar sig på ord som tillhör de öppna ordklasserna och har lexikal betydelse. Beroende på vilken helhet som ska täckas ingår vanligen bara substantiv, adjektiv, verb och adverb.

WordNet har utvecklats vid Princeton University och föreligger för närvarande i version 3.1. Ordnätet har tillämpats på många olika sätt på andra språk, främst genom översättning. Att WordNet översätts implicerar ett antagande om att det skulle handla om ett universellt system som låter sig översättas. Också EuroWordNet baserar sig på WordNet, men för EuroWordNet har man utvecklat en egen ontologisk högsta nivå, *Top Ontology*. Frågan om i vilken mån det är meningsfullt att översätta WordNet har debatterats flitigt (t.ex. Anderson m.fl. 2010, Lindén & Carlson, 2010). Resultatet av den nyligen gjorda översättningen av WordNet till finska diskuteras av Martola (2011).

Namnet till trots är det inte ord som utgör beskrivningsenheter i WordNet utan begrepp och entiteter. Eftersom ett ord i ett språk kan stå för olika begrepp, dvs. ett ord kan ha flera betydelser, leder detta till att ett och samma ord kan representera flera olika begrepp i ordnätet. Varje begrepp ingår i en egen hypernymkedja. I tryckta ordböcker måste man i sådana fall använda sig av hänvisningar. Som elektronisk källa kan WordNet göras användarvänligare tack vare klickbara länkar.

3.2. Lagercrantz begreppssystem jämfört med WordNet

I det följande ska jag visa på skillnaderna mellan begreppssystemen i Lagercrantz och i WordNet med hjälp av två exempel. Det ena är det astrakta begreppet ”skönhet”, det andra är det för samisk kultur mycket centrala begreppet ”ren”. Lagercrantz klassifierar begreppet ”skönhet” enligt figur 9.

metodologiska begrepp
 sinnesförmimmelser
 modalitet
 allmänbegrepp för modalitet och orsak
 substantiv för modalitet och orsak
 skönhet

Figur 9: Hypernymkedjan för begreppet ”skönhet” i Lagercrantz.

I WordNet ser klassificeringen ut som i figur 10⁴.

entitet
 abstraktion, abstrakt entitet
 attribut
 kvalitet
 utseende, visuell aspekt
 skönhet

Figur 10: Hypernymkedjan för begreppet ”skönhet” i WordNet.

De båda klassificeringarna har lika många nivåer men kategorierna ser olika ut. Vissa kategorier är dock jämförbara. Kategorierna ”sinnesförmimmelser” hos Lagercrantz och ”utseende, visuell

4 WordNets hierarki presenteras här vänd upp och ner i jämförelse med WordNets egen presentation, för att de båda systemen ska vara lättare att jämföra.

aspekt” i WordNet motsvarar varandra i stort sett, men i de hierarkiska systemen befinner de sig i motsatta ändar: hos Lagercrantz högt uppe och i WordNet lågt nere som närmast överordnade begrepp till ”skönhet”.

Tvärtom förhåller det sig med den nivå som hänför sig till grammatiska lösningar. I Lagercrantz särskiljs de olika ordklasserna först i nedre ändan av hierarkin medan WordNet gör indelningen högst uppe. Begreppet ”entitet” hänför sig uttryckligen till substantiv.

Begreppssystemen i Lagercrantz och WordNet verkar skilja sig radikalt från varandra. I vissa fall blir över- och underbegrepp närmast diametralt motsatta. Hur är det möjligt? WordNet ska ju vara en allmängiltig och universell beskrivning. En viktig orsak till skillnaderna är hur begreppssystemen kommit till. Lagercrantz utgick från en samling begrepp som han grupperade och kategoriserade. Han sammanställde alltså systemet nerifrån, induktivt. WordNet däremot har skapats uppifrån och ner, alltså deduktivt.

I båda systemen ingår en grammatisk klassificering: i WordNet är entiteterna alltid substantiv, medan Lagercrantz särskiljer ordklasser först på lägsta nivå. På samma nivå som klassen ”substantiv för modalitet och orsak” ligger klasserna ”adjektiv för modalitet och orsak”, ”adverb för modalitet och orsak” och ”verb för modalitet och orsak”. Jämförelsen mellan systemen visar alltså, att de grammatiska kriterierna kan användas på diametralt motsatta ställen i hierarkierna.

Intressant är också att WordNets ”utseende, visuell aspekt” är ett underbegrepp medan Lagercrantz ”sinnesförmimmelser” är ett överbegrepp för det generella begreppet ”modalitet och orsak”.

Vad är då utfallet om vi jämför de två systemen med avseende på ett helt annat begrepp, ett kulturbegrepp som är centralt för samerna, nämligen djuret ren? Hierarkin i Lagercrantz ses i figur 11, WordNets mycket djupare begreppsserie⁵ framgår av figur 12.

5 Jag har gallrat bort all onödig information och vänt upp och ner på serien så att den går uppifrån och ner i likhet med Lagercrantz hierarki.

naturbegrepp
 levande natur
 animaliskt liv
 människor och djur
 djur; husdjur
 ren, allmänt

Figur 11: Hypernymkedjan för begreppet ”ren” i Lagercrantz.

entitet
 fysisk entitet
 objekt, föremål
 helhet, enhet
 levande varelse
 organism
 djur
 ryggsängsdjur, kordater
 ryggradsdjur, kraniedjur, kranier, vertebrater
 däggdjur
 placentadäggdjur
 hovdjur, ungulater
 partåiga hovdjur, klövdjur
 idisslare
 hjortdjur, hjortar
 ren

Figur 12: Hypernymkedjan för begreppet ”ren” i WordNet.

Där Lagercrantz klarar sig med sex nivåer behöver alltså WordNet hela sexton. Särskilt många gemensamma begrepp innehåller klassificeringarna inte. Gemensamt är självfallet det begrepp som ska klassificeras, eller ”ren”. Bland de överordnade kategorierna är det bara två som är desamma: ”djurriket” och ”levande natur/varelse”.

Utifrån WordNets klassificering vet vi med största exakthet vad en ren är rent naturvetenskapligt. Utifrån Lagercrantz klassificering vet vi något helt annat. Under begreppet ”ren, allmänt” anför Lagercrantz en synonymgrupp med sammanlagt 24 uttryck som syftar på renar av olika slag, till exempel ”avskilt levande ren”, ”utvilad ren (som kan utnyttjas som dragdjur)”, ”ren med benägenhet att dra åt ena sidan”. Se också avsnitt 2.1 figur 7.

3.3. Är WordNet en ontologi?

I hypernymkedjorna för ”tomato” i WordNet (figur 13) ser vi att den vänstra, taxonomiska serien har två noder mera än den högra, icke-vetenskapliga serien (0–9 resp. 0–7). Den ontologiska klassificeringen i entiteter börjar uppifrån (0–). Den taxonomiska klassificeringen i taxoner börjar nerifrån (9–) liksom den intuitiva klassificeringen i begrepp (7–). I den vänstra kedjan utgör noderna 6–9 den biologiska taxonomin medan den översta delen (0–6) bildar ontologin. Nod 6 eller entiteten och taxonen ”växt” utgör länken mellan den ontologiska och den taxonomiska delen. Den högra serien är likartad. Intuitiva vardagsbegrepp är grönsak (nod 5), jordbruksprodukter (4) och livsmedel (3). Däremot är det snarast en taxonomisk uppgift att tomat hör till kärlväxterna (6). Den ontologiska delen av den högra hypernymkedjan består av noderna 0–3, och nod 3 ”föda” utgör länken mellan de ontologiska entiteterna och de språkbundna begreppen.

<i>taxonomiska</i>	<i>icke-vetenskapliga</i>	
<i>hypernymkedjans topp</i>	<i>hypernymkedjans topp</i>	
0 entity	entity	0
1 physical entity	physical entity	1
2 object, physical object	matter	2
3 whole, unit	food, solid, food	3
4 living thing, animate thing	produce, green goods, green	
5 organism, being	groceries, garden truck	4
6 plant, flora, plant life	vegetable, veggie, veg	5
7 vascular plant, tracheophyte	solanaceous	
8 herb, herbaceous plant	vegetable	6
9 tomato	tomato	7

Figur 13: Två (förkortade) hypernymkedjor för begreppen 'tomat' i WordNet.

Utifrån det ovan sagda kan vi konstatera att den översta nivån i WordNet:s hierarki verkligen är ontologisk och baserar sig på logiska entiteter. Nerifrån räknat består hypernymkedjorna däremot snarast av språk- och kulturbegrepp. Den ontologiska delen kan utan vidare översättas till andra språk, eftersom det som ska översättas är benämningar på logiska entiteter vilkas innehåll inte ändras vid översättning. Den begreppsliga delen av WordNet är däremot språkbunden, och att översätta den ord för ord till ett annat språk leder inte till ett rimligt resultat. Orden i WordNet är inte heller "ord" utan de är begrepp, och översättningen borde därför ske begrepp för begrepp, inte ord för ord. Men det är en omöjlighet, eftersom begreppssystemen i olika språk är olika.

För EuroWordNet har ontologin (Top Ontology) från början utformats som en självständig del, som inte är beroende av de enskilda språken. Begreppen i de ingående språken är kopplade till varandra via den språkoberoende begreppsapparat som i sin tur är kopplad till ontologin. Ett språkoberoende begreppssystem är i ljuset av det ovan sagda en självmotsägelse, och att försöka bygga

upp den delen av EuroWordNet har inneburit stora problem (se Rodríguez 1998; Vossen 1998).

4. Sammanfattning

I denna artikel har jag genom exempel försökt klargöra hur moderna ontologier förhåller sig till WordNet och begreppsordböcker. Begreppsordböcker och begrepps-system har en flerhundraårig historia medan ontologierna trampar i barnskorna. Man kan i alla fall tydligt se att det informationstekniska arbetet bär frukt. Det främjar inte bara datorernas förmåga att hantera semantik – något som möjliggör mer intelligenta webbsökningar och extrahering av viktig information ur stora textmassor – utan det berikar och befrämjar också språkforskning, i synnerhet, semantik, lexikologi och lexikografi.

Det var under upplysningstiden som begreppet encyklopedisk vetenskap uppkom. I encyklopediska publikationer hade man som strävan att samla all information om hela världen. I och med att kunskapsmängden ökar enormt vartefter den vetenskapliga forskningen framskrider har sådana föresatser sedermera visat sig vara omöjliga att uppnå, hur omfattande en redaktion än är. Möjligen kan det visa sig att de nuvarande ansträngningarna att få datorer att via en universell ontologi förstå all existerande och framtida kunskap om världen blir mer framgångsrika.

Litteratur

Ordböcker

Dornseiff, F. 1933–1940: *Der deutsche Wortschatz nach Sachgruppen*. Berlin. [Åtskilliga senare upplagor.]

- Nagy, G. 1978: *Magyar szinonimaszótár*. Budapest.
- Roget, P. M. 1856: *Thesaurus of English Words and Phrases, Classified and Arranged so as to Facilitate the Expression of Ideas and Assist in Literary Composition*. London. [Åtskilliga senare upplagor.]
- Wehrle, H. 1942: *Deutscher Wortschatz. Ein Wegweiser zum treffenden Ausdruck*. Stuttgart. [Åtskilliga senare upplagor.]

Annan litteratur

- Anderson, W., Pretorius, L., Kotzé, A. 2010: Base Concepts in the African Languages Compared to Upper Ontologies and the WordNet Top Ontology. I: *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*, 3757–3764. www.lrec-conf.org/proceedings/lrec2010/pdf/247_Paper.pdf
- Christiansson, J., Zimmerman, Z. 2006: *Word sense disambiguation med Svenskt OrdNät*. Magisteruppsats i biblioteks- och informationsvetenskap vid Biblioteks- och Informationsvetenskap/Bibliotekshögskolan 2006:34. Högskolan i Borås. bada.hb.se/bitstream/2320/1449/1/06-34.pdf
- Gruber, T. R. 1993: A translation approach to portable ontology specification. I: *Knowledge Acquisition* 5,2, 199–220.
- Lagercrantz, E. 1939: *Lappischer Wortschatz*. I–II. Lexica Societatis Fenno-Ugricae VI. Helsinki.
- Lindén, K., Carlson, L. 2010: FinnWordNet – WordNet på finska via översättning. I: *LexicoNordica* 17, 119–140.
- Linné, C. von 1753: *Species plantarum*. Stockholm.
- Lorhard, J. (Jacobus Lorhardus) 1606: *Ogdoas Scholastica, continens Diagraphen Typicam artium: Grammatices (Latinae, Graecae), Logices, Rhetorices, Astronomices, Ethices, Physices, Metaphysices, seu Ontologiae, Sangalli*. Apud Georgium Straub.

- Martola, N. 2011: FinnWordNet och kulturbundna ord. I: *LexicoNordica 18* (i denna volym).
- Niles, I., Pease, A. 2001: Towards a Standard Upper Ontology. I: *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001)*, Chris Welty and Barry Smith, eds, Ogunquit, Maine, October 17-19, 2001. home.earthlink.net/~adampease/professional/FOIS.pdf
- Niles, I., Pease, A. 2003: Linking Lexicons and Ontologies: Mapping WordNet to the Suggested Upper Merged Ontology. I: *Proceedings of the IEEE International Conference on Information and Knowledge Engineering (IKE 2003)*, s. p. home.earthlink.net/~adampease/professional/Niles-IKE.pdf
- Pedersen, B.S. 2010: Semantiske sprogressourcer – mellem sprogteknologi og leksikografi. I: *LexicoNordica 17*, 163–180.
- Rodríguez, H. m.fl. 1998: The Top-Down Strategy for Building EuroWordNet: Vocabulary Coverage, Base Concepts and Top Ontology. I: *Computers and the Humanities* 32, 117–152.
- Viberg, Å. 2002: Svenskt OrdNät. I: *Nordisk språkteknologi 2002 – Nordic Language Technology 2002*, 135–144. www2.lingfil.uu.se/personal/viberg/VibergNordSprktekPDF.pdf
- Vossen, P. m.fl. 1998: *The EuroWordNet Base Concepts and Top Ontology*. Version 2, Final. January 22, 1998. Contributors: Piek Vossen, Laura Bloksma, Horacio Rodriguez, Salvador Climent, Nicoletta Calzolari, Adriana Roventini, Francesca Bertagna, Antonietta Alonge, Wim Peters. Deliverable D017, D034, D036, WP5. EuroWordNet, LE2-4003. www.vossen.info/docs/1998/D017.pdf

Internethänvisningar (januari 2011)

- DanNet = DanNet – Det danske wordnet. wordnet.dk/dannet
 DanNet ”porre” = porre – andreord.dk/ord/8057-porre

FinnONTO = National Semantic Web Ontology Project in Finland. www.seco.tkk.fi/projects/finnonto/
Index of WordNetMappings = Index of /KBs/WordNetMappings. sigmakee.cvs.sourceforge.net/sigmakee/KBs/WordNetMappings/
KulttuuriSampo = KulttuuriSampo – suomalainen kulttuuri semanttisessa web 2.0:ssa. www.kulttuurisampo.fi/
SUMO = Suggested Upper Merged Ontology (SUMO) – Ontology Portal. www.ontologyportal.org
SUMOclasses = SUMOclasses.gif. www.ontologyportal.org/images/SUMOclasses.gif
SUMOFAQ = Ontologt Portal – FAQ. www.ontologyportal.org/FAQ.html
WordNet = About WordNet – WordNet – About WordNet. wordnet.princeton.edu/
YSO = ONKI3 | Finnish General Upper Ontology – YSO/ALLSO. www.yso.fi/onto/yso/

Klaas Ruppel
forskare, föreståndare för ordboksavdelningen
Forskningscentralen för de inhemska språken
Berggatan 24
FI-00100 Helsingfors
klaas.ruppel@kotus.fi

Översättning från finska: Nina Martola

Dansk begrebsordbog: baggrund og metode

Lars Trap-Jensen

A Danish thesaurus project is currently underway at the Society for Danish Language and Literature in Copenhagen. Funded by The Carlsberg Foundation, the project runs 2010-2013 with a staff of two full-time editors. It does not start from scratch but draws on already existing lexical data: main input for the thesaurus will be the semantic descriptions of The Danish Dictionary (DDO, 116,000 senses) and the Danish WordNet (DanNet, 65,000 synsets), the latter resource itself built on DDO data. Following an initial analysis of existing thesauri, a macro-structure was adopted close to the one used by Franz Dornseiff: *Der deutsche Wortschatz nach Sachgruppen* (2004). As a side benefit, this choice makes it easier for the editors to utilize the contents of the only existing thesaurus for Danish, Harry Andersen's *Dansk Begrebsordbog* (1945), which itself is no less structurally dependent on the first edition of Dornseiff (1934). One methodological aspect is given particular attention: what should be the main ordering principle within an entry? A final decision has yet to be taken as it is primarily a problem of presentation with no immediate bearing on the editing process, but eventually a conclusion must be reached as the thesaurus is going to appear as a print dictionary, too.

1. Baggrund

På Den Danske Ordbog gjorde vi for snart en del år siden den samme erfaring som alle andre der arbejder med elektroniske ordbøger, nemlig at ophævelsen af papirets alfabetiske ordningsprincip åbner for andre måder at systematisere materialet og skabe nye forbindelser mellem ord og betydninger som i forvejen er indeholdt i ordbogen. Én følge af denne erfaring var udarbejdelsen af

et dansk wordnet i samarbejde med Center for Sprogteknologi, DanNet (jf. www.wordnet.dk, Pedersen m.fl. 2009). Undervejs i arbejdet med DDO havde vi gjort os nogle – på daværende tidspunkt ikke helt klare – forestillinger om at data kunne udnyttes på den måde, og vi havde derfor indført et felt i ordbogsstrukturen til angivelse af nærmeste overbegreb. Elementet var ikke synligt i den trykte ordbog, men kom os til gode ved etableringen af DanNet fordi det blev en genvej til at etablere hierarkiet af over- og underbegreber, som jo er en central del af et wordnet. Et andet ikke umiddelbart synligt element der har kunnet udnyttes, er et felt hvori redaktørerne skulle angive en betydnings tilknytning til et fagområde hvis det overhovedet var muligt – til forskel fra den mere snævre angivelse af fagbetegnelse. Dette har gjort det muligt at knytte semantiske domæner til mange betydninger i DanNet.

Et punkt hvor ordbogens oplysninger viste sig mindre hensigtsmæssige for DanNet, var i dens – i hvert fald i computersammenhæng – utilstrækkelige systematiske konsekvens. Det er som bekendt helt nødvendigt at være både eksplicit og konsistent i angivelse af de relationer der udvælges i beskrivelsen, når resursen efterfølgende skal kunne bruges af maskiner og i softwareapplikationer. Derfor er de eksisterende data blevet beriget i forhold til udgangspunktet så der ikke opstår tilfældige huller i kodningen.

I slutfasen af DanNet-projektet har vi forsøgt at udnytte disse data også for menneskelige brugere. I Den Danske Ordbogs netudgave er det sket med oplysningstypen “Beslægtede ord”. Figur 1 viser hvilke beslægtede ord der er registreret for *hue* i Den Danske Ordbog. Som hovedregel vises der tre typer af beslægtede ord, hentet fra DanNet: overbegreber (mere generelle ord), underbegreber (mere specifikke ord) og søsterbegreber (ord der deler samme overbegreb, kohyponymer). Der er også mulighed for at se de beslægtede ord i en grafisk visning når man klikker på “vis som grafik (eksternt link)”. Resultatet kan ses i figur 2 (læs nærmere om denne funktion i Johannsen og Pedersen 2011).

BESLÆGTEDE ORD^{HUE} mere generelt: hovedtøj | hovedbeklædning | mere specifikt:
 elefanthue | lue | skråhue | kippa | baskerhue | kolkehue | skindhue | tophue |
 baret | alpehue | narrehue | beret | frygisk hue | kalot | nathue | nissehue |
 garderhue | strikhue | matroshue | pelshue | andre ord med "hovedtøj" -
 hovedbeklædning som overbegreb: hætte | hat | kyse | kappe | turban | kasket |
 bjerneskindshue | tiara | hjelm | sovsekande | bispehue | mitra | ørevarmer ...
 vis som grafik (ekstern link) 

Figur 1: Beslægtede ord for hue i Den Danske Ordbog.

Den visuelle præsentation er dynamisk derved at man kan klikke på hvert af de beslægtede ord, hvorved figuren positionerer sig med det valgte ord som centrum og de dertil hørende relationer. Det fungerer ganske godt for nogle betydningsområder, især dem der opfører sig taksonomisk ordentligt, eksempelvis dette område hvor det drejer sig om konkrete genstande. I andre tilfælde fungerer det mindre godt. Det gør det fx ikke i tilfælde hvor hierarkiet har uoverskueligt mange medlemmer på hvert niveau. Den situation er ikke usædvanlig.



Figur 2: Visuel fremstilling af oplysningerne i figur 1.

I et tilfælde som *filminstruktør* nævnes der stort set kun søsterbegreber til filminstruktør, dvs. ord der alle betegner personer, og

dem er der for mange af til at det er overskueligt og brugbart for en menneskelig bruger. For at hjælpe brugerne en smule har vi derfor prøvet at skabe en ikke-tilfældig rækkefølge således at de formodet bedst egnede beslægtede ord vises først. Det er sket på grundlag af en score som beregnes efter hvor mange relationer der ud over overbegrebet binder to betydninger sammen, fx om de tilhører det samme fagområde, om de peger på den samme betydning uden selv at være forbundne osv.

BESLÆGTETE ORD¹²³ mere generelt: individ | person | hoved | menneske | mand | m/l'er | andre ord med "individ - person - hoved - menneske - mand - m/l'er" som overbegreb: operatør | stuntman | filmklipper | dekoratør | animator | filmoperatør | manuskriptforfatter | drejebogsforfatter | biografdirektør | klipper | biografefjer | biografgænger | instruktør | bedemand | juror | auktionarius | arbejdsformidler | afdelingschef | impresario | administrator | forsikringsmægler | frimærkehandler | kapelmester | orkesterleder | dirigent | børnehavelsesleder | udråber | forligsmand | journalist | skriverkarl | underviser | balletchef | belysningsmester | korrespondent | tasteoperatør | pædagog | logopæd | lærer | afdelingsleder | gæsteforelæser ...vis 140 flere

Figur 3: Beslægtede ord for filminstruktør i Den Danske Ordbog.

Det gør at personbetegnelser inden for domænet 'film' nævnes først (*operatør, stuntman, animator, filmoperatør, filmklipper, dekoratør, manuskriptforfatter, drejebogsforfatter, biografdirektør, klipper, biografgænger ... osv.*), hvad der vel hjælper lidt, men alligevel ikke er helt overbevisende.

Som det er fremgået, er over-underbegrebs-relationen den vigtigste når der skal findes beslægtede ord, og det er den væsentligste grund til at systemet fungerer bedst for konkrete substantiver. Hvis vi ser på en anden begrebstype, abstrakter, er det tydeligt at taksonomien er mindre veludbygget. Der er ikke så meget at sige taksonomisk om abstrakte handlinger: De har ét endnu mere abstrakt overbegreb 'handling' og derudover blot en mængde søsterbegreber som ikke behøver have alverden tilfælles i øvrigt. Et eksempel som *vold* resulterer således i:

ledelse, bevarelse, svinkeærinde, kunstgreb, gengæld, gestus, indledning, ridt, joke, tjeneste, fejltrin, fejltagelse, snuptag, slyngelstreg, skurkestreg, modtræk, gunstbevisning, fejlgreb,

misgreb, hævnakt, hævn, opfølgning, followup, dækmanøvre, beslutning, afgørelse, tyveri, behandling, gemenhed, akt, uanstændighed, indgreb, overraskelsesangreb, idræt, fedteri, operation, idioti, venlighed, judaskys, forbrydelse, ugerning, hug, modsvar, aktion, afværgemanøvre, rævestreg, galmandsværk, angrebsbølge, nederdrægtighed, camouflaje, afledningsmanøvre, forvandlingsnummer, harakiri, sabotage, hyldest, ceremoniel, fingerknips, genistreg, ... (reduceret fra 205 forekomster)

De fleste vil nok bekræfte at man i dette tilfælde ville foretrække dels færre ord, dels ord fra samme semantiske felt, fx *magt, magt-anvendelse, gevalt, herredømme, overgreb, overlast, voldshandling, voldtægt, voldtægtsforbrydelse, voldførelse, brutalitet*. En indikator for at de anførte beslægtede ord kun er svagt beslægtede, ses af at den omtalte score ikke er særlig høj: 10 for de førstnævnte søsterbegreber, mens den er 110 for de første søsterbegreber nævnt under *hue*. Det samme problem gør sig gældende inden for ordklasserne verber og adjektiver: På den ene side får man for mange beslægtede ord, og i andre tilfælde mistænker man at relevante begreber ikke kommer med. Det skyldes formentlig dels at et wordnet og en begrebsordbog ikke har de samme primære anvendelsesmuligheder, dels at resurserne i DanNet ikke har rakt til den nødvendige manuelle finjustering af de semantiske hierarkier. En kritik man ikke så sjældent hører fremført mod wordnet, er at de er for fine i deres semantiske beskrivelser. I sprogteknologiske applikationer vil det ofte være en fordel hvis der om substantiver blot oplyses fx om det er en person, og hvilken ontologisk type det tilhører. I DanNet har vi gjort en pointe ud af kun at operere med mellemniveauer hvis der findes ord i sproget for dette niveau. Pointen er at wordnettet skal afbilde det sproglige inventar i dansk og ikke være en begrebsontologi der står uden for sproget, sådan som man kender det fra fx faglige terminologier.

En anden – og beslægtet – forskel på wordnets og begrebsordbøger er at det i wordnetverdenen hører til sjældenhederne at en betydning indgår i flere begrebskategorier på samme tid. Det skal ikke forstås sådan at det er forbudt; det er det på ingen måde, og i DanNet forekommer det da også at et synset har flere overbegreber. Ofte er det fremkaldt af at der i en definition i DDO har været nævnt flere overbegreber. Fx defineres *datter* i DDO som en “pige eller kvinde som man er far eller mor til”, og derfor er *datter* blevet tilordnet både overbegrebet ‘pige’ og overbegrebet ‘kvinde’. Det kan også være at DanNet-redaktøren har besluttet at en betydning tilhører to overbegreber af systematiske grunde. Det er sket i tilfældet *blomkål*, som dels er placeret under ‘planter’, dels under ‘grøntsager’. Sproget indeholder mange af den slags eksempler uden at det i alle tilfælde er blevet udskilt som selvstændige betydninger i DDO. Men selvom det altså er tilladt at kode flere overbegreber for samme synset, er det ikke noget der er taget i anvendelse specielt ofte. I den nuværende version af DanNet er der lige under 500 tilfælde af synsets med flere overbegreber – ud af et samlet antal på 65.000 synsets.

Hvis man derimod slår op i registerdelen i en begrebsordbog, vil man hurtigt opdage at det er reglen snarere end undtagelsen at et opslag er placeret i flere kapitler. I den forbindelse skal man naturligvis huske at tage højde for at registeret ikke kan skelne mellem form og indhold, men kun har ordets betegnelse til rådighed og derfor ikke sondrer mellem homografer og forskellige ordbetydninger. Det er derfor begrebet ‘synset’ er vigtigt i wordnetverdenen, mens der ikke findes en pendant i almensproget. Men selv hvis man tager højde for den forskel, er det almindeligt i begrebsordbøger at samme ord i samme betydning er placeret i flere forskellige kapitler i begrebsordbogen – som en afspejling af at ord og betydninger normalt ikke entydigt kan tilordnes ét og kun ét betydningsområde. Det er den væsentligste årsag til at begrebsordbøgers indekser indeholder færre ord end der optræder

i kapitlernes systematiske del (jf. opgørelserne i Nimb 2011).

Vi kan se på et enkelt eksempel: I DDO er der bare én betydning af *køb* som verbalsubstantiv ‘handlingen at købe’ (foruden en konkret betydning: ‘det man køber’, dvs. = vare). I DanNet er der angivet ét overbegreb (‘anskaffelse’) og ét underbegreb (‘import’), mens Harry Andersen 1945 anfører *køb* i tre forskellige indgange: 28 *Ombytning*, 808 *Handel* og 811 *Køb*. Det er den samme betydning der optræder alle tre steder, men der er nuanceforskelle alt efter om man har fokus på købet som et bytteforhold, på handel i almindelighed eller på den specifikke købsituation. Det er der ikke noget usædvanligt i; det er snarest den almindelige situation når menneskelige brugere griber til en begrebsordbog: Vi kan opfatte den samme betydning som hørende til forskellige områder, alt efter hvilken synsvinkel vi anlægger. Og det er også derfor jeg tidligere har karakteriseret det som problematisk når en betydning kun tilordnes én kategori i en tesaurus (jf. fx Trap-Jensen 2010:300ff.). Hvis man fx slår *guinea pig* op i Macmillans netordbog, optræder marsvinet kun ét sted, nemlig i en kategori der benævnes: ‘Animals found on more than one continent’, og det er langt fra sikkert at det er den kategori brugeren er interesseret i. Lige så sandsynligt er det at brugeren interesserer sig for ‘kæledyr’, ‘gnavere’ eller måske ‘dyr med pels’, ‘små pattedyr’ eller ligefrem ‘dyr der kan spises’. Pointen er at det er meningsfuldt at placere marsvinet i flere forskellige kategorier – der er ikke altid kun én relevant kategori for en bestemt betydning.

En vigtig opgave består derfor i at forbedre funktionen Beslægtede ord ved at oprette flere meningsfulde kategorier med et overskueligt antal relevante medlemmer i hver. Det var ikke mindst dette der motiverede os til at gå i gang med en egentlig begrebsordbog.

2. Metode

2.1. Makrostruktur

I sammenligning med andre begrebsordbøger er vores projekt et discountprojekt: 3 år med en bemanning på 2 redaktører er ikke meget i forhold til de livstidsprojekter som de fleste forgængere har været. Til gengæld begynder vi ikke fra bunden, men har et omfattende og systematiseret materiale som udgangspunkt. I en vis forstand drejer det sig om finpudsning af det eksisterende materiale og en præsentation af det i en form der er hensigtsmæssig for ordbogsbrugere.

Ét sted hvor vi har valgt en nem løsning, er med hensyn til makrostrukturen. I stedet for at gentænke en konceptuel opdeling af verden, har vi valgt at overtage en eksisterende. Mest oplagt ville det nok være at bruge Rogets opdeling i ca. 1000 (2002-udgaven: 990) grupper. Ikke alene er den med over 40 millioner solgte eksemplarer blandt de 3 mest sælgende ordbøger i historien (ifølge Wikipedia kun overgået af Webster's American Spelling Book (100 mio.) og Merriam-Webster's Collegiate Dictionary, 55 mio.), men vigtigere endnu er den pga. sin store kommercielle succes kommet i et utal af nye udgaver og revisioner i løbet af de over 150 år der er gået siden den udkom første gang. Af samme grund har Roget været forbillede og udgangspunkt for de fleste begrebsordbøger der siden er udkommet. I dansk leksikografi finder vi klassikeren Harry Andersens *Dansk Begrebsordbog* fra 1945, og den lægger sig igen ganske tæt op ad den tyske klassiker, Franz Dornseiffs *Der deutsche Wortschatz nach Sachgruppen*. På tidspunktet for udgivelsen af Andersen 1945 var Dornseiff en forholdsvis ny udgivelse, idet den udkom første gang i 1934. Dornseiff er også inspireret af Roget, men den kopierer ikke strukturen fra Rogets. I stedet opererer den med 22 større kapitler der igen er inddelt i 906 tematiske grupper. Det mærkes at Dornseiff er udkommet over 80 år senere

end Roget: Den er ikke på samme måde forankret i romantikkens opfattelse af sprogets organiske sammenhæng og traditionen fra Leibniz og bagud til Aristoteles. Dornseiff er for tysk nærmest blevet hvad Roget er for engelsk, og den er på tilsvarende vis kommet i mange senere revisioner, senest i en 8. udgave i 2004 med cd-rom.

I Det Danske Sprog- og Litteraturselskab er vi nået til samme konklusion som Harry Andersen og har besluttet at lægge os op ad makrostrukturen hos Dornseiff. Dens opdeling af verden virker mere klar og mere intuitivt gennemskuelig end Rogets ældre inddeling. Det har den yderligere fordel at det er nemt for os at supplere med relevant materiale fra Harry Andersens begrebsordbog.

I stedet for at genopfinde den dybe tallerken kopierer vi derfor som udgangspunkt makrostrukturen fra Dornseiff, hvorefter alt indhold importeres fra DanNet og DDO, jf. Nimb (2011). Derved får vi også hele mediostrukturen forærende. En vigtig del af en begrebsordbog er jo – som vi så før – at der dels oplyses i indekset om alle de relevante grupper et ord står i, dels at der henvises fra én tematisk gruppe til andre betydningsbeslægtede grupper fordi samme ord eller betydning kan optræde i forskellige grupper alt efter hvilken synsvinkel det betragtes under.

Vi indrømmer gerne at det er en discountløsning. De uhenigtsmæssigheder der måtte være i Dornseiff og Harry Andersen, importeres automatisk til vores begrebsordbog. Den risiko følger med, men det samme gælder naturligvis også den sikkerhed der ligger i de revisioner og tilføjelser der løbende er foretaget frem til 8. udgave, som er vores udgangspunkt. Og man er nødt til at sammenholde det med de resurser der skulle have været investeret i at koncipere en alternativ struktur. Vi tror ikke det ville føre til et væsentlig bedre resultat end det foreliggende i Dornseiffs 8. udgave. Og det ville være ærgerligt at bruge mange resurser på at nå frem til et resultat der måske ikke ville afvige meget fra det foreliggende. Så det har egentlig været en nem beslutning at tage.

Løsningen fører selvfølgelig det væsentlige spørgsmål med sig:

I hvilken udstrækning lader den verdensopdeling som Dornseiff bruger, sig overføre fra et sprog til et andet? Kan man være sikker på at de grupper der byder sig til på tysk, også gør det på dansk? Nej, det kan man ikke; begreber er naturligvis ikke kongruente på tværs af sprog. Uden at gå helt tilbage til Sapir-Whorf og sammenhængen mellem sprog og tanke turde det være klart at der er både sproglige og kulturelle forskelle som man skal være opmærksomme på. Vi kan se på et par eksempler.

Dornseiffs gruppe 17.24 hedder *Spiele*. Det er formentlig uproblematisk på tysk at grupperne *Handpuppe*, *Eisenbahn* og *Schaukelpferd* i samme gruppe som *Monopoly*, *Schach* og *Bridge* fordi man på tysk har ét fælles begreb og ikke skelner – som vi gør på dansk – mellem ‘spil’ og ‘leg’.

Det kan også gå den anden vej: at der er to tyske begreber, hvor vi på dansk kun vil have ét. Og endelig er der alle de kulturbetingede forskelle mellem en dansk og en tysk virkelighed som slår igennem i begrebsopdelingen: Dornseiff har fx et selvstændigt afsnit om industri og har tillige udskilt to særlige områder, bilindustri og medicinalindustri, i selvstændige afsnit. Her ville det i en dansk sammenhæng være mindst lige så oplagt at udskille fx fødevarerindustri eller turisme. Og Dornseiff har et stort afsnit 18.10 *Verwaltungsbehörde* ved siden af et andet afsnit 18.11 om *Vertretungsausschuss*, *Ratsversammlung*. De mange begreber der afspejler tysk samfundsstruktur på forbunds- og delstatsniveau, har ikke nødvendigvis sidestykke i en dansk sammenhæng.

Dansk og tysk er altså ikke kongruente i deres konceptualisering af verden. Alligevel er det vores vurdering at det samlet set drejer sig om en beskeden del af begrebsgrupperne. Det bliver derfor nødvendigt at foretage visse redaktionelle tilpasninger i makrostrukturen efterhånden som behovet opstår: Nogle grupper skal slås sammen, og andre opdeles i flere. Redaktørerne skal i den forbindelse være opmærksomme på de ændringer i henvisningssystemet det medfører.

2.2. Mikrostruktur

En enkelt beslutning har redaktionsinternt voldt særlige problemer. Det drejer sig om præsentationen af data på det øverste niveau i mikrostrukturen.

Der lader til at være enighed mellem de fleste eksisterende begrebsordbøger om følgende prioritering af kriterier for ordning af mikrostrukturen:

1. ordklassebestemt
2. semantisk: afsnit – gruppe – undergruppe
3. alfabetisk

Hos Roget er en artikel først opdelt i ordklasser: Substantiv, adjektiv, verbum og interjektion. Hver ordklasse er underinddelt i mindre grupper bestemt af betydningsnuance og placeret i selvstændige afsnit indledt af et nøgleord i kursiv. Internt i gruppen er inventaret ordnet i mindre grupper adskilt af semikolon og med begreberne anført i logisk rækkefølge. I slutningen af sådan en semikolongruppe kan der være henvisninger til andre nøgleord, anført i kursiv og med det pågældende nummererede afsnit.

Hos Dornseiff anvendes en mikrostruktur der minder om Rogets, men alligevel adskiller sig. Henvisninger optræder ikke inde i de enkelte grupper, men er placeret indledningsvis i en selvstændig rubrik. Ordklasseopdelingen er ikke eksplicit, men er dog også her det primære kriterium for inddelingen i mikrostrukturen, med den forskel at interjektioner kommer først, derefter substantiver, adjektiver/adverbier og til sidst verber. Ved ikke at have eksplicitte ordklasseangivelser undgår Dornseiff klassifikationsproblemer inden for systemet, fx når det drejer sig om at placere præpositionsforbindelser, talemåder, rutineformler osv. Til gengæld er det sværere for brugerne at orientere sig i det enkelte afsnit.

Inden for den enkelte gruppe i Dornseiff er der plads til be-

tydningsnuancer. De er adskilt af en større firkant, mens de enkelte medlemmer i undergrupperne adskilles af en midtstillet prik. Grupper adskilt af store firkanter svarer til Rogets paragraffer. Også hos Dornseiff har det første ord i hver af de større grupper karakter af nøgleord, blot er det ikke kursiveret som hos Roget. Dets status som nøgleord viser sig ved at det står uden for den alfabetiske rækkefølge som ellers er det ordnende princip internt i hver undergruppe.

Harry Andersens artikelstruktur minder i sagens natur mest om Dornseiffs, dog er brugen af henvisninger i en artikel mindre hyppig, og hvis der bruges henvisninger, kan de stå både på artikelniveau og i de enkelte grupper. Ordklasseopdelingen er gjort eksplicit, men der forekommer kun tre ordklasser, benævnt S., A. og V. Mens S. for substantiver og V. for verber er forholdsvis uproblematisk, fungerer A. som en restkategori. Den dækker over adjektiver, adverbier, pronominer, præpositioner, konjunktioner, interjektioner og talord – foruden fraselignende enheder: “til at brække sig over”, “Gud være lovet” osv. De enkelte undergrupper svarer i øvrigt til Dornseiffs, men adskilles af en tankestreg, og Harry Andersen bruger ikke nøgleord, men alene alfabetisk ordning af begreberne i hver undergruppe.

Selvom der altså er mindre forskelle i fremstillingen i de tre værker, er der enighed om det centrale spørgsmål: Ordklasserne er det vigtigste ordningsprincip. Man må derfor spørge: Er det nu så selvfølgeligt? Derved kommer man jo til at adskille ord der står hinanden semantisk nær, blot fordi de tilhører forskellige ordklasser.

Svaret på det spørgsmål har ifølge Wiegand (2004) med konsultationssituationen at gøre. Han skriver herom:

Bei der oben formulierten Charakterisierung des genuinen Zwecks von großen einsprachigen Sachgruppenwörterbüchern wurden bei den Möglichkeiten, die Wörterbücher

dieses Typs dem Benutzer eröffnen, drei verschiedene Arten [unterschieden:

- (i) Konsultationssituation wegen Ausdrucksfindungsschwierigkeiten
- (ii) Konsultationssituation wegen äquivalentbezogener Angemessenheitszweifel
- (iii) Konsultationssituation wegen Wissensbedarf über einen bestimmten Aspekt des deutschen Wortschatzes]

Die erste Möglichkeit ... wird wahrscheinlich am meisten genutzt. Das ist auch der Grund dafür, warum die Wörterbuchartikel im Formteil nach Ausdrucksklassen und damit auch nach Wortarten gegliedert sind. Die kundigen Benutzer, die Ausdrucksfindungsschwierigkeiten haben, werden durch diese artikelinterne Datenanordnung präferiert.

Wiegand, 2004: 59 (i-iii sammenfattet fra 57-58)

Jeg er ikke bekendt med undersøgelser der empirisk har registreret brugen af begrebsordbøger, men Wiegands opfattelse harmonerer godt med min egen fornemmelse: Begrebsordbøger er først og fremmest produktionsordbøger der bruges til fremfinding af paradigmatiske alternative ord og udtryk i en bestemt sproglig kontekst. Kun sekundært bruges den til at skaffe overblik over sprogets inventar inden for et bestemt semantisk felt, og det samme gælder formentlig oversættelsessammenhængen, der til nød kan opfattes som et undertilfælde af (i). Man kan naturligvis indvende at alternative formuleringsvarianter ikke altid har form af paradigmatiske ledalternativer, og at det søgte ord eller udtryk derfor heller ikke nødvendigvis tilhører samme ordklasse. Dette synspunkt har utvivlsomt noget for sig, men kan næppe tillægges afgørende vægt.

Man kan også have den opfattelse at (iii) skal prioriteres, altså at det er vigtigere at strukturen prioriterer semantikken således

at de enkelte undergrupper opregner alle ord inden for samme semantiske nuance uanset hvilken ordklasse det drejer sig om. Det har yderligere den fordel at det redigeringsteknisk er enklere eftersom det eksisterende materiale i forvejen er klassificeret efter semantiske kriterier, jf. Nimb (2011). Som det også nævnes dér, er det dog primært et præsentationsproblem som principielt er uafhængigt af redigeringsprocessen. Denne foregår først og fremmest efter semantiske principper, men alligevel kan indholdet forholdsvis enkelt præsenteres ordklasseopdelt i den færdige publikation. Det siger sig selv at valget først og fremmest er relevant når det drejer sig om en papirudgivelse. Begrebsordbogen skal dels udkomme som en traditionel ordbog på papir, dels indgå i netudgaven af Den Danske Ordbog til forbedring af den onomasiologiske komponent. Hvordan redigeringen af de enkelte kapitler finder sted, beskrives i Nimb (2011).

Litteratur

Ordbøger

- Andersen, Harry 1945: *Dansk Begrebsordbog*. København: Munksgaard.
- DDO = Ebba Hjorth, Kjeld Kristensen m.fl. 2003-2005: *Den Danske Ordbog*. København: Gyldendal og Det Danske Sprog- og Litteraturselskab. Onlineversion: ordnet.dk/ddo (februar 2011).
- Dornseiff, Franz 2004: *Der deutsche Wortschatz nach Sachgruppen*, 8. Auflage. Berlin/New York: Walter de Gruyter.
- Macmillan = Macmillan Online Dictionary. www.macmillandictionary.com
- Roget, Peter Mark: *Roget's Thesaurus*, 150th anniversary edition edited by George Davidson 2002. London: Penguin.

Anden litteratur

- Johannsen, Anders & Bolette S. Pedersen 2011: "Andre ord" – a wordnet browser for the Danish wordnet, DanNet. I: Proceedings from 18th Nordic Conference of Computational Linguistics, NODALIDA 2011, Riga, Latvia. Northern Association for Language Technology, University of Tartu, Vol. 11, 295-298.
- Nimb, Sanni 2011: Semantiske relationer i en ny dansk begrebsordbog: genbrug på tværs af ordbøger. I: *LexicoNordica 18* (i dette bind).
- Pedersen, Bolette S., S. Nimb, J. Asmussen, N. Sørensen, L. Trap-Jensen, H. Lorentzen 2009: The challenge of compiling a WordNet for Danish by reusing a monolingual dictionary. I: *Language Resources and Evaluation, Computational Linguistics Series*, Volume 43, Number 3: 269-299.
- Trap-Jensen, Lars 2010: Access to Multiple Lexical Resources at a Stroke: Integrating Dictionary, Corpus and Wordnet Data. I: Sylviane Granger, Magali Paquot (eds.): *eLexicography in the 21st Century: New Challenges, New Applications. Proceedings of eLex 2009, Louvain-la-Neuve, 22-24 October 2009*, Presses universitaires de Louvain • Cahiers du CENTAL 2010, 295-302.
- Wiegand, Herbert Ernst 2004: Lexikographisch-historische Einführung. I: Dornseiff 2004, 9-91.

Internethenvisninger

andreord.dk = www.andreord.dk ved Anders Johannsen. Viser data fra DanNet.

DanNet = Pedersen, Bolette S. m.fl. 2011: *DanNet*. www.wordnet.dk

Lars Trap-Jensen
 ledende redaktør, cand.mag., MPhil.
 Det Danske Sprog- og Litteraturselskab
 Christians Brygge 1
 DK-1219 København K
 ltj@dsl.dk

Den første norske tesaurus: Ivar Aasens *Norsk Maalbunad*

Lars S. Vikør

This article contains a discussion of the first onomasiological dictionary for Norwegian: Ivar Aasen's *Norsk Maalbunad* (published posthumously in 1925). Aasen (1813–1896) was the founder of the dialect-based “Nynorsk” variety of Norwegian, and Norway's first advanced dialectologist and lexicographer. *Norsk Maalbunad* was in fact primarily intended by Aasen as an aid in his other lexicographic work, and therefore not published by him (although completed in 1876). Its model was Roget's Thesaurus, but Aasen's choice of words and their arrangement were taken from the world of Norwegian farmers, although he also coined abstract words for concepts not covered in everyday popular speech at the time. The article analyses the structure of *Norsk Maalbunad* and discusses some of the problems Aasen encountered in the development of an adequate modern vocabulary for the new written language.

1. Innleiing

Ivar Aasen: *Norsk Maalbunad* (utgitt 1925) er den første tesaurusen som vart laga for norsk, og var lenge den einaste. Men den vart aldri ferdig frå Aasens hand, og var heller ikkje primært meint for publisering; boka vart gitt ut posthumt nesten femti år etter at Aasen hadde lagt siste hand på manuskriptet i 1876. Ivar Aasen var Noregs første store leksikograf, og i det omfattande ordboksarbeidet sitt hadde han bruk for semantiske klassifikasjonar som eit middel til å handtere det ordmaterialet han samla inn. Det var til hjelp i definisjonspraksisen hans.

I denne artikkelen tar eg for meg bakgrunnen og tilblivings-

historia til *Norsk Maalbunad*, heretter forkorta NMB, og så går eg inn på struktur og innhald i ein kulturell og historisk kontekst, med konkrete punktdrøftingar av nokre små utdrag. Eg gjer ei kort samanlikning mellom Aasen og det truleg viktigaste forbildet hans, Rogets *Thesaurus*, og drøftar resepsjonen av NMB i ettertida. Til slutt prøver eg å plassere NMB i norsk leksikografihistorie.

2. Bakgrunnen

I 1873 fullførte Ivar Aasen sitt største og viktigaste verk, *Norsk Ordbog med dansk Forklaring*, heretter forkorta NO73, der han la fram heile ordtilfanget han fann i dialektane. (Tittelen *Norsk Ordbog / Ordbok* dekker fleire verk, deriblant det planlagde tolvbandsverket *Norsk Ordbok*, som er i arbeid med tanke på fullføring i 2014, og som eg av og til også refererer til, da med forkortinga NO2014.)

Etter dette arbeidde han vidare på fleire andre prosjekt, deriblant NMB, som han fullførte i ei foreløpig form tidleg i 1876. Men så vart det ikkje meir. Han tenkte på å gå vidare og gjere arbeidet klart til publisering, men prioriterte andre oppgåver til han ikkje lenger hadde krefter til slikt arbeid. Dermed vart manuset til NMB liggande i papira hans.

Først 29 år etter Aasens død, i 1925, kom NMB ut på trykk, i ei utgåve ved Sigurd Kolsrud, professor i nynorsk og norske dialektar ved Universitetet i Oslo. Kolsrud har ei grundig innleiing om bakgrunnen for dette arbeidet, og Aasens biografar Kjell Venås og Stephen Walton, som leverte kvar sin store biografi i «Aasen-året» 1996 (hundreårsmarkeringa for Aasens død), har også utførlege framstillingar som i mykje bygger på Kolsrud, men som også omfattar korte innhaldsanalysar (Venås 1996:282–289, Walton 1996:567–579). Eit nytt opptrykk av NMB kom i 1975, men det er fotografisk etter Kolsruds utgåve og inneheld ingenting nytt.

Det første spørsmålet ein kan stille når ein skal gå inn på NMB,

er: Kva er *målbunad*? Ordet står ikkje i NO73, det vil seie at Aasen ikkje hadde funne det i dialektane, men laga det sjølv. I dag ville vi tolke det som ‘språkdrakt’. *Mål* er i alle fall ‘språk’. Men *bunad*, som i dag betyr ‘folkedrakt’, hadde før ein langt meir komplisert tydingsstruktur, blant anna kunne det bety reiskapar, verktøy, utstyr; jf. artikkelen *bunad* i NO2014 og analysen av denne artikkelen i Vikør 1997:43–50 (kortare og ganske likelydande i Vikør 2006:20–21 og Fjeld og Vikør 2008:200–201). Eg trur nok at også slike nyansar må tolkast inn i *målbunad* her.

I ordboksarbeidet sitt laga Aasen heile tida lister og oversikter over begreppssystematikk, først for dyr og planter, seinare, allereie kring 1840, også over andre begrepsfelt. Kolsrud har tatt med ei slik oversikt i starten av NMB-utgåva si: «Samling af Synonymer og Ord som i Betydningen staae nær ved hinanden», frå 1843. I denne oversikta har Aasen artiklar som dekker 26 begrepsfelt (på dansk) og nokre titals enkeltord under kvart (på norsk i ei førebels normalform han hadde sett opp). I 1856 laga han ei ny og større oversikt, også no med danske overbegrep og norske underbegrep, etter at den første utgåva av ordboka hans, *Ordbog over det norske Folkesprog*, var ferdig i 1850 og han arbeidde vidare for å supplere og auke ordmaterialet sitt med tanke på ei ny og større utgåve (det som skulle bli NO73).

Desse oversiktene var altså berre arbeidsreiskapar for Aasen i ordboksarbeidet. Dei var ei nødvendig strukturering som grunnlag for utforming av definisjonar, og ikkje minst var det viktig for han å få klarheit i dei strukturelle skilnadene mellom norsk og dansk semantikk. Ordbøkene hans var tospråklege. Oppslagsorda var på norsk, i ei rettskrivingsform Aasen sjølv utforma på grunnlag av komparasjon mellom dei dialektformene han fann i bruk, det som skulle bli grunnlaget for målforma landsmål eller nynorsk. Ekvivalentar og definisjonar var på dansk, som også var metaspråket i ordbøkene, det språket fororda i ordbøkene var skrivne på, tittel osv.

Men i tillegg til dette hadde Aasen også fått i oppdrag å lage ei dansk-norsk ordbok, og den tanken låg i bakhovudet hans heile tida. Det gjorde denne kontrastive semantiske struktureringa, norsk vs. dansk, enda viktigare. Denne dansk-norske ordboka vart likevel stadig utsett på grunn av meir presserande oppgåver. Først fleire år etter fullføringa av NMB i 1876 tok Aasen dansk-norsk-prosjektet opp igjen og skreiv eit førsteutkast, som var ferdig i 1883. Men så vart det ikkje meir, og den rimelegaste forklaringa er at syttiåringen Aasen ikkje lenger hadde igjen krefter til eit nytt stort prosjekt. Men manuskriptet vart utgitt i 2000 (Aasen 2000, heretter kalla DO).

NMB er blitt eit verk som etter Waltons utreknig omfattar bortimot 40 000 ord, og det gjeld nesten berre enkeltord, ikkje frasar. Der NO73 i prinsippet berre omfatta ord som levde i dialektane, prøvde Aasen i NMB også å innarbeide moderne kulturbegrep som dialektane ikkje hadde ord for. Det var ikkje minst med tanke på DO-prosjektet, for dansk hadde desse kulturorda, og Aasen ville utvikle norsk til å fylle same funksjonane som dansk, men utan å låne dei danske orda. Derfor måtte han ty til ein «islandsk» nyordspolitikk.

Det viktigaste forbildet for NMB var Rogets *Thesaurus*, som kom ut i 1852 og ser ut til å ha komme til Aasens kjennskap kort tid etterpå. Skilnadene mellom Roget og Aasen måtte likevel nødvendigvis bli store. Roget beskrev eit høgt utvikla kulturspråk med hundreårs skriftradisjon bak seg, mens Aasen skulle bygge opp eit nytt språk der han sjølv måtte utvikle eit moderne ordtilfang. Det viser seg med ein gong ein moderne lesar prøver å orientere seg i verket: Store delar av Aasens ordtilfang er eksperimentelt, av grunnar eg nettopp har nemnt, og ikkje utan vidare forståeleg for eit moderne publikum fordi utviklinga av språket gjekk i ei anna retning. Mange av ordformene til Aasen slo ikkje igjennom, og er dermed ukjent i dag. Ikkje minst for svenske og danske lesarar, sjølv om dei er språkekspertar, blir verket nokså ugjennomtrenge-

leg. Den som vil bruke NMB, må ha NO73 for handa, for der vil han finne definisjonane til dei orda Aasen hadde funne i dialektane – men mange av dei blir heller ikkje brukt lenger.

3. Struktura i NMB

Makrostrukturen er enklare i NMB enn i Rogets tesaurus. Tekst-inndelingar er i tre gradar, som har fått noko ulike termar hos dei som har skrive om NMB. Eg legg meg opp til terminologien som blir brukt om Rogets verk, og kallar gradane frå topp til botn *klasse* (Kolsrud like eins, Venås: hovudemne, Walton: hovudbolck), *seksjon* (Kolsrud og Venås: bolck, Walton: kategori) og *artikkel* (Kolsrud: stykke, Venås: (einskild)emne, Walton: område).

Det er fire overordna *klasser*, med følgjande titlar:

- A *Frumverket*
- B *Emne*
- C *Umstøda*
- D *Mannsfora*

Desse klassene, markert med A, B, C og D, står berre oppført i det innleiande oversynet, ikkje i sjølve teksta. Der er det det neste nivået, 30 *seksjonar*, som utgjer høgaste eining, markert med romartal og fortløpande nummerering frå I til XXX. Seksjonane består i sin tur av *artiklar*, totalt 315, også dei fortløpande nummererte, men med arabiske tal. Dei enkelte artiklane er ofte igjen inndelte i avsnitt med *undergrupper* av ord, helst utan tittel, nokre gonger med. Eksempel på dette vil bli gitt nedanfor.

Det er eit kompliserande moment at Aasen i det innleiande artikkeloversynet ofte gir opp to eller (sjeldan) tre stikkord på dei enkelte artiklane, mens han berre har eitt på sjølve artikkelen. Dei fem siste artiklane står såleis i innhaldslista referert slik:

- 311. Samhelde. Fred.
- 312. Usemjå. Strid.
- 313. Hollskap. Godvilje.
- 314. Uholkskap. Illvilje.
- 315. Fredgjerd. Soning.

Men i sjølve «brødteksta» står berre det første ordet i kvart par som artikkeltittel. Når ikkje noko anna er sagt, refererer eg artiklane på denne korte måten, med berre eitt stikkord.

Heilt til slutt i boka er det to alfabetiske register, eitt på norsk og eitt på dansk – det sistnemnde var til bruk i DO-prosjektet. Eg kjem tilbake til registra under 4.3.

4. Nokre gløttar inn i NMB

4.1. Klasser og seksjonar

Den første klassa er altså *Frumverket*. I NO2014 er det definert som «opphavleg (skapar)verk», med NMB som einaste kjelde; ordet er altså laga av Aasen, men har aldri komme i reell bruk. Denne klassa omfattar desse seksjonane:

I Land, II Vatn, III Luft, IV Jordfang, V Vokster (Plante), VI Dyr, VII Mann, VIII Likam, IX Ande.

Første seksjon er altså I *Land*, som har desse artiklane:

1 Land, 2 Landsbolk, 3 Grunn, 4 Leite, 5 Høgder, 6 Dypter, 7 Marker, 8 Landslag, 9 Føre.

Aasen gjer det fleire stader slik at første underinndeling under ei høgare inndeling får same tittel som det høgare nivået. Truleg

impliserer det at han først gir opp dei mest generelle termene innanfor feltet, og så tar han spesialitetane kvar for seg. Slik er det gjort her, som vi ser. Både seksjonen (I) og den første artikkelen (1) heiter «Land».

Her er det naturformasjonar som står i sentrum. Den første artikkelen, *Land*, gjengir eg og drøftar meir i detalj under 4.2. Den neste, *Landsbolk*, går på 'stykke land i motsetning til andre land-område'. Her er det både naturleg avgrensa landområde: *Dalføre*, *Fjord*, *Koll* 'bakke', og slike område som er avgrensa på kulturelt eller administrativt grunnlag, som *Grend*, *Herad*, *Skipreida*, *Sokn*, *Tinglag*. Dei neste artiklane handlar om ulike landskapstypar og terrengforhold.

Vidare er det naturen som blir behandla i klasse A, under seksjonane *Vatn*, *Luft*, *Jordfang* (dvs. geologiske forhold), *Planter* og *Dyr*. Deretter kjem *VII Mann* (dvs. 'menneske'). Her er det samfunnet som blir beskrive, med termar som gjeld den måten menneska fungerer på i samfunnet og kva slags sosiale forhold dei står i til andre menneske, som *Skyldfolk* 'slektningar', *Sivjungar* 'inngifta slektningar', og så følgjer ulike yrke, menneske etter personlege eigenskapar – her inngår det også skjellsord – grupperingar av menneske etter kor dei kjem frå, folkeslag og andre kollektivitetar. Neste seksjon, *Likam* 'kropp', går på det fysiske og anatomiske, men òg på kroppsslyte og trivsel. Endeleg har vi den korte bolken *Ande*, som handlar om religion og det overnaturlige. Det høyrde også med til språket, uansett kva ein elles måtte meine om det.

Neste klasse er B *Emne*, og denne overskrifta bruker han på tre nivå, både på første seksjonen innanfor klasse B (seksjon X) og på første artikkelen innanfor seksjonen (artikkel 83). Denne artikkelen er generell, og så går Aasen vidare med å gruppere emna etter bruksverdi: Først bruksemne og forråd, og så får vi emna grupperte slik det var naturleg i det gamle landbruks- og naturalhus-haldet. Deretter tar han for seg *Slag*, det vil seie kategorisering av

emne i forhold til kvarandre, termar for jamføring osv. Så går det på *Mengd* og talforhold, på *Maal* og gradering av ulike slag (bl.a. lengd, breidd, vekt og verdimål), og på *Skap* (form og skikkelse). Til slutt får vi XV *Lag* (kvalitet, som også omfattar sanseintrykk: *Lit* ‘farge’, *Døkkleike* ‘mørke’, og så har vi lyd, *Ange* ‘lukt’, smak, og ulike eigenskapar som hardheit, styrke osv. oftast sett i forhold til ei antonymgruppe).

Klasse C kallar Aasen *Umstøda*, definert i NO73 som «Omstændigheder, Forhold», og det omfattar seksjonane (frå XVI til XXI) *Tid, Stad, Støde* (situasjon), *Rørsla, Tilfelle* (som han definerer som ‘indtræffende Omstændighed’) og *Tilstand*.

Den siste klassa, D, kallar Aasen *Mannsfora*, eit ord som heller ikkje står i NO73, men *Fora* ‘Færd, Atfærd, Behandling’ står. «Menneskeleg åtferd’ er truleg den beste moderne attgjevinga av begrepet. Denne klassa omfattar desse seksjonane (mine forklaringar i hakeparentes):

XXII *Vit* [dvs. tankeliv i det heile]

XXIII *Vilje* (*Hug*) [som omfattar humør, lyst og ulyst, trivsel, tolmod, og hovmod og arroganse]

XXIV *Vitring* [ser ut til å dekke kommunikasjon allment, også kunstnarleg, her kjem t.d. stev, song og spel inn]

XXV *Gjerd* [’gjerning’, men også ‘måte, skikk’, og vi ser begge tydingar avspegla i artiklane]

XXVI *Vedgjerd* [i NO73 definert som ‘Forbedring, Istandsettelse’]

XXVII *Bureide* [går på hus og husbygging og det som høyrer til i huset]

XXVIII *Handreide* [går på reiskapar og verktøy]

XXIX *Eign* [går på eigedomsforhold]

XXX *Samliv* [må her tolkast vidare enn det moderne ordet *samliv*, omtrent som ‘forhold mellom menneske’]

Heile vegen ser vi kor nøye og allsidig det rurale arbeids- og samfunnslivet i Noreg på 1800-talet er dekt, men også kva for ei utfordring det var for Aasen å skape nye ord for abstrakte begrep som ikkje var kvardagslege for vanlege folk. Når slike ord ikkje står i NO73, så var det fordi den er empirisk basert på Aasens talemålsundersøkingar og ikkje inkluderte konstruerte ord. Men han visste at slike konstruerte ord måtte til, og han var sterkt puristisk og ville i prinsippet helst følge ei islandsk line, med å lage nye ord av heimlege ordrøter og unngå innlånte element. På det området er NMB eit ærgjerrig verk, men oppgåva han stilte seg, var nok for stor for berre éin mann.

4.2. Artiklar: Nokre eksempel

Eg går no ned på det mest konkrete nivået og gir nokre eksempel på enkeltartiklar.

Hovuddisposisjonen i NMB går frå det konkrete og materielle til det meir abstrakte. Eg skal no gjengi tre artiklar, blant dei den første og den siste i heile verket, og drøfte dei litt vidare, for dei eksemplifiserer denne kontrasten ganske slåande. Først artikkel 1: *Land*.

1. *Land*.

a) *samfengt*. – Berg. Bygd. Dal. Fastland. Fjell. Grunn. Heim. Holm. Jord. Land. Landjord. Lende. Mark. Nes. Rike. Skage. Strand. Uppland. Vidaatta. Øy. Øyland. –Verd(i).

b) *serlege*. – Austland. Nordland. Vestland etc. – Utland. Norig. Sverige. Danmark. Holland. Engelland.

c) *etter Bunad*. – Buland. Bygdarlag. Fødebygd, Fødeland. Grannbygd, -lag, land. Grend. Grendarlag. Heimbygd, -land. Nybygd. Storbygd. Umkverv. Øydeland, -mark.

d) *etter Tillægje*. – Avbygd. Avdal. Bergland. Fjellbygd. Fjordbygd. Havland. Høglend. Innland. Jadar. Laagland.

Midland. Sjøland, -bygd. Tverdal, -fjord. Utbygd.

e) *etter Lag*. – Aldeland. Baleland. Berrsnøyda. Farlende. Fiskeland. Forland. Frostland. Grasland. Grjonbygd. Høyland. Kornbygd, -land. Markland. Matland. Myrlende. Skogbygd, -land. Uland. Vaalende.

Denne artikkelen deler Aasen altså i fem undergrupper:

a) *samfengt* betyr ‘felles’, det er dei grunnleggande termene som også er overbegrep, ordna alfabetisk, som vi ser. Dei fleste orda er fellekkandinaviske. Ordet *Heim* betyr her «Verden, Jorden», lik som islandsk *heimur*. Han nemner òg *Verd(i)* særskilt til slutt.

b) *serlege* går ut på at det her er snakk om proprium, namn på spesielle område i motsetnad til andre. Her grupperer han ikkje alfabetisk, men etter innhaldskriterium.

c) *etter Bunad*: *Bunad* betyr i dag folkedrakt, men hadde for Aasen ei langt vidare betydning, knytt til verbet *bu*, som både betydde ‘bu’ og ‘bygge’; vi ville i dag skrive *busetjing* for Aasens *Bunad* her (jf. kap. 2).

d) *etter Tillægje* betyr om lag ‘etter plassering i landskapet’: I NO73 står ikkje ordet *Tillægje*, så det har Aasen laga.

e) *etter Lag*: *Lag* er eit svært mangtydig ord, men kan her truleg tolkast som ‘etter eigenskapar’, sett ifrå gardbrukarens synspunkt. Enkeltorda her gir anten opp type produkt som landet eignar seg til: *Alde* tyder ‘frukt’, og *Fiskeland*, *Grasland*, *Høyland*, *Kornbygd*, *Kornland* forklarar seg sjølv. Eller han gir opp naturtype: *Berrsnøyda* ‘skoglaus mark’, *Farlende* truleg ‘lende som det ikkje er for vanskeleg å reise i’, *Myrlende*, *Skogbygd*, *Skogland*. Somme av orda har ikkje danske ekvivalentar, så Aasen må forklare dei omstendeleg i NO73: *Baleland*, til verbet *bala* ‘streve, slite’: ‘en Kyst som er vanskelig at komme til i Stormveir’, *Forland* ‘Strand hvor det er vanskelig at lande; en Havkyst uden Indløb eller Bugter’, *Matland* ‘Land eller Sted med hensyn til Fødemidler: *Eit godt Matland*’, *Uland*, som altså er noko anna enn *u-land* i dag, det er:

‘et uheldigt Land for Ferdsel og Ophold; en farlig Kyst, en Strand uden Landingssteder o.s.v.’

Så ser vi på 129 *Ring*:

129. *Ring*.

a) *Lykkja*. – Aatrenne. Band. Bendel. Bugt. Draglykkja. Fatl. Gyrda. Halvkast, -stikk. Hanke. Havald. Helda. Hempa. Hespera. Hogold. Homp. Honk. Hork. Horvelde. Kast. Knut (Kjerringknut, Raabandsknut). Kurra. Kurv. Kverva (=Korva). Lykkja. Mella. Moske. Nesting. Ora. Ranglykkja. Rennelykkja, -snar. Ride. Ridtaatt. Snar. Snarlykkja, -rukka. Snur. Sperr. Stikk. Strumpa. Taumsauke. Tull. Tvilla. Urva. Vidja. Vidjespenning. Yrgja.

b) *Kring* (Ring). – Baug. Belte. Bug. Fella. Gard. Gjerde. Gjord. Hjulfella, -ring. Holk. Huv. Kneppe. Krans. Kring. Kringel. Kringla. Krins, Krinsel. Krull. Kruna. Kveil. Kverv, Kvervel. Lekk. Lur. Null. Ring. Ringvid. Rull. Rundel, -ing. Skjæl. Skruvgang. Svarv. Sveiv. Sverkel. Sverv. Svervel. Sving. Sylgja. Trjosk. Tull. Vinding. Ving.

Adj. – bugtutt. dragvind. idutt. kaurutt. kringlutt. krullad. lykkjutt. ringlagd, ringutt. snarad. snurutt. sveiputt. vind. vunden.

V. – bugta. bulla. garda. kringa. krinsa. krulla. kurla. kverva. ringa. snara s. snura. sveiva. sverva. svinga. sviva. tulla. tunna. tvilla seg. vinda s.

M: hespa. kveila. nysta.

Denne artikkelen står under seksjon XIV «*Skap* (Form)» i klassa B *Emne* (ved sida av *Flata*, *Klump*, *Lengja*, *Grein*, *Bøyg*, *Krok*, *Opning*, *Hol*, og andre). I innholdslista er overskrifta: «129. Ring. Lykkja», og Aasen er mest opptatt av praktiske bruksgjenstandar som kan vere ring- eller løkkeforma (smykke spela nok ei perifer rolle). Under a) har han termar knytte til tau og reip, under b) reknar han

opp slikt som er eller kan vere rundt eller bøygd utan å gjere ei meir systematisk undergruppering. Vi kan merke oss at ordet *Tull* står både under a) og b). I NO73 går det fram at dette ordet har ulike tydingar, og dei som er relevante her, er: «Hvirvel, Kreds-gang» og «noget som er sammenrullet», som begge går under b), og «en Bylt, Pakke, et sammenbundet Knippe», som ser ut til å passe best under a) (det er mi tolking, i alle fall).

Så føyer han til kortare rekker av adjektiv og verb som kan assosierast inn i den same (vide) tydingskrinsen. Det var eit problem for han korleis han skulle stille seg til ordklassegruppering i forhold til ei reint semantisk basert rekkefølge av orda på tvers av grammatiske kategoriar. Det er eit generelt problem for alle som lagar begrepsordbøker. Aasen kom altså til at den semantiske grupperinga måtte vere overordna, men at det på nivået under var nyttig å skilje ut verb og adjektiv og i somme tilfelle andre ordklasser. Substantiva markerte han på den da konvensjonelle måten ved å gi dei stor førebokstav.

Forkortinga M i siste lina betyr «Merk», eller «NB», om ein vil.

Etter kvart blir emnefeltet meir abstrakte, og som eit eksempel på det tar eg med artikkel 315 *Fredgjerd*, den siste i boka:

315. *Fredgjerd*.

a) *Uppgjerd*. – Avgreiding. Avrekning. Bot, Bøting. Fullnad. Fullnøgje. Greida. Grid. Jamning. Kvitting. Likning. Millomlag. Nøgje. Raadbot. Samføring. Samnad. Skul-ting(?). Tiljamning. Uppgjerd. Utjamning. Utrekning.

V. – freda seg. gjera Fred. greidast. jamna. kvitta (av). lata seg lita (l. seg nøgja). likna ut. nøgja seg. samføra seg. semja seg. skarva. skulta. skynsama.

Adj. – avgjord. avgreidd. klaarad. kvittad. uppgjord. utjamnad. – Jf. fullnøgd. litande. nøgd (aatnøgd). samd. samtykt. saatt(?).

b) *Soning*. – Avløysning. Ettergiving. Forlating. Frikjen-

ning. Gløyming. Løysning. Miskunn. Naade. Nedsoning. Orsaking. Rettferding(?). Semja. Soning. Sætt, -ing. Tilgiving. Utsoning. (Jf. 313). – »Forlik.«

V. – avløysa. frikjenna. fritaka. heilaga. kannast ved. miskunna. naada. orsaka. setja fri. sleppa (undan). taka fri (t. til Naade). tilgiva. tægjast. vaarkunna.

forlata. giva etter. gløyma. leggja ned. sona ned.

Jf. liva. skona. spara. vægja. – blidna. daast. kløkka. linnast. mildast. mjuka. rørast. sevast.

Adj. – avløyst. frikjend. orsakad. saklaus. skuldfri. – Jf. fredsam. godtolug. langmodug. linn. mild. miskunnsam. naadug. raadmild. skynug. tolug. tøyg. vægjen.

I NO73 er *fredgjerd* definert som 'Fredslutning', med opplysninga «sjelden» (sjå nedanfor). Her er det mange abstrakte begrep som det har vore vanskeleg for Aasen å finne norske ord for. Tre av orda er forsynt med spørsmålsteikn. Det er nok ordframlegg som Aasen har vore usikker på. Vi må stadig ha i minnet at heile NMB er eit resultat av Aasens diskusjonar med seg sjølv, han har ikkje prøvd ut orda på andre språkbrukarar, så langt eg kjenner til. Også elles i NMB er det ein del ord med spørsmålsteikn, dei fleste i dei delane som dekker det abstrakte ordtilfanget.

4.3. Registra

Å finne fram i NMB er ikkje så lett. Aasen laga, som nemnt, to ordregister til dei ulike artiklane, eitt med (ny)norske oppslagsord og eitt med danske (med tanke på DO-prosjektet). Vi skal undersøke dei litt nærare, og dermed også indirekte få eit visst innsyn i dei terminologiske problema Aasen hadde og korleis han prøvde å løyse dei.

Eg har gjort det slik at eg i første omgang har gått gjennom registra og registrert alle referansane til dei tre artiklane eg har

sitert ovanfor. Oppstillinga skal lesast slik at når ein slår opp på for eksempel «Uppgjerd» i det norske registeret eller «Forlig» i det danske, blir ein i begge tilfelle vist til artikkel 315, altså *Fredgjerd*. Slik:

Uppfostring. 302.

Uppgjerd. 315.

Upphav. 154.

Forlegen. 204.

Forlig. 315

Form. 76. 121. 284

Og så ser vi på referansane til dei tre artiklane. Først det norske registeret:

1. *Land*: Jord, Land

129. *Ring*: Kring, Lykkja, Ring

315. *Fredgjerd*: Fredgjerd, Soning, Uppgjerd

Så det danske:

1. *Land*: Jord, Jordstrøg, Land

129. *Ring*: Bøile, Kreds, Omkreds, Ring, Runding

315. *Fredgjerd*: Forlig, Forsoning, Fred, Fyldestgjørelse, Naade, Opgjør, Qvittering, Tilgivelse, Ædelmodighed

Totalt har det danske registeret 1444 oppslag etter mi teljing, mens det norske har berre 710. Det viser at Aasens primære problem da han utarbeidde registra, var begrepsklassifiseringa av dei *danske* orda og ekvivalering av dei på norsk. Og når ein ser på oppstillinga ovanfor, ser ein kor dei største vanskane låg, nemleg i det abstrakte og samfunnsmessige ordtilfanget. Det var der det landbruksba-

serte norske folkemålet trong mest bearbeiding for å utvikle eit moderne begrepsapparat og ordtilfang.

Oppslagsorda i registeret er for det meste seksjons- og artikkeltitlar, men det kan òg vere enkeltord som representerer undergrupper eller finst på enda lågare nivå. Slike enkeltord er det mest av i den danske lista, og dei fleste av dei viser til dei siste klassene, særleg D, altså dei mest abstrakte begrepsfelta.

Reint praktisk må det ha fungert slik at når brukaren slo opp på for eksempel *Tilgivelse* i den danske lista og vart vist til nummeret 315, ville han finne eit nynorsk synonym til dette ordet under *Fredgjerd*. Men der måtte han berre leite i listene for å finne det beste synonymet; i stor grad var det altså brukarens skjønn som ville avgjere om det vart eit akseptabelt resultat. Så lenge Aasen var einaste brukaren, var dette sjølv sagt ikkje noko eigentleg problem, utover det at han ikkje hadde nokon eksisterande språkbruk å forhalde seg til. I DO har Aasen ført opp *Tilgivelse* med ekvivalentane «Ettergiving, Miskunn». Begge desse orda står under b *Soning* i artikkel 315 i NMB. Men i dag vil nok begge desse orda ha andre tydingsvalørar enn *tilgjeving*, som i seg sjølv er blitt eit vanleg ord også på nynorsk. Og slik var det truleg også for folk på Aasens tid, for det religiøst-moralske begrepsapparatet for bygdefolk på 1800-talet var jo prestens danske språk. Først seinare utvikla det seg eit nynorsk kyrkjepspråk, men det hadde Aasens upubliserte alderdomsverk neppe nokon innverknad på.

4.4. Nokre vanskelege begrep

For å illustrere Aasens problem med abstrakte begrep, listar eg opp titlane på artiklane i seksjon XXX *Samliv*. I parentes etter kvar tittel set eg definisjon(ar) og bruksopplysningar i NO73. Minusteikn (÷) tyder at ordet ikkje finst i NO73, og dermed truleg er nylaga av Aasen. Somme ord er oppførte med fleire tydingar i NO73; eg refererer da den som verkar mest relevant her.

XXX. *Samliv.*

299. Skyldskap. (Slægtskab)
300. Samnøyte. (Bordselskab)
301. Heimlag. (÷)
302. Uppfostring. (ikkje oppslagsord, men nemnt som vanleg synonym under *Uppfødsla*)
303. Lagfesting. (÷)
304. Lagskipnad. (÷)
305. Samgreida. (÷)
306. Samkoma. (Sammenkomst)
307. Hjelp. (Hjælp, Understøttelse (osv))
308. Vyrndnad. (Agtelse, Anseelse)
309. Velde. (Vælde, Magt, Raadighed)
310. Tevling. (Kappestrid, Veddekamp; Strid, Anstrængelse)
311. Samhelde. (Enighed, Sammenhold, God Forstaaelse.)
312. Usemjá. (Uenighed)
313. Hollskap. (Huldskab. Velvillie. (Lidet brugt.))
314. Uhollskap. (÷)
315. Fredgjerd. (Fredslutning (Sjeld.))

Her ser vi altså fem minusteikn på sytten oppslag, og ved to av orda med danske ekvivalentar er det markert at dei er «sjeldne» eller «lidet brugt».

Ved 300 *Samnøyte* er «Bordselskab» openbert ein for snever definisjon. Aasen underbygger det i NO73 ved å føye til uttrykket *hava Samnøyte* 'spise sammen' frå Hallingdal. Men han nemner òg at norrønt *samneyti* betydde «Samfund etc.». Og i eit tillegg med rettingar i NO73 presiserer han: «*Samnøyte* n. (rettere): Fællesskab i en vis Nydelse; Selskab, samfund; især Bordselskab etc.» Men i NMB er det openbert meininga at ordet skal dekke eit vidare tydingsfelt, om lag 'menneskeleg samkvem' reint allment. Artikkelen *Samnøyte* er delt i to bolkar: a) *Samlag*, som nok langt på veg

er synonymt med overbegrepet Samnøyte, og b) *Kjendskap*, som dekker ulike gradar av og former for det som no mest vanleg blir uttrykt med bokmålsordet «berømmelse».

Når det gjeld *Uppfostring*, så er det altså ikkje oppslag i NO73, men nemnt som eit vanleg synonym til *Uppfødsla*. *Opfostring* var vanleg i dansk, og Aasen har kanskje ønskt å prioritere *Uppfødsla* som det norske ordet i NO73, men så endra meining i NMB og likevel akseptert *Uppfostring* som hovudordet. Men dette er også berre mi tolking.

Eg kommenterer dei fem orda med minus kort her.

Heimlag er ikkje belagt nokon annan stad enn akkurat her, så det må reknast som eit mislykka forsøk på eit nyord. Artikkelen er delt i undergruppene *Lydlag*, *Lydskort* og *Einvære*, og ingen av desse orda er med i NO73. Ifølgje NO2014 er *einvære* sjeldan brukt, men tydinga er 'einsemd' ('ensomhet'); dei to andre er berre belagt i NMB. I Aasens danske register viser oppslaga *Eenlighed*, *Familie*, *Folkemangel*, *Huusfolk* til artikkel 301. Elementet *Lyd-* betyr altså 'folk' (norrønt *lýðr*), og dermed får vi desse tentative definisjonane: *Lydlag* 'samvær med folk'. *Lydskort* 'mangel på folk'. Og ordutvalet i sjølve artikkelen tyder på at det handlar om nettopp det, da helst «folk» som arbeidshjelp (som i det gamle bondesamfunnet vart rekna som del av hushaldet) og familiemedlemmer. *Heimlag* er da kanskje, på dansk, kombinasjonen av «Familie» og «Huusfolk».

Lagfesting er òg berre belagt i NMB; i NO2014 tolka som 'innsetjing (i eit embete o.l.)' og 'vigsle'. Det passar med ordutvalet i artikkelen; det omfattar også innføring i ein ny (kyrkjeleg og sosial) status (dåp, konfirmasjon, og andre typar innviing). Dei relevante oppslagsorda i det danske registeret (som viser til artikkel 303) er *Ansættelse*, *Godvillie*, *Indvielse*, *Vielse*. Kanskje *Lagfesting* da kan definerast som 'innføring i ein fellesskap'.

Heller ikkje *Lagskipnad* er belagt andre stader enn i NMB og må dermed reknast som eit mislykka forsøk. Men meininga må vere noko i retning av 'samfunnsordning'. Det er tre undergrup-

per: *Styring, Lydnad og Rett*, og dei peikar ganske eintydig i den retninga. Orda i det danske registeret som viser til artikkel 304, er ganske mange: *Anordning, Befaling, Bestemmelse, Føielighed, Lydighed, Ordning, Pligt, Protest, Regel, Regjering, Ret, Skyldighed, Straf, Styrelse, Valg, Ærefrygt*.

Samgreie betyr ifølgje NO2014 ‘noko ein gjer eller har saman’, og det er nemnt eit belegg hos ein forfattar frå Aasens eiga tid, men elles er dette ordet heller ikkje brukt. Også her er det mange oppslag i det danske registeret som viser til artikkel 305: *Afgjørelse, Contract, Forretning, Gjenstand, Mellemværende, Overenskomst, Regnskab, Sag, Samraad, Underhandling, Ærinde*. Det ser ut som *Mellemværende* er det danske ordet som er mest dekkande som overbegrep; Aasen har *Millomvære* med som enkeltord i sjølve artikkelen. Men *hopehav* er kanskje det moderne ordet som best dekker dette begrepet; det står også i NO73, men her i artikkel 305 har han berre med langforma *hopehavande*.

Uhollskap har nok Aasen nylaga som motsetning til *Hollskap*; også dette ordet er berre belagt i NMB. Det går på positive vs. negative kjensler overfor andre menneske. Dette viser korleis Aasen markerer antonymi ved å samle motsetningar i separate artiklar som følgjer på kvarandre, det positive først og det negative etterpå, som kan reknast som ei «naturleg» rekkefølgje. Rett framfor *Hollskap* står 311 *Samhelde*, i innhaldslista presisert med *Fred*, og 312 *Usemjå*, presisert med *Strid* (sjå kap. 3).

Andre stader samlar Aasen motsetningane i same artikkel. 307 *Hjelp* inneheld såleis undergruppene *Beine* og *Ubeine*, og *Beine* betyr da ‘teneste’, mens *Ubeine* er eit (igjen nylaga og seinare ubrukt) ord for det å vere tverr og lage vanskar for andre. 308 *Vyrndnad* inneheld undergruppene *Vyrding* og *Uvyrdnad* (hån, spott, forakt o.l.), og 309 *Velde* inneheld undergruppene *Velde* (Magt) og *Undermagt* (i NO73 definert som «svagere Magt»); ‘underordna stilling’ vil vi vel seie i dag).

Atter andre stader blir antonymien markert på eit enda låga-

re nivå. Eit eksempel på det har vi under 300 *Samnøyte*, bolk *b Kjendskap* (jf. ovanfor). Her skil han ut adjektiva frå substantiva, og så får vi denne oppstillinga:

Adj. – folkegjengd. folkekjend, -van. heimholl(?). kjend.
kunnug. mannkjend, -tamd. reken. vidfaren. vidkjend.
(Mots.) framand. langfaren. nykomen. ukjend. ukunnug.

Markering av antonymi har i det heile vore eit viktig problem for Aasen, som det er for alle som arbeider med begrepsordbøker. Han gjorde det altså på ulike nivå, truleg også styrt av kor mange ord som var inne i bildet i kvart tilfelle.

5. Roget og Aasen: Likskapar og skilnader

Sidan Roget var det fremste forbildet for Aasen, vil ei samanlikning mellom dei vere interessant, trass i dei uvanleg ulike utgangspunkta, med verdspråket engelsk mot eit nynorsk som enno var på teiknebrettet. Men Hüllen (2005:335–337) gir ei interessant belysning ved å jamføre Roget med ein av forgjengarane hans, John Wilkins, som i 1668 gav ut ei tesaurisk oversikt over engelsk ordtilfang. Hüllen viser at der Wilkins var empirikar og sette opp konkrete nomenklaturar over dyr, planter, tal osv., baserte Roget seg på abstrakte idear og førte opp termar og ord med konkret tyding mest som illustrasjon til ideane. Til dømes hadde Wilkins under *relations*, altså slektskapstermar, *Progenitor/Descendant*, *Parent/Child*, *Uncle/Nephew*, *Brother/Half Brother*, *First Cosin/Cosin*. Roget hadde derimot overbegrepet *consanguinity* med termar som *kindred*, *blood*, *parentage*, *cousin*, *brother*, *sister*, *brotherhood*, *sisterhood*, *cousinhood* og mange andre.

Det er vel lite truleg at Aasen visste om Wilkins, men det var hans spor han følgde i. I artikkelen 55 *Skyldfolk* står såleis først un-

dergruppa *Mann*, med ord som *Broder, Broderason, Dotterson* osv. og seinare undergruppa *Kvinna: Broderdotter, Dotter, Dotterdotter* osv. – rekkefølga i kvar undergruppe er alfabetisk. I den tredje undergruppa, *samfengt*, står *Barn, Barnebarn, Eigefolk* osv. Vi har sett nok av eksempel på dette ovanfor: Aasen baserte seg på det konkrete språket og livsverda til norske bygdefolk, og hadde, som nemnt, atskilleg strev med å bygge opp ein meir abstrakt terminologi, som han visste var nødvendig i den moderne verda, og som dessutan ville vere heilt nødvendig når han skulle lage ei ordbok frå det meir utvikla kulturspråket dansk til det nye norske skriftspråket – det prosjektet som han hadde langt framme i tankane, men som til sjuande og sist ikkje vart realisert.

I innleiinga til tesaurusen la Roget vekt på to hovudmål:

1) Sjølve struktureringa av det menneskelege begreppssystemet og innkledinga av det i ord, med Rogets eigne ord (sitert etter Hüllen 2004:327):

The purpose of an ordinary dictionary, is simply to explain the meaning of words; and the problem of which it professes to furnish the solution may be stated thus: – The word being given, to find its signification, or the idea it is intended to convey. The object aimed at in the present undertaking is exactly the converse of this; namely, – The idea being given to find the word, or words, by which that idea may be most fitly and aptly expressed.

2) Bruksverdien: Det skulle vere ei bok der engelske morsmålsbrukarar skulle finne det eksakte ordet for det dei tenkte og ønskte å uttrykke.

I kor stor grad Roget lykkast i dette, ligg utanfor emnet her. For Aasens del har det praktiske føremålet vore det dominerande,

nemleg å bygge opp ein semantisk struktur som skulle hjelpe han i ordboksarbeidet, og også gi han støtte i arbeidet for å utvikle nynorsken som moderne kulturspråk. Rogets mål 2, å lage ei bruksbok for andre språkbrukarar, har nok Aasen også tenkt på, men han kom altså ikkje i gang med å bearbeide manuskriptet sitt til det formålet. Og da NMB låg føre som trykt bok frå 1925, vart ho stort sett liggande eller ståande i hyller, nedstøva og uoppskoren.

6. NMB i ettertida

NMB er blitt ståande i språkhistoria og forskingshistoria som eit ikon, og ikkje ei bruksbok. Begge Aasen-biografane undrar seg over det. Venås peikar på at boka nok ikkje var lett å bruke, utan sidetal i det innleiande oversynet og eit register baki som berre tok med overbegrep, ikkje enkeltord. Ho er ikkje så lett å finne fram i, utan at ein har gjort ein forholdsvis grundig jobb med å setje seg inn i korleis ho er organisert. Dessutan er det nok ein hemsko for ei slik bok at det ikkje er med definisjonar og forklaringar. På engelsk går det, for Roget kunne stole på at dei potensielle brukarane utan vidare kjente dei enkeltorda han gav opp (men han kan jo ha overvurdert dei). Aasen kunne absolutt ikkje ta det for gitt. Ingen brukar av NMB i dag kan la vere å ha NO73 ved sida av seg for å vere sikker på kva dei enkelte orda betyr. Til og med ei moderne nynorskordbok er ikkje nok, for orda kan ha – og har ofte – endra tyding i forhold til det Aasen la i dei, og det er det siste som er relevant når ein vil tolke NMB.

I tillegg spelar det ei rolle at NMB beskriv eit urealisert språk og eit ordtilfang som berre delvis vart realisert, eller som for stadig større delar av folket vart utdatert på grunn av urbaniseringa. Den nynorsken som etter kvart voks fram i bruk, vart annleis enn den Aasen framstilte i NMB. Det var nok blitt klart alt i 1920-åra, og enda meir seinare. Og som sagt: Aasen kom aldri til å redigere

verket med tanke på allmenn bruk. Så NMB er og blir primært eit faghistorisk dokument.

Men i heilt moderne tid har det likevel skjedd noko: I samband med oppbygginga av det digitale tekstkorpuset til prosjektet Norsk Ordbok 2014 (ved Universitetet i Oslo), vart også NMB digitalisert ca. 1996 av det daverande Dokumentasjonsprosjektet ved Universitetet i Oslo (i dag: Eining for digital dokumentasjon). Det er også i gang eit arbeid med eit alfabetisk register over alle ord i NMB med tilvising til dei NMB-artiklane der ordet er oppført (ved Christian-Emil Ore).

Walton på si side vil forklare situasjonen ved å peike på den uvanleg svakt utvikla leksikontradisjonen i Noreg, som har hemma framvoksteren av ein norsk synonymikk, i staden for å framheve svakheiter i NMB. Dette er eit viktig poeng, men først og fremst med tanke på eit anna spørsmål: Korfor har ingen – med eller utan inspirasjon frå NMB – gått i gang med å lage ein moderne norsk tesaurus – for bokmål eller nynorsk eller begge – gjennom heile 1900-talet? Det er i grunnen meir merkeleg enn at Aasens verk ikkje vart meir brukt.

No har Johan Hammond Rosbach gitt ut ein ny norsk tesaurus – for bokmål – i 2001, og det var ganske seint, samanlikna med Sverige og Danmark. Rosbach gir opp at han er inspirert av Roget, men nemner ikkje NMB. Tesaurusen hans omfattar eit moderne norsk ordtilfang ved tusenårsskiftet, og er sjølvsagt tilgjengeleg for vanlege brukarar på ein heilt annan måte enn NMB, men ein systematikk som kan måle seg med Aasens finn vi ikkje hos Rosbach. Både prinsippa for klassifisering, utvalet av ord og assosiasjonane som ligg til grunn for kategoriseringa av dei enkelte begrepa, verkar ofte svært tilfeldige (sjå omtale i Guttu 2002).

Ein god og dekkande tesaurus for moderne norsk, med ei god og gjennomtenkt organisering av ordmaterialet, manglar vi altså framleis. Det ville kanskje vere ein god idé å ta fram att *Norsk Maalbunad*, ikkje for å gjenutgi eller kopiere denne boka, men

for å vurdere om ein ut frå den systematiske grunnstrukturen i den kunne bygge opp ein modernisert begrepsstruktur som ein så kunne setje det noverande ordtilfanget i bokmål, nynorsk eller begge, inn i.

Litteratur

DO = Aasen 2000.

Fjeld, Ruth Vatvedt og Lars S. Vikør 2008: *Ord og ordbøker. Ei innføring i leksikologi og leksikografi*. Kristiansand: Høyskoleforlaget.

Guttu, Tor 2002: En leksikografisk sjeldenhet. [Melding av Rosbach 2001.] I: *Ord om ord* 8, 156–158.

Hüllen, Werner 2004: *A History of Roget's Thesaurus*. Oxford: Oxford University Press.

NMB = Aasen 1925.

NO2014 = *Norsk Ordbok*.

NO73 = Aasen 1873.

Norsk Ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet. 1966–. Oslo: Det Norske Samlaget.

Roget, Peter Mark 1852: *Thesaurus of English Words and Phrases*. London: Longman, Brown, Green and Longmans.

Rosbach, Johan Hammond 2001: *Norsk tesaurus*. Oslo: Pax forlag.

Venås, Kjell 1996: *Då tida var fullkomen. Ivar Aasen*. Oslo: Novus.

Vikør, Lars S. 1997: Kva kan Norsk Ordbok lære oss? I: Lars S. Vikør (red.): *Norsk Ordbok – nynorskens leksikografiske kanon?* Oslo: Universitetet i Oslo, Institutt for nordistikk og litteraturvitskap.

Vikør, Lars S. 2006: Dei diakrone aspekta ved Norsk Riksmålsordbok og Norsk Ordbok. I: *LexicoNordica* 13, 7–23.

Walton, Stephen 1996: *Ivar Aasens kropp*. Oslo: Det Norske Samlaget.

Aasen, Ivar 1850: *Ordbog over det norske Folkesprog*. Christiania.
Ny utgåve 2000 ved Terje Aarset og Kristoffer Kruken, Oslo:
Det Norske Samlaget.

Aasen, Ivar 1873: *Norsk Ordbog med dansk Forklaring*. Christiania.
Fleire opptrykk; ny utgåve 2003 ved Terje Aarset og Kristoffer
Kruken, Oslo: Det Norske Samlaget.

Aasen, Ivar 1925: *Norsk Maalbund*. Utgitt av Sigurd Kolsrud,
Oslo: Det Norske Samlaget. Fotografisk opptrykk 1975, Ber-
gen: Norsk Bokreidingslag.

Aasen, Ivar 2000: *Dansk-norsk Ordbog*. Utgitt av Terje Aarset, Jarle
Bondevik og Oddvar Nes, Bergen: Norsk Bokreidingslag.

Lars S. Vikør
professor
Norsk Ordbok 2014
Institutt for lingvistiske og nordiske studium
Universitetet i Oslo
Postboks 1021 Blindern
NO-0315 Oslo
l.s.vikor@iln.uio.no

IKKE-TEMATISKE BIDRAG

Hvad hedder det på dansk? Polske studerendes brug af ordbøger mellem polsk og dansk¹

Marcin Overgaard Ptaszynski, Mikołaj Sobkowiak & Patrick Leroyer

The paper is a research study into how Polish students of Danish use dictionaries when translating from Polish into Danish and when writing texts in Danish. The methodology of the study relies on three assumptions: (1) the distinction between subjective and objective information needs, (2) the impact of educational level on lexicographic behaviour, (3) the differences in terms of users' needs and their satisfaction during translation and free text production. By combining various data generating and gathering methods, the study provides an insight into differences between both kinds of needs. It uncovers the dictionary users' objective needs and examines how their subjective needs are satisfied, i.e. which dictionaries are consulted, for what purposes, what determines their choice, and how useful they are.

1. Indledning: behov for ordbøger mellem polsk og dansk

I 2010 udkom anden udgave af Jan A. Nowaks *Polsk-Dansk Ordbog* (herefter PDO), der med sine ca. 70.000 opslagsord er den største ordbog mellem dansk og polsk. Det giver anledning til at kigge nærmere på leksikografi mellem de to sprog. I betragtning af at interessen for polsk i Danmark og for dansk i Polen slet ikke kan måles med interessen for store europæiske sprog – først og fremmest engelsk, tysk og fransk – ser udbuddet af ordbøger med dansk og polsk imponerende ud. Ikke alene findes der flere almenordbøger

1 Vi vil gerne takke Thomas Mathiasen fra Adam Mickiewicz Universitet i Poznań for hjælp med at gennemføre undersøgelsen.

(f.eks. Sørensen og Widding 1989, Frank-Oborzynska 2008, Ste-mann 2008, Nowak 2010, Kuciewicz (uden årstal)) og fagordbø-ger (f.eks. Okoniewski 1999), men også grammatik-opslagsbøger (f.eks. Rajnik 1999), parlører (f.eks. Michalska og Król 2007), ord-lister integreret i lærebogssystemer (f.eks. Bostrup og Sochańska 2006) og en billedordbog (Schothuis og Helbo (uden årstal)). Fæl-les for alle de nævnte værker er, at de kun forekommer i papirud-gave (dog kan parlørerne være ledsaget af cd'er med lydoptagelser af indholdet). Egentlige elektroniske ordbøger mellem dansk og polsk hører stadigvæk fremtiden til.

Men hvem har overhovedet brug for dansk-polske ordbøger? Ud over turister, polsk arbejdskraft i Danmark (håndværkere og medicinsk personale) og danske forretningsfolk i Polen er der en stor, stabil gruppe potentielle ordbogsbrugere i begge lande: folk som vil lære dansk eller polsk på en systematisk måde. I skrivende stund er der to universiteter i Polen, hvor man kan læse dansk som hovedfag (i Poznań og Gdańsk) og to (i Wrocław og Warsz-awa), hvor systematisk undervisning i dansk tilbydes inden for an-dre studieretninger (hhv. som led i tyskstudiet og som et valgfrit sprogkursus). Det anslås, at omkring 150 polske studerende lærer dansk som fremmedsprog nu, og at flere hundrede har afsluttet deres uddannelse som cand.mag. i dansk siden 1975. I Danmark kan man læse polsk som hovedfag på Københavns Universitet. Derudover findes der en række sprogskoler i begge lande, som ud-byder undervisning i dansk hhv. polsk.

Frem for at undersøge, hvad der findes på markedet af ord-bøger med dansk og polsk, og hvem der kunne have gavn af dem, har vi undersøgt, hvilke (typer af) ordbøger, der rent faktisk bliver anvendt, og om de er passende til deres formål. Vi fokuserer på én type ordbogsbrugere og på to typer situationer, hvori brugerne oplever informationsbehov: polske studerende med dansk som hovedfag og deres brug af ordbøger i forbindelse med oversættelse fra polsk til dansk og ved fri tekstproduktion på dansk.

Vores undersøgelse har tre mål. Det første er at afklare de behov, der opstår i forbindelse med oversættelse og fri tekstproduktion. Det andet er at skabe indsigt i, hvilke (typer af) ordbøger der anvendes for at løse informationsproblemer, hvorfor dette sker og med hvilket resultat; særlig fokus er rettet mod PDO i sammenligning med andre ordbøger. Det tredje og sidste mål er at bidrage til den metodologiske udvikling af leksikografiske brugerundersøgelser (jf. Tarp 2008). Dette har vi forsøgt at gøre ved at udvikle og anvende en ny metode. Metoden er kendetegnet ved kombinationen af fire dataindsamlingsmetoder – spørgeskemaer, protokoller, interviews og tekstretning – og ved samkøring af resultaterne. Vi tager udgangspunkt i brugernes behov og valg af ordbøger snarere end i tilstedeværelsen og præsentationen af specifikke datatyper i konkrete ordbøger.

Afslutningsvis efterlyser vi mere omfattende leksikografiske brugerundersøgelser, der ved at reproducere samme metode kunne generere endnu flere resultater.

2. Undersøgelsens baggrund, design og metode

2.1. Baggrund

Selv om traditionen for at udforske anvendelse af ordbøger er næsten hundrede år gammel (Lew 2011:1), er langt de fleste leksikografiske undersøgelser blevet gennemført i de seneste tre årtier. En systematisk, kritisk oversigt af 200 undersøgelser for perioden 1962-2006 bringes i Welker (2010). Forskningsemnet synes også at være meget populært for tiden. Et af de seneste hæfter af *International Journal of Lexicography* (24.1) er et temahæfte, som udelukkende beskæftiger sig med brugerundersøgelser. Emnet har også tiltrukket sig nordiske leksikograferes opmærksomhed. Undersøgelser af ordbogsbrug udgjorde temasektionen i *LexicoNordica* 15,

og de forekommer med jævne mellemrum i rapporter fra NFL-konferencer.

Mange leksikografiske brugerundersøgelser beskæftiger sig med anvendelse af ordbøger ved tekstproduktion, og hertil regner man ofte også ordbogsbrug ved oversættelse. Her kan nævnes bl.a. Varantola (1998) og Nord (2002). Tono (2001:33ff.) omtaler 11 studier i ordbogsbrug ved tekstproduktion på fremmedsprog (herunder oversættelse til fremmedsprog) og sammenfatter deres konklusioner og de anvendte metoder. Efter vores mening er det ikke særlig hensigtsmæssigt at inkludere oversættelse og tekstproduktion i samme kategori og dermed antage, at de behov for information, der opstår i de to situationer, er ens. Begge situationer indebærer tekstproduktion, men oversættelse er mere kompleks, da den også forudsætter reception af den originale tekst. Desuden er produktionen af teksten på målsproget ikke fri, men afhængig af kildetekstens indhold og struktur. De to situationer kan dermed afføde forskellige leksikografiske behov. Denne hypotese er indbygget i funktionsteoriens rammer, jf. Tarp (2006). Vi har søgt nye veje og har undersøgt ordbogsbrug både ved oversættelse og ved fri tekstproduktion på fremmedsprog, hos samme informanter og i de samme omgivelser. Os bekendt findes der ingen andre undersøgelser af denne art.

Vi har ladet os inspirere af flere studier. Et af formålene med vores undersøgelse har været at finde frem til brugernes valg af konkrete (typer af) ordbøger i forbindelse med konkrete aktiviteter. Dette har også været tilfældet hos f.eks. Nord (2002), Varantola (1998), Frankenberg-Garcia (2005) og Pilke (2008). Til forskel fra Pilke har vi ikke brugt spørgeskemaer for at udvinde denne information, da metoden ikke er særlig egnet til kvalitativ forskning. Lige som Varantola og Frankenberg-Garcia anvendte vi protokoller,² men for at få maksimalt udbytte supplerede vi dem

2 I modsætning til Varantola og Frankenberg-Garcia, som anvendte mundtlige protokoller, har vi valgt at anvende skriftlige protokoller.

med interviews. En kombination af protokoller og interviews med udvalgte informanter blev også anvendt af Duvå og Laursen (1994), men dér blev interviewene først foretaget to uger efter den afsluttede aktivitet. Vi gjorde det straks efter, så aktiviteterne var i frisk erindring.

Varantola (1998), Duvå og Laursen (1994) og Frankenberg-Garcia (2005) anvender alle protokoller for at samle data om, hvad der nøjagtigt giver anledning til ordbogskonsultation. Det gjorde vi også, men både spørgsmålenes udformning i vores protokoller og muligheden for at besvare dem er forskellige fra, hvad man ser i protokollerne brugt i de nævnte studier. I Varantolas undersøgelse blev informanterne efter hver konsultation spurgt om, hvorfor de havde slået et bestemt ord op, og svaret skulle vælges blandt 3 på forhånd fastsatte svarmuligheder: “du kender ikke ækvivalenten på målsproget”, “du tror, du kender ækvivalenten på målsproget, men du vil tjekke det” og “du har brug for anden hjælp til at bruge ækvivalenten på målsproget – vær specifik, om muligt” (Varantola 1998:182). Et lukket sæt af svarmuligheder foreligger ikke i Duvå og Laursens protokoller (1994:58), men “manglende ækvivalent på spansk” anføres som et eksempel på et problem. Informanterne i denne undersøgelse blev også bedt om at specificere, hvad der havde voldt dem problemer med oversættelsen (Duvå og Laursen 1994:249), men det præsenterede uddrag fra en protokol viser, at problemfeltet kun blev udfyldt ved at anføre de(t) konkrete ord, der var problematisk(e). Vi synes, at man ved at angive eller antyde svarmuligheder i høj grad risikerer at påvirke svarene, så vi valgte at give vores informanter frie muligheder til at forklare, hvad deres problemer var. Dette var tilsyneladende også, hvad Frankenberg-Garcia havde gjort, da det relevante felt i hendes protokol kun hedder “hvad forsøgte du at finde ud af?” (2005:338). Vi har dog ikke set eksempler på de svar, som hendes informanter angav. Dermed kan vi heller ikke udelukke, at svarene ikke blev givet ud fra de problemkategorier, som præsenteres i omtalen af undersøgelsens

resultater, som f.eks. “at finde ækvivalent på målsproget” eller “at tjekke stavemåden”.

Ved planlægningen af undersøgelsen har vi forsøgt at tage højde for den hidtidige kritik af videnskabelige brugerundersøgelser. Der findes en lang række kritiske analyser af specifikke studier (f.eks. Humblé 2001) og af metodevalg generelt (f.eks. Tono 2001, Tarp 2008 og Welker 2010). Vi har på flere punkter holdt os til Tarp (2008), da hans kritik – os bekendt – er den eneste, som er forankret i en sammenhængende leksikografisk teori.

2.2. Design

Da vores studium handler om brugernes informationsbehov og den måde, hvorpå disse bliver tilfredsstillet ved hjælp af ordbøger, har vi designet undersøgelsen med udgangspunkt i den leksikografiske funktionsteori, som sætter brugerne og deres behov i centrum. Vi følger således Tarps anbefaling om, at “undersøgelser af ordbogsbrug (...) ikke kun bør dreje sig om, **hvordan** man bruger en ordbog, men også om **hvem** brugerne er, **hvor**, **hvornår** og **hvorfor** de bruger ordbøger, og med **hvilket** resultat” (2008:9).

I henhold til Tarp (2008:23-27) har vi for det første valgt at lave en kvalitativ undersøgelse, dog med den modifikation, at vi bringer udvalgte kvantitative data for at tydeliggøre de kvalitative iagttagelser. Vi er nemlig enige med Tarp i, at rent kvantitative undersøgelser, som udgør det overvejende flertal af alle publicerede undersøgelser, synes at hvile på et tvivlsomt statistisk grundlag og dermed at have begrænset relevans for den fortsatte udvikling af den leksikografiske teori. For det andet har vi fokuseret på den konkrete brugssituation snarere end på bestemte datatyper. Med andre ord har vi sat fokus på, hvorfor brugerne benytter sig af et konkret opslagsværk, og hvad de får ud af det, snarere end hvilke data, de tjekker. For det tredje har vi kombineret forskellige datagenereringsmetoder. Vi har anvendt spørgeskemaer for at af-

dække brugerprofilen og dernæst brugt både skriftlige protokoller og dybdeinterviews for at undersøge selve ordbogsbrugen. Retningen af informanternes tekster blev brugt for at afdække de behov, som ikke blev erkendt af informanterne under selve skriveprocessen. Ved at rette fokus både på brugssituationen, på brugeradfærd og på resultatet af aktiviteterne har vi forsøgt at afdække brugerens subjektive såvel som objektive behov.

Vores første skridt var at finde frem til vores informanter og de typer af situationer, hvori de ville få brug for ordbøger. Vores valg af polske studerende med dansk som hovedfag ved Adam Mickiewicz Universitet i Poznań (AMU) og af oversættelse og tekstproduktion har vi truffet af flere grunde. For det første er brugere af denne type i kraft af deres uddannelse beskæftiget på fuld tid med systematisk sprogtilegnelse over en årrække. For det andet har det givet os mulighed for at inddele dem i tre forskellige grupper, der hver især befinder sig på forskellige trin i uddannelsesforløbet. For det tredje ligner de ekstra-leksikografiske situationer, som danner udgangspunkt for vores undersøgelse, de situationer, som vores informanter oplever i deres uddannelse. Dermed er der tale om en undersøgelse af adfærdsmønstre, som er naturlige og ikke er specielt konstrueret til projektets formål. Undersøgelsen bestod af fire faser, hvori der genereredes og indsamledes data om:

1. Brugerprofiler (spørgeskemaer)
2. Brugerbehov under selve testen: oversættelse og tekstproduktion (skriftlige protokoller)
3. Brugernes valg af ordbøger (post-test-interviews)
4. Brugernes produktion (evaluering ved opgaveretning)

De 19 deltagere har vi inddelt i tre brugerprofiler – herefter kaldt BP1, BP2 og BP3 – ud fra antagelsen om, at viden om ordbøger og deres brug forøges i løbet af brugerens uddannelsesforløb:

1. BP1-informanter (8 deltagere) har læst dansk i 1 til 2 år,
2. BP2-informanter (4 deltagere) i 2 til 4 år og
3. BP3-informanter (7 deltagere) i min. 4 år.

Hver informantgruppe fik udleveret en kort, ikke-faglig tekst, som de skulle oversætte fra polsk til dansk. Efterfølgende skulle de hver især forfatte en kort, ikke-faglig tekst på dansk. Under opgaveløsningen skulle informanterne udfylde en protokol. Et uddrag fra protokol præsenteres i figur 1:³

Nr.	Hvad er grunden til, at du griber til et hjælpemiddel?	Det anvendte hjælpemiddel	Har du fundet den information, du søgte? (JA/NEJ)?	Var det nemt at finde informationen? (JA/NEJ) (Hvis NEJ, hvorfor?)
1.	Jeg ved ikke, hvad "królowa" hedder på dansk	Słownik MINIMUM	nej	Ordet "królowa" findes ikke
2.	Jeg ved ikke, hvad "już" hedder på dansk	ditto	ja	ja
3.	Jeg ved ikke, hvad "dostał" hedder på dansk	ditto	ja	ja
...
12.	Hvordan gradbøjer man den danske ækvivalent af adverbiet "chętnie"?	ditto	nej	I tillegget til ordbogen står der intet om adverbier (selv om andre ordklasser bliver omtalt)
...

Figur 1: Uddrag fra protokol anvendt ved oversættelse fra polsk til dansk.

Som det tydeligt fremgår af protokollen, registrerer vi ikke, hvilke data folk søger efter, men hvilket problem der er årsag til ordbogs-konsultationen.

Forud for undersøgelsen blev vores informanter orienteret om undersøgelsens design og formål, og de blev bedt om at medbringe alle de hjælpemidler, som de mente, de ville få brug for.

Undersøgelsen bygger også på en antagelse om, at PDO ville blive foretrukket, hvis brugerne fik adgang til den. Derfor fik halvdelen af informanterne i hver gruppe udleveret et eksemplar af ordbogen. Vi gjorde dog informanterne opmærksomme på, at de ikke var forpligtet til at bruge PDO.

3 Undersøgelsens metasprog er polsk. Al gengivelse af både vores materiale og informanternes data præsenteres her i vores egen oversættelse til dansk.

Efter løsningen af oversættelses- og tekstproduktionsopgaverne deltog fire informanter fra hver gruppe, hvoraf to havde haft adgang til PDO, i et interview. Interviewene er opbygget af to sæt åbne spørgsmål, der går på valg af ordbøger i forbindelse med (1) oversættelsen og (2) tekstproduktionen. De, der havde adgang til PDO, fik yderligere et par spørgsmål (se figur 2):

Spørgsmål:
1. Ud fra din protokol kan vi se, at du har brugt følgende hjælpemidler ved oversættelsen af den udleverede tekst fra polsk til dansk. Hvorfor anvendte du netop disse hjælpemidler?
2. Er der noget, som du ville tilføje eller ændre i de hjælpemidler, du har brugt? Hvad og hvorfor?
3. Har du yderligere kommentarer vedrørende de hjælpemidler, du har brugt ved oversættelsen fra polsk til dansk?
4. Ved oversættelsen fra polsk til dansk havde du mulighed for at anvende Jan Nowaks <i>Polsk-dansk Ordbog</i> . Har du benyttet denne mulighed? Hvorfor?
5. Mener du, at <i>Polsk-dansk Ordbog</i> er et godt hjælpemiddel for dig, når det gælder oversættelse fra polsk til dansk? Hvorfor?
6. Synes du, at <i>Polsk-dansk Ordbog</i> er bedre/værre end de andre hjælpemidler, du har anvendt ved oversættelsen fra polsk til dansk? Hvorfor?

Figur 2: Uddrag fra et interview vedrørende præferencer for brugen af ordbøger.

2.3. Metode: objektive vs. subjektive behov

Den tredje og sidste hypotese, som vores undersøgelse bygger på, er, at man metodologisk bør skelne mellem afklaring af subjektive og objektive informationsbehov (Tarp 2008). Subjektive behov er de behov, som brugeren selv erkender i den ekstraleksikografiske situation og forsøger at tilfredsstille. Objektive behov er alle de behov, der opstår i en given situation, uanset om brugeren erkender dem eller ej. Protokollerne og interviewene har vi brugt til at registrere brugernes erkendte, dvs. subjektive, behov. Opgaveretningen har vi til gengæld brugt til at søge de faktiske, dvs. objektive behov afdækket. Vores kombinerede metode har klart påvist, at bl.a. behovet for at finde danske ækvivalenter i mange tilfælde er blevet overset. Dette gælder faktisk for alle vores informanter.

3. Analyse af undersøgelsens resultater

I det følgende præsenterer vi resultaterne af vores undersøgelse. Da der er overlap i mange resultater for både oversættelse og tekstproduktion, har vi valgt at bringe disse samlet. Til gengæld har vi udarbejdet en særskilt præsentation af de resultater, der er afhængige af opgavetype.

3.1. Informationsbehov

Det er næppe overraskende, at det behov, der i langt de fleste tilfælde gælder, er behovet for at finde passende danske ækvivalenter til polske ord og udtryk. Behovet erkendes af samtlige informanter. Imidlertid viser vores opgaveretning, at behovet ikke blev opdaget i alle de tilfælde, hvor det opstod. Dette bekræfter, at det er nødvendigt at skelne mellem subjektive og objektive behov. Opgaveretningen viser også, at den hyppigste fejltype er forkert valg af ækvivalent. Dette taler naturligvis for udarbejdelse af nye, pålidelige bilingvale ordbøger mellem polsk og dansk.

Selv om behovet for andre informationskategorier (stavning, genus, pluralisformer, præpositionsforbindelser og verbernes bøjningsformer) kun forekommer sporadisk, kan dette registreres hos de fleste af vores informanter. Bilingvale ordbøger bør derfor også yde hjælp til dette.

3.2. Valg af ordbøger

Alle vores informanter bruger bilingvale ordbøger, for det meste mellem polsk og dansk og mellem dansk og engelsk samt, dog mindre hyppigt, mellem dansk og tysk og polsk og engelsk. Kun én monolingval ordbog blev brugt, og det var *Nudansk Ordbogs* elektroniske udgave (herefter NuDO-e). Tabel 1 viser det samlede antal opslag i de forskellige slags ordbøger for hver opgavetype.

ordbogstype	antal opslag	
	over-sættelse	tekst-produktion
bilingval: polsk-dansk/dansk-polsk	96	34
bilingval: engelsk-dansk/dansk-engelsk	35	21
bilingval: tysk-dansk/dansk-tysk	7	5
bilingval: polsk-engelsk	1	0
monolingval: NuDO-e	29	21
andet	2	2
kombineret	18	5

Tabel 1: Antal opslag i forskellige ordbogstyper i forhold til opgavetype.

Tabel 2 relaterer valg af ordbøger til brugerprofilerne:

ordbogstype	antal opslag		
	BP1	BP2	BP3
bilingval: polsk-dansk/dansk-polsk	74	32	24
bilingval: engelsk-dansk/dansk-engelsk	22	14	20
bilingval: tysk-dansk/dansk-tysk	11	0	1
bilingval: polsk-engelsk	0	0	1
monolingval: NuDO-e	0	0	50
andet	0	2	2
kombineret	1	3	19

Tabel 2: Antal opslag i forskellige ordbogstyper i forhold til brugerprofil.

Det viser sig, at brugeradfærden påvirkes af uddannelsestrinet: Jo højere trin, jo mindre tilbøjelig er man til udelukkende at bruge bilingvale ordbøger. BP1-informanterne bruger udelukkende bi-

lingvale ordbøger, helst med dansk og polsk. Det samme gør flere BP2-informanter, idet især PDO er en af de foretrukne ordbøger. Informanterne på højeste trin (BP3) benytter sig derimod hyppigt af den monolingvale NuDO-e, især for at løse problemer med grammatiske og morfologiske oplysninger og med stavning. Valget af monolingvale resurser bekræfter delvis, hvad andre forskere har fundet frem til. At mere sprogligt (og dermed også mere leksikografisk) bevidste brugere vælger monolingvale ordbøger oftere end brugere med lavere sproglig bevidsthed bekræftes bl.a. af Lew (2004).

BP3-informanterne udskiller sig igen ved systematisk at inddrage brug af bilingvale ordbøger med engelsk. Disse ordbøger bliver brugt næsten lige så hyppigt som ordbøger mellem dansk og polsk. Dette skyldes, ifølge informanterne selv, mangel på gode ordbøger til at oversætte direkte fra polsk. Nogle informanter har simpelt hen vænnet sig til at oversætte via et tredje sprog.

Tilgængeligheden – enten på universitetets bibliotek eller i ens private samling – havde også betydning for informanternes valg af ordbøger. Dette gjaldt for det meste bilingvale ordbøger med engelsk, tysk og polsk. Manglende adgang til konkrete ordbøger blev tit nævnt som grunden til, at informanterne ikke havde medbragt dem til undersøgelsen. Dette vedrørte for det meste PDO (som kun én af informanterne ejer, og som biblioteket kun har til læsesalsbrug) samt internetressurser, især DDO, KorpusDK og sproget.dk, som flere informanter mente, at de ville have brugt, hvis de havde haft adgang til dem.

Enkelte informanter angav derudover en række tekniske årsager til, at de brugte de pågældende ordbøger, heriblandt handyness (“ordbogen er lille og dermed nem at bære rundt på”), layout og elektronisk format.

3.3. De anvendte ordbøgers brugbarhed

Ideelle ordbøger findes der næppe nogen af. Dette gælder, uanset hvilke og hvor mange ordbøger man har valgt for at løse et konkret problem. Kun få af vores informanter har fundet brugbare løsninger på alle de problemer, de selv har opdaget (to ved oversættelse og fem ved fri tekstproduktion). Her har uddannelsestrinet, især for oversættelse, igen betydning for, hvor godt de er i stand til både at vælge den rigtige ordbog og få deres informationsbehov dækket. Jo større sprogbevidsthed, jo flere løsninger kunne man finde på sine problemer, og jo oftere viste løsningerne sig at være korrekte. Dette gjaldt, uanset hvilket hjælpemiddel man benyttede sig af.

De fleste fejl relateret til ordbogsbrug, som vi har registreret ved at samkøre data fra protokollerne og fra opgaveretningen, skyldes forkert valg af ækvivalent. Dette problem kan der være flere forklaringer på. Der kan være tale om, at den rigtige ækvivalent ikke findes i det pågældende opslagsværk, eller at informanten ikke har været i stand til at vælge rigtigt blandt flere angivne ækvivalenter. Hvis oversættelsen foregår via et tredje sprog, kan der også være tale om manglende kundskaber i dette sprog.

Forsøg på at dække andre informationsbehov (stavning, genus, bøjningsformer af substantiver og verber) viste sig ikke at volde nævneværdige problemer. Vores undersøgelse peger på, at det kan være vanskeligt at finde frem til de rigtige kollokationer og præpositionsforbindelser. Vores datamateriale er dog for spinkelt til at belyse problematikken.

Informanterne havde flere forslag til ændringer i de ordbøger, de brugte. De fleste ville have ordbøgerne udvidet. Dette betyder tilføjelse af opslagsord, ordforbindelser, eksempler på opslagsordenes kontekstbrug (især til de små bilingvale ordbøger med polsk, men det sidstnævnte gjaldt også PDO). Ellers var der forslag om tilføjelse af flere grammatiske og morfologiske data (for det meste genus samt substantivernes pluralisformer, verbernes

datidsformer og participiale former samt adjektivernes og adverbiernes gradbøjningsformer) og et enkelt ønske om en elektronisk udgave af PDO. En af informanterne ville også supplere NuDO-e med data fra internetressurser.

3.4. PDO's påvirkning af brugeradfærd

Som tidligere nævnt fik halvdelen af informanterne udleveret PDO. Ved begge opgaver har ordbogen klart haft indvirkning på brugeradfærden i alle informantgrupperne. Hos BP3-informanterne var ordbogen den eneste mellem dansk og polsk, som de valgte at bruge. PDO blev dog brugt i mindre omfang end hos de tilsvarende BP1- og BP2-informanter.

Mange informanter anfører, at de har brugt PDO, fordi ordbogen er stor (dvs. har mange opslagsord). Det er derfor ikke overraskende, at ordbogen næsten udelukkende blev brugt til at finde danske ækvivalenter til polske ord. Til dette formål synes den at være ganske god, især for brugere på de højere uddannelsesstrin. I de allerfleste tilfælde var informanterne i stand til at finde en ækvi-valent, der var anvendelig i den pågældende kontekst. Vores objektive og subjektive målinger viser dog, at PDO hverken er markant bedre eller dårligere end de andre ordbøger, som blev anvendt til at løse samme problem.

3.5. Oversættelse og fri tekstproduktion – forskelle i brugeradfærd

Overordnet set ligner de typer informationsbehov, der opstår ved fri tekstproduktion, de typer behov, som opstår under oversættelsen. Alle informanterne greb til ordbøger, hovedsageligt fordi de ikke kendte de danske ord til at udtrykke de ideer, som de havde formuleret på polsk. Dette bekræfter, at fri tekstproduktion på fremmedsprog ofte indebærer en skjult oversættelsesfase (jf.

Hausmann og Werner 1991:2741). Opgaveretningen viser dog, at behovet opstår betydeligt sjældnere end ved oversættelse (vi har registreret det i alt 140 gange ved tekstproduktion og 296 gange ved oversættelse). Analyser af protokollerne peger i samme retning. Samlede antal foretagne opslag viser, at behovet for en ækvi-valent kun blev erkendt 55 gange ved tekstproduktion mod hele 121 gange ved oversættelse. Dette bekræftes af informanternes udsagn om, at de ved fri tekstproduktion hovedsageligt benytter sig af det ordforråd, de kender i forvejen. Det er tankevækkende, at antallet af opslag falder ved tekstproduktion, også set i lyset af at de frit producerede tekster uden undtagelse var markant længere end oversættelserne. Man kunne nemlig forvente, at en længere tekst vil medføre flere ordbogsopslag.

På trods af at de typer informationsbehov, som forekommer ved fri tekstproduktion, i høj grad overlapper med de typer behov, som opstår ved oversættelse, er der væsentlige forskelle i den grad, de bliver erkendt af informanterne. Generelt er erkendelsesgraden meget lav ved tekstproduktion. Bortset fra behovet for at finde et passende dansk ord – hvad hedder det på dansk? – blev behovene kun sjældent opdaget af informanterne. Endvidere, mens evnen til at erkende egne behov synes at stige med uddannelsestrinet i forbindelse med oversættelse, kunne vi slet ikke observere samme tendens ved fri tekstproduktion. Dette forhold bør efter vores mening indgå i fremtidig leksikografisk teoridannelse og praksis. Der er ingen tvivl om, at de behov, som opstår sjældent, men hos mange brugere inden for en bestemt brugerprofil, bør kunne dækkes ved hjælp af passende ordbøger. Spørgsmålet er så, hvad leksikografen skal stille op med de behov, som kun sjældent erkendes, men som ikke desto mindre er helt centrale i den pågældende type situation. Man kan nok lave en behovstilpasset ordbog, men det er svært at argumentere for, at brugere ville gribe til den, medmindre de opdager, at de rent faktisk har behov for den.

Vi ville gerne finde ud af, hvad årsagen til det markant lavere

opslagsantal ved fri tekstproduktion kunne være. Vi spurgte alle de interviewede informanter om det. Alle svarene var ens. Informanterne mente, at det skyldtes, at de ved fri tekstproduktion udtrykte deres egne tanker, og at de derfor kunne udtrykke sig, som de havde lyst til. Dette skyldes naturligvis, at de bruger de ord og vendinger, som de i forvejen kender og er fortrolige med. Det fører automatisk til færre tilfælde af opslag. Denne antagelse står lidt i modsætning til dataene fra protokollerne, hvor mangel på en passende ækvivalent blev angivet som den hyppigste årsag til, at informanterne brugte ordbøgerne.

Selv om ordbøgerne blev brugt markant sjældnere ved tekstproduktion end ved oversættelse, skete det med bedre resultater. I Tabel 3 viser vi det samlede antal opslag foretaget ved begge opgaver og anfører, hvor mange der var vellykkede (dvs. informanten har fundet frem til løsningen på sit problem). Disse data, samlet fra protokollerne, har vi suppleret med data fra opgaveretningen, hvorved vi afdækkede, hvor mange af de fundne løsninger der var brugbare (dvs. om den fundne information passede til formålet, idet det blev brugt rigtigt i teksten).

	oversættelse	tekstproduktion
antal opslag	181	87
heraf vellykkede	151 (83,4%)	78 (89,7%)
heraf brugbare	117 (64,6%)	62 (71,3%)

Tabel 3: Udbytte af ordbogsbrug i forhold til opgave.

De samlede tal viser en generel forskel på oversættelse og fri tekstproduktion, men skjuler en anden vigtig faktor, nemlig at antallet af vellykkede og brugbare opslag varierer med uddannelsestrinet i højere grad ved oversættelse end ved fri tekstproduktion. Tabel 4 viser, hvordan de samlede resultater fordeler sig på de tre brugerprofiler:

	oversættelse			tekstproduktion		
	BP1	BP2	BP3	BP1	BP2	BP3
antal opslag	76	33	72	28	13	46
heraf vellykkede	57 (75,0%)	24 (72,0%)	70 (97,2%)	23 (82,1%)	12 (92,3%)	43 (93,5%)
heraf brugbare	38 (50,0%)	21 (63,6%)	58 (80,6%)	19 (67,9%)	10 (76,9%)	33 (71,7%)

Tabel 4: Udbytte af ordbogsbrug i forhold til opgave og brugerprofil.

Tallene viser, at jo højere uddannelsesstrin er, jo bedre udbytte får man af at bruge ordbøgerne; dog er disse forskelle ikke så tydelige ved fri tekstproduktion som ved oversættelse. Samme tendens gælder erkendelsen af egne informationsbehov. BP3-informanter erkender deres behov langt hyppigere ved oversættelse end ved tekstproduktion.

4. Slutbemærkninger

Formålet med denne artikel var at afdække de polske universitetsstuderendes ordbogsbrug ved oversættelse fra polsk til dansk og fri tekstproduktion på dansk. Vi valgte ikke at præsentere vores subjektive vurdering af en række ordbøger mellem polsk og dansk, men besluttede i stedet at gennemføre en empirisk undersøgelse af de to situationer, hvor nogle af disse ordbøger ville blive taget i brug.

Objektivt set synes de ordbøger, der blev brugt, ikke at passe særlig godt til brugernes behov. Vores undersøgelse, især opgaveretningen, påviser, at alle informanterne har begået en del sproglige fejl. Nogle af disse fejl hidrører fra manglende sprogkompetence generelt. Imidlertid kan mange fejl spores tilbage til de anvendte ordbøgers kvalitet (eller mangel på samme) – her kan der være tale om både deciderede fejl og mangler i ordbogen, og om brugerens utilstrækkelige evne til at benytte den. Informanternes subjektive

vurdering af ordbøgernes kvalitet peger i samme retning. De er nemlig også selv klar over, at de ordbøger, de har valgt at bruge, ikke yder den optimale hjælp.

Undersøgelsen bekræfter til fulde, at det er en hvad-hedder-det-på-dansk-tankegang, som styrer informanternes leksikografiske adfærd, ikke bare ved oversættelse, men også ved fri tekstproduktion. Derfor er det næppe overraskende, at informanterne langt hen ad vejen i deres studieforbøb foretrækker at bruge bilingvale ordbøger (dog ikke på øverste trin i uddannelsesforløbet). Det undrer os ikke, at de informanter, som fik udleveret et eksemplar af PDO, brugte den flittigt – ordbogen indeholder nemlig langt flere opslagsord end alle andre ordbøger mellem dansk og polsk til sammen. Desværre indeholder den næsten ingen af de øvrige datakategorier, som vores informanter havde brug for ved oversættelse til dansk eller ved fri tekstproduktion på dansk, først og fremmest genusangivelser, substantivernes pluralisformer, verbernes og adjektivernes bøjningsformer. Tilgangen til data vedrørende stavning af danske ord eller præpositionsforbindelser synes at være ret besværlig, da det først er nødvendigt at finde det rigtige polske opslagsord (og håbe på, at det pågældende danske ord eller udtryk angives). For at løse problemer vedrørende denne slags sproglige data valgte nogle brugere at benytte sig af en monolingval ordbog, men det har vi, som allerede nævnt, kun observeret på øverste uddannelsestrin.

Det utilstrækkelige udbud af pålidelige og funktionelle ordbøger mellem dansk og polsk har også andre konsekvenser. Der anvendes ordbøger mellem dansk og hhv. engelsk eller tysk. Dette fører til, at det mest udbredte informationsbehov ved oversættelse og tekstproduktion, nemlig behovet for at finde en dansk ækvivalent til et polsk ord, ikke altid kan tilfredsstilles, fordi informanterne ikke er fuldt fortrolige med det pågældende tredjesprog. Derudover er disse ordbøger normalt konciperet til danske brugere, hvilket betyder, at nogle data vedrørende danske ord, så som genus og bøjningsendelser af substantiver, ikke forefindes.

Vores leksikografiske iagttagelser stammer fra en eksplorativ undersøgelse og må dermed betragtes som foreløbige observationer. Vi ser frem til mere omfattende brugerundersøgelser, hvor datagenerering og samkøring af forskellige dataindsamlingsmetoder kan kaste mere lys over relationerne mellem objektive og subjektive informationsbehov og dermed underbygge vores resultater eller udfordre dem. Systematiske brugerstudier af de leksikografisk relevante aspekter af oversættelse og af fri tekstproduktion på fremmedsprog vil forhåbentlig kaste endnu mere lys over forskelle og ligheder i de to situationer og bidrage til at generere ny praktisk og teoretisk viden om tekstproduktionsprocesser i et interlingvalt perspektiv.

Litteratur

Ordbøger og øvrige opslagsværker

- Bostrup, Lise og Bogusława Sochańska 2006: *Aktivt dansk. Język duński dla początkujących, gramatyka i słownictwo*. København: Alfabet.
- Frank-Oborzynska, Elżbieta 1970: *Słownik minimum duńsko-polski i polsko-duński* (1. udgave; 7. udgave i 2008). Warszawa: Wiedza Powszechna.
- Kuciewicz, Magdalena (uden årstal): *Słownik polsko-duński i duńsko-polski*. Kraków: Level Trading.
- Michalska, Urszula og Maja Król 2007: *Rozmówki polsko-duńskie*. Warszawa: Kram.
- NuDO-e = Christian Becker-Christensen (red.) 2010: *Politikens Nudansk Ordbog* (21. udgave, med cd-rom). København: Politikens Forlag.
- Okoniewski, Grzegorz 1999: *Dansk-Polsk Juridisk-Merkantil Ordbog*. København: Gads Forlag.

- PDO = Jan Nowak 2010: *Polsk-Dansk Ordbog* (2. udgave). København: Gyldendal.
- Rajnik, Eugeniusz 1999: *Gramatyka języka duńskiego*. Poznań: Wydawnictwo Naukowe UAM.
- Schothuis, Teresa Jaskólska og Erik Helbo (uden årstal): *Duński od podstaw, część 1. Polsk Billedordbog, del 1*. Dąbrowa Górnicza: Wydawnictwo Granpok System.
- Stemann, Ingeborg 2008: *Polsk-Dansk, Dansk-Polsk Ordbog* (14. oplag, uændret optryk fra 1939 og 1949). København: Jørgen Paludans Forlag.
- Sørensen, Wanda Strange og Lilly Widding 1989: *Dansk-Polsk Ordbog* (1. udgave; 7. oplag i 2009). Herning: Special-Pædagogisk Forlag.

Anden litteratur

- Duvå, Grete og Anne Lise Laursen 1994: Translation and LSP Lexicography. I: Burkhard Schaefer og Henning Bergenholtz (red.): *Fachlexikographie. Fachwissen und seine Repräsentation in Wörterbüchern*. Tübingen: Gunter Narr Verlag, 247-268.
- Frankenberg-Garcia, Ana 2005: A Peek into What Today's Language Learners as Researchers Actually Do. I: *International Journal of Lexicography* 18.3, 335-355.
- Hausmann, Franz Josef og Reinhold Otto Werner 1991: Spezifische Bauteile und Strukturen zweisprachiger Wörterbücher: eine Übersicht. I: Franz Josef Hausmann, Oskar Reichmann, Herbert Ernst Wiegand og Ladislav Zgusta (red.): *Wörterbücher/Dictionaries/Dictionnaires. Ein internationales Handbuch zur Lexikographie [...]*. Berlin / New York: Walter de Gruyter, 2729-2769.
- Humblé, Philippe 2001: *Dictionaries and Language Learners*. Frankfurt am Main: Haag und Herchen.

- Lew, Robert 2004: *Which Dictionary for Whom? Receptive Use of Bilingual, Monolingual and Semi-bilingual Dictionaries by Polish Learners of English*. Pozna : Motivex.
- Lew, Robert 2011: Studies in Dictionary Use: Recent developments. I: *International Journal of Lexicography* 24.1, 1-4.
- Nord, Britta 2002: *Hilfsmittel beim Übersetzen. Eine empirische Studie zum Rechercheverhalten professioneller Übersetzer*. Frankfurt am Main: Peter Lang.
- Pilke, Nina 2008: Ordboksanvändning hos blivande språkexperter och hos professionella översättare. I: *LexicoNordica* 15, 135-153.
- Tarp, Sven 2006: *Leksikografien i grænselandet mellem viden og ikke-viden. Generel leksikografisk teori med særlig henblik på lærerleksikografi*. Doktorafhandling. Århus: Center for Leksikografi.
- Tarp, Sven 2008: Kan brugerundersøgelser overhovedet afdække brugernes leksikografiske behov? I: *LexicoNordica* 15, 5-32.
- Tono, Yukio 2001: *Research on Dictionary Use in the Context of Foreign Language Learning (Lexicographica Series Maior 106)*. Tübingen: Niemeyer.
- Varantola, Krista 1998: Translators and Their Use of Dictionaries: User Needs and User Habits. I: Beryl T. Sue Atkins (red.): *Using Dictionaries: Studies of Dictionary Use by Language Learners and Translators*. Tübingen: Niemeyer, 179-192.

Internethenvisninger

- DDO = *Den Danske Ordbog. Moderne dansk sprog*. Det Danske Sprog- og Litteraturselskab. <http://ordnet.dk/ddo/> (april 2011).
- KorpusDK = *KorpusDK. Sproget i brug*. Det Danske Sprog- og Litteraturselskab. <http://ordnet.dk/korpusdk> (april 2011).
- Sproget.dk = *Sproget.dk. Indgangen til det danske sprog*. <http://sproget.dk/> (april 2011).

Welker, Herbert Andreas 2010: *Dictionary Use: A General Survey of Empirical Studies*. Brasília. http://www.let.unb.br/hawelker/dictionary_use_research.pdf (april 2011).

Marcin Overgaard
Ptaszynski
adjunkt, ph.d.
Center for Leksikografi
Aarhus Universitet
Fuglesangs Allé 4
DK-8210 Aarhus V
maop@asb.dk

Mikołaj Sobkowiak
adjunkt, ph.d.
Katedra Skandyna-
wistyki
Uniwersytet im. Adama
Mickiewicza
al. Niepodległości 4
PL-61-874 Poznań
miksobko@amu.edu.pl

Patrick Leroyer
lektor, ph.d.
Center for Leksikografi
Aarhus Universitet
Fuglesangs Allé 4
DK-8210 Aarhus V
pl@asb.dk

Partiklar inn i ordbøkene? Om tradisjonell og nyare kategorisering av preposisjonar

Leiv Inge Aa

Norsk Ordbok (NO) (The Norwegian dictionary of spoken language and the Nynorsk written language) and also other dictionaries use traditional terminology for grammatical categories, while in most contemporary linguistic literature and e.g. *Norsk referansegrammatikk* ('The Norwegian reference grammar', NRG), more up-to-date terms are used. In this article, I look at the distinction between *prepositions* and *adverbs* in traditional and modern terminology, and discuss whether NO would be better off adopting the NRG system. In short, my answer is *yes*, with some provisos. I advocate the use of the term *particle* where NO refers to a preposition switching to an adverb. I also question the use and relevance of the term *intransitive preposition* (from NRG) in a dictionary setting. Eventually, I discuss some problems connected with introducing new normative terms in a descriptive dictionary. Since NO is a project that has already existed for many years (first volume printed in 1966) and is completed in three years, I do not see a change of grammatical terms as an option for the printed dictionary. However, the discussion is of great importance in principle – and of practical importance for the upcoming NO (or any) online dictionary.

1. Innleiing

Større ordboksverk som *Norsk Ordbok* (NO) og SAOB møter på utfordringar av mangt slag, då mykje har endra seg i språk og samfunn sidan ein i si tid tok fatt på *a*. NO sitt materialarbeid starta på 1930-talet, og det første bandet, som dekte *a-doktrinær*, vart trykt i 1966. Ikkje berre har det kome endringar i ordtilfanget i den ak-

tuelle alfabetbolken sidan, men over ein så (høvesvis) lang periode kjem det gjerne òg endringar i terminologi. Ordklasseinndelinga som NO nyttar, er den same i 2011 som ho var i 1966, men skulegrammatikken har innført nokre endringar i mellomtida. Den viktigaste ålmennterminologiske milestolpen gjennom desse åra er kanskje likevel utgjevinga av *Norsk referansegrammatikk* (NRG), som nyttar andre inndelingskriterium enn dei tradisjonelle. Til dømes høyrer dei tidlegare possessive og demonstrative pronomena til under *determinativa* i NRG, og underordna konjunksjonar skil seg så mykje frå dei sideordna at dei (underordna) er vortne omdøyppte til *subjunksjonar*. Også det gamle infinitivmerket *å* er kome med i denne kategorien. Endringa eg vil diskutere her, er likevel ikkje ei av dei ovannemnde, men snarare omgrupperinga av ei rekkje adverb til preposisjonskategorien. I NRG tenkjer ein seg at preposisjonar slik som verb varierer mellom å vere transitive og intransitive; nokre tek utfylling og andre ikkje – og dei fleste kan variere mellom ei transitiv og ei intransitiv realisering. Slik får vi ei gruppe med typisk transitive preposisjonar (som rett nok òg blir realiserte intransitivt), som svarar til dei gamle, «ekte» preposisjonane, dvs. dei som blir noterte som preposisjonar i tradisjonelle grammatikkar og typisk i ordbøker. Døme på desse er *av, bak, blant, etter, for, hjå, med, mellom, over, på, til, under, ved* og fleire. Av dei typisk intransitive preposisjonane finn ein ei stor gruppe lokativar (*borte, framme, heime, inne, nede, oppe, ute ...*) og ei mindre gruppe med direksjonale nemningar (*att, bort, fram, heim, hen, vekk*) (sjå meir om inndelinga i NRG s. 411 ff.). Til den siste gruppa må ein òg plassere *inn, ut, opp* og *ned*, sjølv om desse har transitiv bruk i tillegg. Nettopp det siste er det viktige poenget i denne samanhengen. Ei rad av desse funksjonsorda blir realiserte både med og utan utfylling, noko som gjev grunn for eit ordklassebyte i NO, medan NRG vil kalle dei transitive og intransitive variantar av det same ordet.

Innan lingvistikken er NRG-termene alt godt etablerte; pre-

posisjonar blir analyserte både med og utan utfylling. Det bør heller ikkje overraske ein filolog av noko slag at termar endrar seg, slik bruken av eit ord gjer. Kva skal så ein ordboksredaksjon gjere med dei? Gjev dei nye termene ei betre og meir sameinande framstilling av funksjonsorda enn dei gamle? Kan det nye omgrepsapparatet gje oss så vel faglege som pedagogiske føremoner? Desse spørsmåla kjem ikkje berre beint ut av mitt eige hovud, men kan seiast å ha vorte aktualiserte av Jan Terje Faarlund sjølv (den eine NRG-forfattaren) i ei grammatisk forvitneleg melding av det åttande bandet av NO (*mugg-ramnsvart*) i *LexicoNordica* 17 (2010). Faarlund tek opp fleire problem enn eg vil gjere her, og han argumenterer for at NO kunne teke i bruk NRG-systemet ved oppstarten av prosjektet NO 2014 for om lag ti år sidan. At eg er deltidstilsett i NO sjølv, vil ikkje seie at eg har til føremål å forsvare praksisen der. Tvert imot har eg til liks med Faarlund grammatisk bakgrunn og er van med kategoriseringar som liknar NRG si. Målet med denne artikkelen er å vurdere kva som er den beste prinsipielle (og pedagogiske) løysinga når ein kategoriserer preposisjonar (og adverb) i ei ordbok. Etter kvart vil eg argumentere for at *partikkelomgrepet* kan vere tenleg å nytte seg av. Eg tek utgangspunkt i NO fordi det er den ordboka eg kjenner best, og diskuterer til slutt kva som er dei beste praktiske løysingane vidare når ein allereie har godt etablert det tradisjonelle systemet. Diskusjonen bør såleis vere relevant både for etablerte og nyoppstarta papirordbøker, og særleg for nettordbøker.

2. Ulike (tenkte) ordboksføringar

Føringa for lemma som varierer mellom å bli realiserte som preposisjonar og adverb, er ikkje konsekvent i NO, og kan nok forvirre både filologar og andre. Er materialet til eit ord stort nok, slik som med *på* og *i*, synest preposisjons- og adverbbruken å bli ho-

mografseparerte, og då får adverbet romartal I, og preposisjonen romartal II (jf. døme 1 under). Hierarkiordninga er fast, uavhengig av materialfordelinga. Har ordet mindre materiale som er jamt fordelt på dei to bruksmåtene (jf. samansetjinga *oppover* i døme 2), står dei to under same oppslag, med preposisjonen under tyding A og adverbet under tyding B (altså motsett av rekkjefølgja for homografane). Når det gjeld dei retningsorda som oftast er realiserte utan utfylling, men somtid med, står adverbtydinga først (A), følgt av ei meir marginal preposisjonstyding (B) (døme 3). Her styrer altså materialet rekkjefølgja. Dei ulike føringsvariantane i NO er skisserte under:¹

- (1) **I** på adv (49 liner): *Johan sette hatten på.*
II på prep (759 liner): *Johan sette hatten på hovudet.*
- (2) **oppover** prep A (22 liner): *dei gjekk oppover lia.*
B adv (25 liner): *koma, reisa, bera oppover.*
- (3) **opp** adv A (418 liner): *gå, stiga, flyga opp.*
B prep (8 liner): *laksevandringa opp fossen.*

Inkonsekvensen her er ikkje heilt uproblematisk. Motivasjonen for å splitte opp det store oppslaget *på* må vere for at det skal vere lettare å finne adverbvarianten i ein artikkel som fyller fleire sider. Konsekvensen er at *I på* og *II på* gjerne blir tolka som homonym, noko som er å ta vel hardt i. No markerer ikkje alltid ein ny homograf at vi har med homonymtilfelle å gjere, men det er lett å tolke det slik spontant. I tilfella *opp* og *oppover* kan ein tenkje seg at ein har ein «primærordklasse» i tyding A, og at ein finn ein meir marginal variant i tyding B, sjølv om dei er jamstore i *oppover-*

1 I dei første banda av NO kunne ein føre fleire ordklassar under same oppslaget, t.d. **bak** prep, adv og **for** prep, adv, konj. Denne praksisen har ein lagt bak seg.

tilfellet. I alle høve er det pragmatiske kriterium som har vorte teke i bruk når *på*, med sitt enorme materiale, er behandla ulikt det meir overkomelege *oppover*. Denne prinsipielle inkonsekvensen er uheldig, då ein slik føringsregel er vanskeleg å gjennomskode for andre enn dei som lagar regelen. I ei ordbok som var basert på NRG-inndelinga, kunne ein tenkje seg følgjande oppslag for *på*:

- (4) **på** prep A tr: *Johan sette hatten på hovudet.*
 B intr: *Johan sette hatten på.*

Ved å kategorisere begge variantane som preposisjon, har ein den føremonen at det ville vere unaturleg å homografseparere dei to variantane, og den spontane reaksjonen min på det er positiv. Med (4) kan det sjå ut til at vi har fått ei meir sameint beskriving av *på* enn i (1).

3. Føremoner med NRG-systemet?

Det er ein openberr pedagogisk føremon å kunne generalisere preposisjonar på same måte som verb. Om vi har eit system som går att, blir det mindre å lære, og dess meir kan ein forstå. Likevel kan ein spørje seg om termen *intransitiv* ('utan objekt') er eigna for *preposisjonar* ('det å stå framfor noko'). Samanstillinga *intransitiv preposisjon* blir såleis eit paradoks. Dette er så opplagt at NRG sjølv òg kommenterer det (s. 412), men avviser det som eit problem, då ordklassetermane på det jamne langt frå er gjennomsiklige. Det har dei sjølvsagt rett i, men så er det slik at *preposisjon* til no var ei av dei termene som *gav* mening. Spesielt påfallande kan det kanskje synast ved dei gruppene av ord som vil bli kategoriserte som *obligatorisk intransitive preposisjonar*, jf. rada av dei lokative og direksjonale orda nemnde i innleiinga. Dette treng ikkje vere eit stort problem, men paradokset blir òg nemnt i den lingvistiske

innføringsboka Bye, Trosterud & Vangsnes (2003:303), der ein løyser problemet med å nytte den nøytrale, kategorifleksible nemninga *partikkel*. Eg skal kome attende til at det kan vere ein eigna utveg òg for ordbøkene. Men det kan vere verdt å merke seg at Ivar Aasen òg omtala preposisjonar med og utan objekt i grammatikkane sine (jf. Aasen 1848:§ 334). Det ville då gjelde dei «ekte» preposisjonane i tradisjonell terminologi, som kan realiserast utan utfylling. Retningsorda som *opp*, *ned*, *ut* og *inn* gjeikk under adverb hjå Aasen, sjølv om desse som nemnt òg kan realiserast *med* utfylling.

Eit anna reint pedagogisk problem er at preposisjon/adverb-skiljet som NRG nyttar, *ikkje* vart teke omsyn til i Språkrådet (den gongen Norsk språkråd) og Utdanningsdirektoratet si tilråding om «Grammatiske termer til bruk i skoleverket» (mai 2005), som skulle vere retningsgjevande for læreplanen frå 2006. Det må rek-nast som eit visst nederlag for NRG at mykje av terminologien og inndelingskriteria dei nyttar, primært er (og blir) noko fagfolk kjenner til. Utan at eg kan byggje under denne påstanden anna enn med stikkprøvar, tvilar eg sterkt på at lektorar rundt omkring kjenner til NRG-systemet særleg godt heller.

Eit siste, pedagogiske poeng er spørsmålet om det verkeleg er så avgjerande om eit ord som viser seg å vere syntaktisk og semantisk mangslunge, må beskrivast med same ordklassen i alle tilfella. NRG har òg tilfelle av ord som er preposisjonar i eitt tilfelle, men som må kategoriserast som subjunksjonar i andre (der dei innleier leddsetningar). I ei NRG-basert ordbok ville *innan* og *til* kanskje fått denne representasjonen:

- (5) **innan** prep A *innan jul* B subj *innan du kjem heim, er julematen klar.*
- (6) **til** prep A *til jul* B subj *julematen er klar til du er heime.*

B-tydingane i slike tilfelle er klassifiserte som konjunksjonar i NO, men dei må altså skifte ordklasse òg i NRG-systemet.

4. *Setje hatten på* – frå (in)transitivitet til partikkelnemning

Eg vil gå attende til den «ekte» preposisjonen *på*, og gjennom syntaktiske testar sjå korleis den transitive og intransitive varianten oppfører seg overfor ein annan. Vi tek utgangspunkt i setningsparet frå (1) og (4) ovanfor, repetert som (7):

- (7) a. Johan sette hatten på.
b. Johan sette hatten på hovudet.

Den første påfallande ulikskapen mellom *på* i (7a) og (7b) er at berre den første varianten, den intransitive, kan stå framfor *hatten*:

- (8) a. Johan sette på hatten.
b. *Johan sette på hatten hovudet.

Faktisk er det slik at ein i dei fleste norske dialektane (særleg på Austlandet, i Trøndelag og eit stykke nedover Vestlandet) føretrekkjer å ha den intransitive varianten framfor objektet (Aasen 1848:§ 334–335, 1864:§ 334, Western 1921:§ 454, Sandøy 1976, 1991:102).² Det vil seie at (8a) då er den umarkerte ordstillinga – medan (7a) er markert, og for mange faktisk heilt ubrukbar. Og sidan utfyllinga i (8b) er heilt umogleg, kan begge døma i (8) gje oss ein indikasjon på at også (7a) og (7b) er meir ulike enn å ha eigenskapane ±utfylling.

Noko anna som kunne tyde på at den intransitive og transitive varianten er distribuerte ulikt, er at den venstredistribuerte varianten i (8a) kan stå saman med den transitive varianten i (7b),

2 Eg refererer til *hatten* som «objektet» her, som er tradisjonell terminologi. I lingvistisk litteratur er det rett nok like vanleg å sjå på han som eit subjekt i den resultatative, reduserte setninga (småsetninga) *hatten på (hovudet)* – som i sin heilskap fungerer som objekt til *sette*.

men *ikkje* med den høgre-distribuerte intransitive varianten i (7a):

- (9) a. (8a + 7a): *Johan sette på hatten på.
 b. (8a + 7b): Johan sette på hatten på hovudet.

(9b) er ikkje vanleg i skrift, men heilt gjengs i munnleg mål. Dette kan ein tolke slik at (9a) har same *på* brukt to gonger, medan (9b) har to ulike *på*. Men om intransitive og transitive *på* har ulik distribusjon, skulle ein òg vente at den høgre-distribuerte intransitive varianten i (7a) kan stå saman med den transitive i (7b) – noko som slett ikkje er umogleg i munnleg mål:

- (10) (7a + 7b): Johan sette hatten på på hovudet.³

Det kan altså sjå ut som at *på* (og andre preposisjonar) har fleire og andre viktige eigenskapar enn å ta ei nominal utfylling eller ikkje. Men varianten som ikkje tek nominal utfylling, og som kan stå på begge sider av objektet, *kan* ta ein preposisjonsfrase (PP) som utfylling, slik (9b) og (10) syner. Det er vanleg å referere til den *på*-varianten som anten tek PP eller ingenting til utfylling, som *partikkel*, og vi skal sjå vidare at det kan vere ein fleksibel og semantisk eigna term å nytte seg av for ei ordbok.

5. Partiklar

Døma i (7–10) kan alle kallast *resultative*. I det ligg det at *på* i (7a), *på hovudet* i (7b) og *på på hovudet* i (10) oppstår som resultat av verbalhandlinga. *Hatten* kjem/blir *på* (*på hovudet*) som resultat av setjinga. I døma under tyder sløyfeparentesane {1},{2} at 1 og 2 al-

3 For dialektane på Austlandet og i Trøndelag (særleg dei nær svenskegrensa) som har store problem med (7a) i utgangspunktet, vil (10) òg vere tilsvarende problematisk.

ternerer, og aldri opptrer saman, sjølv når ingen av dei er markerte som ugrammatikalske (jf. føringa i Svenonius 1996).

- (11) a. Enkel partikkelkonstruksjon:
 Johan sette {på} hatten {på}.
- b. Kompleks partikkelkonstruksjon:
 Johan sette {på} hatten {på} på hovudet.
- (12) Preposisjonskonstruksjon:
 Johan sette hatten på hovudet.

Medan partikkelkonstruksjonar oftast er resultative, gjeld ikkje det på langt nær for alle preposisjonskonstruksjonar. (12) er direksjonal; den nominale utfyllinga til *på* er då målet eller resultatet til *hatten*. Har vi ein *lokativ* variant av same setninga, *Johan held hatten på hovudet*, er *hatten* lokalisert i (eller på) utfyllinga til preposisjonen (jf. Svenonius 2008). Og nettopp det lokative (lok) versus det resultative (res) er i mange tilfelle det som skil partikkelkonstruksjonar frå preposisjonskonstruksjonar:

- (13) a. Res: Ta [PART. på] bordet (= dekkje (maten) på bordet)
 b. Lok: Ta [PP på bordet] (= vere borti bordet)
- (14) a. Res: Skubbe [PART. frå] båten (slik at båten fer frå (land))
 b. Lok: Skubbe [PP frå båten] (den som skubbar, sit i båten)
- (15) a. Res: Vente [PART. på] bussen (at bussen skal kome)
 b. Lok: Vente [PP på bussen] (den som ventar, sit på bussen).

Slike minimale par finst det fleire av, og den viktigaste semantiske skilnaden er altså det resultative i a-versjonane mot det lokative i

b-versjonane. Men i mitt eige arbeid med verbpartiklar i norske målføre (Aa under arbeid) analyserer eg faktisk berre den høgre-distribuerte partikkelen (jf. 7a) som reint resultativ, medan den tilsvarande venstredistribuerte partikkelen (jf. 8a, 13a, 14a, 15a), som altså står tettare inntil verbet, er meir *dynamisk*. Skilnaden finst det gode empiriske grunnar til å motivere. Sidan ein venstredistribuert partikkel blir meir dynamisk og dermed «verbaktig», kan han kanskje likevel forståast som eit adverb? Det mest nærliggjande svaret er nok *nei*, for vi kan òg ha eintydige *adjektiv* som venstredistribuerte partiklar. Eit adjektiv får inkjekjønnsbøying når det blir nytta adverbialt, men som partikkel held det på adjektivegenskapane sine. Det blir anten samsvarsbøygd med objektet (oftast når det står til høgre), eller opptrer i ubøygd form (oftast når det står til venstre).

- (16) a. Dei sleppte {laus} hestane {lause}⁴ (Heggstad 1931)
 b. Slepp {fri} dyret {fri(tt)}
 c. Vaske {rein(t)} huset {reint}
 d. Måle {blå} bilen {blå} (Åfarli 1985)

I litteraturen er slike konstruksjonar omtala litt ulikt. Heggstad (1931:§ 609) skriv at adjektivpartiklar («objektspredikatsordet» i terminologien hans) «kann stundom binda seg nærare til verbet og kann då kome fyre objektet og missa bøygningi.» Sandøy (1976:104) hevdar at adjektivet *blir* samsvarsbøygd i begge tilfella i romsdalsmål. (16c) har eg sjølv nytta i utspørjingar av informantar rundt omkring i Trøndelag, og her har det vore lite semje hjå talarane om dei samsvarsbøyer den venstredistribuerte partikkelen eller ikkje. Men at det er ei samsvarsbøying det er snakk om, og ikkje eit adverb, er eg temmeleg viss på. Dersom den venstredistribuerte partikkelen var eit adverb, skulle vi vente inkjekjønns-

4 Heggstad sitt døme er nok ikkje det aller beste, då *laus*, slik som *fri*, ofte står ubøygd til høgre òg.

bøying på han ved alle tilfelle, òg i (16d).⁵ (16) illustrerer dermed godt at partiklar kan nyttast som ei *kategorifleksibel* nemning – dei er ikkje bundne til ein ordklasse. Dette kan gje oss føremoner i ordboksframstillinga. I neste del diskuterer eg korleis ein kunne innlemme partikkelnemninga naturleg i ein preposisjonsartikkel. Eg går ut frå at ho kan nyttast uproblematisk på tilsvarande måte i adjektivartiklar, sjølv om desse blir tona ned.⁶

6. Partiklar i ordboka

Dersom vi innfører det kategorifleksible partikkelomgrepet i ordboka, trur eg det kan få gode konsekvensar særleg for preposisjonsartiklane. Det viktigaste poenget er då at vi slepp å endre ordklasse undervegs i same artikkelen, og at vi får kategorisert ei semantisk *resultativ* tyding. Ein kunne likevel sjå føre seg eit anna scenario, der preposisjonar vart inndelte etter om dei hadde utfylling eller ikkje. Då ville ein måtte kategorisere dei komplekse partikkelkonstruksjonane saman med preposisjonar som tek nominal utfylling, og preposisjonane utan utfylling måtte stå for seg, jf. (17):

- (17) på prep A tr (med nominal utfylling + komplekse partikkelkonstruksjonar) B intr (enkle partikkelkonstruksjonar)

Dette kan vere ein grei måte å kategorisere dei tidlegare «ekte» preposisjonane på, som t.d. *av, blant, for, i, med, på, til* og fleire. Problemet er berre at dei komplekse verbpartikkelkonstruksjonane

5 Nokre svenske kollegaer hevdar at ein konstruksjon som *Johan tvättade rent bilen* er gjengs i svensk, m.a.o. ville det vere nærliggjande å tolke *rent* som adverb her. Slike tilfelle har eg aldri sett eller høyrst i norsk.

6 Adjektivet *rein* kunne t.d. hatt følgjande grovrepresentasjon: **rein** adj A (attrib og som pred): *eit reint hus, huset er reint* B (som partikkel): *Johan vaska {rein(t)} huset {reint}*.

ligg nærare semantisk dei enkle motstykkane sine enn dei ligg dei lokative preposisjonane med nominal utfylling. Derfor tykkjer eg (18) er ei betre løysing. Der blir partiklane skilde ut i eigen bolk på grunnlag av den resultatative semantikken, og prinsippet \pm utfylling blir tona ned.

- (18) på prep A med nominal utfylling: *halde, setje hatten på hovudet* B som partikkel (enkle og komplekse partikkelkonstruksjonar): setje {på} hatten {på} (oppå hovudet).

Eit atterhald ein må ta med ei slik inndeling, er at ein vil finne mange resultatative konstruksjonar òg i tyding A. Det vil gjelde direksjonale variantar (jf. konstruksjonar med *setje* vs. lokative *halde*), og slik sett kunne ein halde fram med eit tredje alternativ, som plasserte desse i tyding B. Rett nok trur eg at skiljet i (18) er fornuf-tig. Det ville vere eit slags kompromiss med den tradisjonelle føringa, bortsett frå at ein ikkje skil ut tyding B med ein ny ordklasse. Kva så med dei typisk intransitive lokative preposisjonane i NRG, dei gamle stadadverba som *borte, framme, heime, inne, nede, oppe, ute* og fleire? Eg har ikkje noka sterk meining om desse bør klassifiserast som adverb eller preposisjonar, anna enn at det som nemnt er noko paradoksalt over ei preposisjonsnemning for ord som (så å seie) aldri står med utfylling. At desse har ein partikkelbruk, slik som dei «ekte» preposisjonane, endrar ikkje på saka.

- (19) a. Halde {inne} katten {inne}
b. Halde {nede} folket {nede}

Partiklar er som nemnt kategorifleksible og skulle slik sett vere uproblematisk å nytte under eit adverb-oppslag òg. Eit argument for likevel å kalle denne gruppa for preposisjonar kunne vere at ho då kan sjåast meir direkte i samanheng med det direksjonale motstykket sitt, nemleg *opp, ned, ut, inn, fram* og fleire. Typisk for

desse er at partikkelbruken av dei er overveldande, medan dei nokre gonger får nominal utfylling. Døme (3) lenger oppe syner at *opp* i NO har 418 liner som adverb og berre 8 som preposisjon, og dette reflekterer nokolunde bruksfrekvensen. Den klart mest vanlege bruken av *opp* er som partikkel til eit verb, og det kan ha å gjere med at *opp* har eit stort spekter av avleidde tydingar i tillegg til den konkrete, og at desse variantane koplar seg til ei rad ulike verb, både innhaldstunge og avbleikte. Men *opp* har altså òg ein meir marginal bruk med nominal utfylling. Dette gjer at ordet vanskeleg kan oversjåast som preposisjon. Skulle ein då gje *opp* ein sameint representasjon, må ein kalle lemmaet for preposisjon. Det gjev oss følgjande framstilling:

- (20) **opp** prep A som partikkel: *lukke opp døra, seie opp jobben* B med nominal utfylling: *laksevandringa opp fossen*

Sjølv om det ser mest elegant ut at eit lemma ikkje skiftar ordklasse undervegs, er det vanskeleg å unngå i nokre tilfelle. Eg nemnde *innan* og *til* ovanfor, og snart kjem NO til den sistnemnde preposisjonen. Då må ein fort ty til ein eigen konjunksjonsbolc (som ville ha vorte subjunksjon med NRG-systemet, jf. døma (5–6) ovanfor). Ei løysing for *til* kan vere at ein vel å skilje denne varianten ut med eigen homograf, utan at eg er spesielt begeistra for ei slik føring, då eg som nemnt trur det gjev galne assosiasjonar om homonymi.

Ein siste føremon med å innføre partikkelomgrepet i ordboka, er konsekvensen det ville ha for dei store verbartiklane. I NO følgjer dei store verbartiklane eit fast skjema (sjå Grønvik 2008), der verb + adverb-kollokasjonane kjem som eigne utheva underoppslag i tyding C i artikkelen. Dette er etter kvart ei fast gjennomført oppskrift i NO. Eit problem er likevel at føringa er noko sprikande

under dei ulike verba. Under tyding C til *leggja* i band 7 heiter det at verbet opptrer «med adv el trykksterk prep».⁷ *Liggja* har formuleringa «i faste samb med adv el prep». I band 9 har eg sjølv brukt denne formuleringa under *skjera*: «i ymse samb med trykksterk prep, adv el adj» – og med det inkludert samband som *skjera klar, rein og sund* i same rekkja. Om ein reknar både retningsorda (*opp, ned, ut, inn* ...), korte preposisjonsfrasar (*bak seg*) og dei nemnde adjektiva som partiklar, kan ein likeins gjere den overordna formuleringa i denne tydingsbolken endå meir generell: «i samb med partikkel».

7. Skilnad på termar i deskriptive og normative ordbøker?

Grammatiske termar må gjerne avgrensast normativt om dei skal nyttast med ei mest mogleg presis tyding. Eit potensielt paradoks for NO som ei deskriptiv ordbok er at det deskriptive kan setje grenser for kva grammatiske termar ein kan eller bør nytte seg av metaspråkleg. Eg tenkjer då på konfliktane som kan oppstå mellom den normative verdien ein har gjeve ein term, og det faktiske innhaldet termen har i litteraturen. La oss seie at ein redaksjon bestemmer seg for at *adverb* tyder X og *preposisjon* tyder Y i den aktuelle ordboka. X og Y viser seg så å vere i pakt med den ålmenne oppfatninga av adverb og preposisjonar. Dersom redaksjonen så innfører Z som eit nytt preposisjonsomgrep, kan ein ikkje lenger nytte termen *preposisjon* metaspråkleg og samstundes gå ut frå at brukarane forstår preposisjonstermen som Z, og ikkje som Y. Det er grunn til å vente ein viss tregleik i systemet, at det tek tid før Z etablerer seg hjå folk. Spørsmålet er då om det er ei *deskriptiv*

7 *Leggja bak seg* er då ein trykksterk preposisjonsvariant, dvs. ein PP med utfylling. Tilsvarande døme har ein i *ta {av seg} skoa {av seg}, sende {om bord} maten {om bord}*

ordbok, som jo skildrar etablert folkemål, si oppgåve å *innføre* nye metaomgrep – eller om ein primært skal ha ei deskriptiv tilnærming også til grammatiske termar.

Nokre metaspråklege termar (òg utanom ordklassetermane) endrar seg gjennom redigeringstida til ei stor ordbok. Eit døme har vi med *komplement*, som tradisjonelt har vorte nytta med tydinga ‘preposisjon fylgd av styring; preposisjonsuttrykk’ (jf. definisjonen i NO, band 6). I meir moderne norsk lingvistisk litteratur tyder komplement ‘tillegg, utfylling til kjerne i frase’ (NO, band 6). Problemet for ordboksredaktøren er at nokre brukarar vil forstå termen med den første tydinga, og nokre med den andre. Dermed blir ordet problematisk å nytte metaspråkleg i ordboka. Og det same kunne ein tenkje seg vil skje med preposisjonar, om dei får den nye tydinga Z i staden for den tradisjonelle Y. Vi kan hevde at føremålet med Z er å gje ein grammatisk meir sameint representasjon av preposisjonskategorien, men i realiteten kan vi ende opp med ein konflikt med den etablerte forståinga av ordet. Med dette prøver eg ikkje å seie at ei deskriptiv ordbok konsekvent skal styre unna terminologiske nyvinningar, men at ein skal vere var på at ein ny term under oppslaget kan gje nye metaspråklege utfordringar. I dette perspektivet har òg kjepphesten min i denne artikkelen om å bruke omgrepet *partikkel* på ein bestemt måte – som mellom anna bryt med den etablerte definisjonen av ordet i NO, band 8 – eit alvorleg problem. Kanskje ville det vere lettare å leggje beslag på *verbpartikkel*. Då får vi òg eit meir gjennomslutlig omgrep. Ut frå delane 4–6 ovanfor kan ein nemleg forstå ein (verb)partikkel som eit lite ord (oftast preposisjon) eller ein kort frase (PP) som står eller bind seg til eit *verb*.

8. Avrunding: ny terminologi aktuelt for NO?

Ovanfor har eg argumentert for at (verb)partikkelomgrepet kan vere eigna til bruk i ordbøker. Det er ei fleksibel nemning som

kryssar ordklassane, og som såleis kan gje eit lemma ei meir sam-eint framstilling, ved at ein slepp å føre opp ny ordklasse undervegs i artikkelen, eller at det unaudsynleg og gjerne misvisande blir splitta opp i to lemma. Likevel ser eg føre meg ei litt meir moderat omlegging enn den Faarlund (2010) forfektar. Tidleg i denne meldinga luftar han eit ålment paradoks, nemleg at ei ordbok må halde seg til det same systemet frå *a* til *å*, men at «det [burde] ha vore mogleg å gå over til eit meir moderne system ved oppstarten av det nye prosjektet Norsk Ordbok 2014» (Faarlund 2010:314). I praksis vil det seie at ein måtte bytt system om lag frå bokstaven *h* og fullført dei åtte siste banda (av tolv) med det nye systemet. Det trur eg NO gjorde lurt i å la vere (om det nokon gong vart vurdert). Dessverre er det slik at det er mykje med den trykte utgåva som i verste fall kan verke avleggs, allereie før ho er ferdigstilt i 2014; mellom anna kunne ein seie det om redigeringsmålet, som følgjer den nynorske 1938-norma. Eg trur i alle høve det er greitt å fullføre det trykte verket med eit mest mogleg konsekvent og rigid *formelt* system, sjølv om nokre endringar er uunngåelege. Filologane vil finne ut av problema sine likevel, medan andre brukarar truleg ville bli ytterlegare forvirra med eit system- (og for den saks skuld norm-)byte. Såleis vil eg paradoksalt nok fråråde NO-redaksjonen å innføre partikkelomgrepet *no*. Heller vil eg tilrå å føre adverba som ei eiga bokstavtyding, slik som i døme (2) ovanfor, og ikkje som ein eigen homograf, som i døme (1). Men dette tyder ikkje på nokon måte at diskusjonen her er mindre relevant. For det første finst det heldigvis fleire ordboksprosjekt enn NO, som då òg kunne vurdere å ta i bruk eit anna system enn det tradisjonelle. Og for det andre bør dette stadig vere aktuelt for *nettutgåva* til NO, som bør, så langt pengane rekk, vere fleksibel og framsynt nok til å kunne endre på formelle nemningar. Å endre redigeringsspråk for ei heil ordbok vil nok vere særskild ressurskrevjande, men å kople oppslagsorda til den gjeldande rettskrivinga bør vere overkommeleg. Ei eventuell omlegging av ordklasser og system må sjølv sagt gjerast

greie for og helst grunnjevast i ei brukarretteiing, sidan det ville bryte med tradisjonell leksikografi. Og likeins kunne ein tenkje seg, når plassomsyn ikkje må takast i same grad lenger, at eit oppslag som hadde fått ein ny ordklasse, fekk nemninga kopla til ein faktaboks eller liknande, som problematiserte endringa. I alle fall bør det i ei nettordbok kunne opnast for meir fleksible løysingar – òg for grammatiske nemningar – enn i ei papirordbok.

Litteratur

- Bye, Patrik, Trond Trosterud & Øystein Vangsnes 2003: *Språk og språkvitskap. Ei innføring i lingvistikk*. Oslo: Det Norske Samlaget.
- Faarlund, Jan Terje 2010: Melding av Norsk Ordbok, band VIII. I: *LexicoNordica* 17, 313–319.
- Grammatiske termer til bruk i skoleverket*. Tilråding fra Norsk språkråd og Utdanningsdirektoratet (2005). <http://skolenettet.no>
- Heggstad, Leiv 1931: *Norsk grammatikk. Større utgåve*. Oslo: Olaf Norlis Forlag.
- NO = *Norsk Ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet*. 1966–. Oslo: Det Norske Samlaget.
- NRG = Jan Terje Faarlund, Kjell Ivar Vannebo & Svein Lie 1997. *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.
- Sandøy, Helge 1976: *Laust samansette verb i vestnordisk: ein samanliknande leddstillingsanalyse for islandsk, færøysk og romsdalsmål*. Magisteravhandling ved Universitetet i Oslo.
- Sandøy, Helge 1991: *Norsk dialektkunnskap*, 2. utg. Oslo: Novus forlag.
- SAOB = *Ordbok över svenska språket utgiven av Svenska Akademien*. 1898–. Lund: Gleerups.

- Svenonius, Peter 1996: The Verb-Particle Alternation in the Scandinavian Languages, <http://ling.auf.net/lingBuzz>
- Svenonius, Peter 2008: Spatial P in English. I: G. Cinque & L. Rizzi (red.): *Mapping Spatial PPs*. Oxford: Oxford University Press, 127–160.
- Western, August 1921: *Norsk riksmåls-grammatikk*. Kristiania: H. Aschehoug & Co.
- Aa, Leiv Inge (under arbeid): *Dynamiske og resultative verbpartiklar*. Ph.d.-arbeid ved NTNU.
- Åfarli, Tor Anders 1985: Norwegian Verb Particle Constructions as Causative Constructions. I: *Nordic Journal of Linguistics* 8, 75–98.
- Aasen, Ivar 1848: *Det norske Folkesprogs Grammatik*. Christiania: Werner & Comp.
- Aasen, Ivar 1864: *Norsk Grammatik*. Christiania: Mallings.

Leiv Inge Aa
ph.d.-kandidat, redaktør i NO 2014
Institutt for nordistikk og litteraturvitenskap
NTNU
NO-7491 Trondheim
leiv.inge.aa@ntnu.no

ANMELDELSER

Den svenske forbindelse, version 2010

Lars Brink & Ken Farø

Kjeld Kristensen (hovedredaktør), Else Bojsen og Pernille Folkmann (redaktører): *Svensk-dansk ordbog*. København: Det Danske Sprog- og Litteraturselskab og JP/Politikens Forlagshus A/S 2010. 1183 sider. Knap 50.000 opslagsord. 1. udgave. 299 DKK.

1. Indledning

De fleste danskere kan i stor udstrækning forstå både talt og skrevet svensk. Det skyldes det nære slægtskab mellem sprogene på bl.a. leksikalsk plan: fx *tjuvstart* – *tyvstart* og *fläskstek* – *fläsksteg*. Men når det er konstateret, er der også store dele af ordforrådet, som slet ikke ligner hinanden. Tag bare disse ordpar: *tomteblöss* – *romerlys*, *sopskyffel* – *fejebakke* og *tvestjärt* – *ørentvist*. Eller hvad med svenske udtryk som *en hel special*, der kan bestilles ved en pølsevogn? Resultatet er to pølser med brød og kartoffelmos. Svensk kan derfor undertiden være helt eksotisk for danskere.

Man kan også konstatere orddannelseskontraster mellem de to sprog, som når de svenske enkeltordskonstruktioner *tårta*, *okeja*, *bolla*, *showa* og *köa* modsvares af danske flerordskonstruktioner: *kaste en lagkage i hovedet på*, *sige okay til*, *spille bold*, *optræde i et show* og *stå i kø*. Om det er en intersproglig regelmæssighed, en tendens eller blot en tilfældighed, som kun vedrører de nævnte eksempler, kan og skal vi ikke afgøre her.

Ovenstående eksempler stammer fra den genstand, vi her vil beskæftige os ret indgående med. Vi belyser et nyt opslagsværk om svensk for danskere fra en række meget forskellige vinkler.

1.1. En helt ny svenskordbog i Danmark

Svensk-Dansk Ordbog kom i 2010. Den var ventet med længsel. Dens eneste kompetente forgænger var fra 1954, og siden er der udkommet nogle få, frustrerende små idiotika. At slå forgæves op i en ordbog giver en tom, afmægtig fornemmelse. Den fik man sjældent i Valfrid Munch-Petersen og Ellen Hartmanns ordbog (som rummer en hel del flere ord end den nye), men efter godt 50 år er det klart, at den ofte meldte pas. Dertil kommer, at den gamle trods sine store fortjenester ikke helt levede op til sin tids bedste leksikografi. I 1957 kom Molde og Ferlovs udmærkede dansk-svenske ordbog, men den har man jo ikke meget brug for som dansker; hvordan et dansk ord udtales, mener alle at vide, og hvad det hedder på svensk, kan være lige meget, eftersom danskere sjældent giver sig af med at tale svensk.

Hvis et ord ikke stod i den gamle svensk-danske, var der en stribe svensk-svenske til rådighed, fra SAOB og nedefter, både kvantitativt og kvalitativt forstået. Men den i svensk ubefæstede kan naturligvis ikke bruge disse, og den avancerede risikerer let at møde et ord eller en nuance i forklaringen, han ikke forstår, eller ligefrem en lumsk lighed (se afsnit 4). Tryk er man først med en stor, up-to-date svensk-dansk. Og derfor skal den nye hilses med trumpetstötär/trompetskrald. Ikke mindst til den pris.

Ordbogen er beregnet for danskere, ikke svenskere. Det siges klart og tydeligt i forordet, og den er lavet til hjælp for tekstreception. Af samme grund kan man undre sig over, at der er lagt vægt på at vise varians på ækvivalentsiden. For det er der, uanset hvad indledningen postulerer (se afsnit 5).

1.2. Om denne anmeldelse

Vi tilstræber videnskabelighed i vores metode. Det gør vi ved at forsøge at sikre a) omfang, b) spredning og c) tilfældighed i vores materiale i form af bl.a. to relativt omfattende stikprøveundersøgelser

af varierende type. Vi har dermed sørget for et minimum af repræsentativitet. Men der er også store dele af anmeldelsen, der bygger på tilfældig bladren i værket. Vi er ikke gået efter noget specifikt.

Fokus er på følgende elementer: layout (afsnit 2), dækningsgrad (afsnit 3), falske venner (afsnit 4), lemmaliste og ækvivalenter (afsnit 5), leksikalske huller og stilmarkører (afsnit 6) og endelig udtale (afsnit 7).

2. Layout

Bogen er indbydende. Bortset lige fra vægten på 2,1 kg, som man dog på den anden side ikke ønsker fordelt på to bind. Indbinding, guldsnit og papirkvalitet er i top. Og de 3 spalter står typografisk ualmindelig klart på siderne. Ideen med at blege klummetitlerne virker god, hvorimod det generer, at begge sidens klummetitler står samlet yderst på siderne. I alt er der 3 slags midterprikker til adskillelse plus firkant plus komma og semikolon. Til gengæld er (adskillel- ses)punktum, noget forbavsende, pensioneret. Systemet er næppe gennemskueligt for brugerne (selvom det forklares), men det giver en god og diskret opdeling af artiklerne. Bokssystemet, som alene indeholder “falske venner”, har vi viet et selvstændigt afsnit (4).

körsång ['kø:ɾ:] *sb.* -en, -er, -erna

• korsang

körtel ['kø:ɾ:əl] *sb.* -n, körtlar, körtlarna

1. kirtel

2. nyre (malmklump af uregelmæssig form)

körtelfeber *sb.* -n

• mononukleose, kirtelfeber

körvel ['kø:ɾ:vøl] *sb.* -n

• kørvel (plante; krydderi)

dansk körvel kørvel, havekørvel (*Anthriscus cerefolium*) **SYN** trädgårds-körvel

köttulle *sb.* -n, ..bullar, ..bullarna

• (især i plur.) frikadelle (lille) □ köttbullar med lingon och potatismos frikadeller med tyttebær(syltetøj) og kartoffelmos - i Sverige förknippas köttbullar ofta med nostalgiska minnen vilket kommer till uttryck i smeknamnet "mammans köttbullar" i Sverige forbindes frikadeller ofte med nostalgiske minder, hvilket kommer til udtryk i kælenavnet "mors frikadeller"

Da kødbolle hedder ikke köttulle på sv., men frikadell.

Figur 1: Uddrag af side 554 i Svensk-Dansk Ordbog.

3. Dækningsgradsundersøgelse

Vi har undersøgt ordbogens leksikalske dækning ved hjælp af følgende metode: Et uddrag af 3 forskellige værker er blevet excerperet. Det drejer sig om (1) Sjöwall/Wahlöö: *Polis, Polis, Potatismos* (1970), (2) *Kalle Ankas Pocket 326, Storm* (2006) og (3) Alexander Solsjenitsyn: *Den första kretsen* (1969/1974).

Udvalget er et forsøg på at differentiere tre forskellige stilniveauer, der i hvert fald tendentielt må give sig udslag i 3 forskellige delmængder af ordforrådet. Og det er et delmål med undersøgelsen at se på, om der er forskel på ordbogens evne til at dække forskellige typer (i dette tilfælde alene fiktionale) tekster. At de to af kilderne er oversat, fra henholdsvis russisk og engelsk, kunne godt være et problem. På den anden side udgør oversatte tekster en væsentlig del af den store tekstmasse, vi hver dag konfronteres med. Og skulle det være et reelt problem, måtte det i givet fald være på den måde, at oversættelserne indeholdt mange ord, som var kunstige, sjældne eller på anden vis markerede. Vi må se, om den antagelse holder stik. Det samme gælder teksternes alder. To af dem er fra 1970'erne. Det er dermed naturligvis i skrivende stund ikke helt nutidige tekster. På den anden side er de heller ikke decideret gamle, og vi kan ikke se noget galt i at teste ordbogen ved hjælp af kilder, der har klassikerstatus. Man må så bare være lidt smidig i det enkelte tilfælde, hvis der skulle vise sig nogle relevante problemer.

Vi har konfronteret de første 50 ord og ordforbindelser fra hver af de tre kilder, som har givet anledning til opslag, med ordbogen. I nogle tilfælde fordi excerptisten ikke kendte ordet eller udtrykkets betydning, i andre fordi der er en markant afvigelse fra dansk, som gør det sandsynligt, at andre brugere af ordbogen kunne finde på at slå udtrykket op. Visse ord er på forhånd sorteret fra, fordi de er så snævert knyttet til den specifikke tekst eller serie, at deres relevans i hvert fald i en trykt ordbog må vige for andre hensyn.

Et eksempel er *sockernäbb*, Anders And's/Kalle Anka's ømme tiltale til Andersine/Kajsa (dansk: *pus*). Dette skal ikke forstås sådan, at fiktionismer (jf. Farø 2005) generelt er irrelevante i en trykt to-sprogsordbog. I værket finder man da også både enkle og flerledede fiktive navne som *Långben*, *Musse Pigg*, *Mårten Gås*, *Kalle Anka*, *Kajsa Anka* samt *Nalle Puh* og *Mumintrollen*. Men man kan trods alt ikke slå fx de lige så veletablerede *Björnligan*, *Ankeborg* og *Stål-Kalle* op.

I tabellerne findes følgende oplysninger: sidetal fra kilden (kun angivet ved sideskift), opslagsordet eller -udtrykket, en eller flere relevante ækvivalenter eller et "0", hvis det ikke er beskrevet. Et "NB" refererer til en kort, løsrevet kommentar under tabellen. For overblikkets skyld har manglende opslag fået mørkegrå baggrund, mens sådanne med en kommentar er lysegrå. En kommentar er ikke nødvendigvis påpegnings af fejl, men ofte mere et forslag til overvejelse. Efter tabellerne sammenfatter og diskuterer vi resultaterne af undersøgelsen.

s. 5: folktom = menne- sketom, folketom	göra sig påmind = begynde at kunne mærkes	luffare = (ung) person der rejser rundt uden bestemt rejserute el. mål; ...	inte så värst ... = ikke særlig ...	snövit = snehvid
barmästare = barchef	tystlåten = tavs, fämält	s. 6: servitör = tjener	fundersam = tankefuld	solenn = højtidelig, festlig, pompos
belamra = fylde, over- fylde	dra sig un- dan = trække sig tilbage	ty = for	reslig = høj og statelig	avlägsna = fjerne
fnittra = fnise (nervøst)	affärsman = forretnings- mand	långrock = overfrakke	svängstol 0 NB	pelare = pille, søjle

kavaj = jakke, jakkesæt	s. 7: annalkende = som nærmer sig	föremål = genstand	räfflad = riflet NB	framstupa = forover
krenelera 0	tvært = skarpt, brat	aj! = av!	s. 8: ungkarlsly = ungkarlehybel	veckoskifte 0 NB
rutig = ternet	Borta med vinden 0	kalsonger = underbukser	mitt emot = lige over for NB	skam til sägandes = beklageligvis; jeg beklager, men desværre... NB
s. 9: polishus 0 NB	var god drøj = vent venligst; et øjeblik	slarvig = sjusket	skockas = stimle sammen; samle sig	sorgset = bedrøvet
stimmig = larmende	fullständiga = fuldstændiggøre; supplerer NB	ledig = afslappet	utanpåskjorta = skjorte der bæres uden på bukser el. nederdel	ett slag = et øjeblik
överblicka = overskue NB	s. 10: dylik = sådan, lignende	klyftig = intelligent	s. 11: häpet = forbløffet	skymt = skjult

Tabel 1: Oversigt over medtagne og “manglende” ord fra *Polis, Polis, Potatismos*.

svängstol: ordet er ikke så udbredt.

räfflad: ikke som adj., men som vb.

veckoskifte: ordet er ikke så udbredt.

mitt emot: strengt taget en lumsk lighed, redaktionen ikke er opmærksom på; det betyder ‘over for’, ikke ‘lige over for’, jf. SAOB.
skam til sägandes: ækvivalenterne passer ikke med den aktuelle kontekst: “Han rörde om i groggen med en gaffel som han

skam till sägandes hade stulit eller råkat få med sig från en restaurang”. Desuden kan *med skam at melde* overvejes.

polishus: ordet er særdeles almindeligt med millioner af hits på internettet. Svarer vel til dansk *politigård* og *politistation*.

fullständiga: her snarere *gøre komplet*.

överblicka: her snarere *forsøge at få et overblik*.

Resultat: en dækningsgrad på 90 %. Vi diskuterer resultaterne samlet efter sidste tabel med kommentarer.

s. 2: Musse Pigg = Mickey Mouse	Stål-Kalle 0	Oppfinnar-Jocke 0 NB	Mårten Gås = Fætter Guf	pocket = pocketbog NB
s. 5: vilse = vild (på forkert vej)	gummi-gluttar 0	Ankeborg 0	s. 6: knycka = hugge, stjæle	Björnligan 0
mogen = moden	nalla = hugge, negle	s. 7: trams = vrøvl, pjat	förvärva = erhverve	klottra = kradse ned
slitgöra = slidsomt arbejde	tät = rig	muta = bestikke	snorkig = hoven, storsnudet	rappa på = rubbe sig
kosing = penge	s. 8: flämt! = gisp!	s. 9: soppa = suppedas	sjöplan 0	envisas = insistere
mallig = storsnudet	gnälla = jamre	s. 11: få kalla kårar = det løber ngn koldt ned ad ryggen	tjosan! = hej!	Kajsa = Andersine
trasa = klud, las, trævl NB	kila = smutte	s. 12: rulla tummarna = trille tommelfingre	urbota = håbløst, umådelig	lurt = mistænkeligt

s. 13: oljud = larm	inte få en blund i ögat = ikke lukke et øje NB	s. 14: gå skallgång = gå frem på linje ved en eftersøgning e.l.; drive klapjagt NB	slå vad = vædde	klubba ner = slå ned
mysa = hygge sig	snoka = snage, snuse	deckare = kriminalroman	s. 16: sjappa = stikke af	Mimmi = Minnie Mouse
spaning = eftersøgning	s. 18: Långben = Fedtmule	tvi = fy, føj	s. 19: fasansfull = frygtelig	s. 24: vad i hela friden = hvad i alverden?

Tabel 2: Oversigt over medtagne og “manglende” ord fra *Kalle Ankas Pocket 326: Storm*.

Oppfinnar-Jocke: heller ikke appellativet findes i ordbogen. Det skyldes, at formen er uformel, normalt *upppfinnare*. Men torsoen *opp-* er faktisk medtaget.

pocket: oversættes med *pocketbog*, men *Kalle Ankas Pocket* hedder *Jumbobog* på dansk.

trasa: *klud* er nok den mest adækvate ækvivalent, men det kræver en betydningsdifferentiering.

inte få en blund i ögat: strider ifølge talrige kilder mod det trade-rede *ögonen* (1:50 ifølge *Google*), så ordbogens pluralisform er helt rimelig.

gå skallgång: evt. også *danne kæde*.

Resultatet er en dækningsgrad på 88 %, altså kun en ubetydelig afvigelse fra før. Det er stadigvæk virkelig tilfredsstillende.

s. 5: jäktade = stressede	överstelöjtnant = oberstlöjtnant	smärt = slank	rya (med) = råbe NB	s. 6: söla = smøle, drysse
bördor = byrder	s. 7: lågsko = sko NB	blötsnö = blød, våd sne; tøsne	farkost = fartøj	trottoar = fortov
knacka = banke	irra omkring = strejfe om NB	s. 8: köa = stå i kø	imma = dug	lur = rør
vara litet på örat = have en kæp i øret	slå ett nummer = dreje et nummer NB	loj = dorsk; sløv	däst = stopmæt; forædt; fed	slött (=) NB
s. 9: (koka av) ilska = raseri NB	gröngöling = grønskolling	yllekostym = ulddragt	skåda = skue, se	hålla hus = opholde sig
snåla = spare, være nærig	s. 12: hedersord = æresord	illa = meget	kulmen = højdepunkt	s. 13: pall = afsats, skammel
brits = briks	åtsmitande = stram	bitterljuv = bittersød	s. 14: öde = skæbne, livshistorie	fläta = flette
neka = benægte	hunnit = nået	nuna = fjæs	sträv = ru, grov	möda = besvær, slid
hydda = hytte	jämka = rette	s. 15: förebråelse = bebrejdelse	s. 16: vispgrädde = flødeskum	krånglig = besværlig
då och då = af og til	s. 17: buga = bukke	anlända = ankomme	s. 18: puta = bule (ud); strutte	ihålig = hul (adj.)

Tabel 3: Oversigt over medtagne og “manglende” ord fra *Första kretsen*.

rya med: (“de biträdande ryade med byråcheferna”) den konkrete form med præposition mangler, men ækvivalenten vil ofte være tilstrækkelig til at regne ud, at det er *råbe* og *skrigende adkommandere med*, der er aktuelle.

lågsko: opslagets supplerende oplysning (*modsat støvler*) kan måske virke mystificerende uden yderligere forklaring.

irra omkring: *rende rundt* passer nok bedre her (jf. *irra runt*).

slå ett nummer: her kunne have været medtaget den mere nutidige *taste*.

slött: kan man ikke slå op. Det kræver indsigt i svensk grammatik at vide, at man skal fjerne *-tt* og slå op under *slö*, som bl.a. betyder ‘sløv’.

koka av ilska: står ikke som udtryk, men separat. Når først *ilskas* betydning er forstået, giver resten sig selv, for dansk har samme metafor.

Resultat: 100 % dækning.

3.1. Konklusioner på dækningsgradsundersøgelsen

Kvantitativt ser det virkelig flot ud: 93 % succesopslag i gennemsnit er *langt* bedre, end vi havde forventet. Ser man på teksttyperne, der dækkes, så danner der sig et billede af, at ordbogen dækker høj stil lidt bedre end mellem-lav: Jumbobogen scorer lavest, og den højletterære roman højest; det kunne måske tyde på, at ordbogen bygger på lidt ældre ordbogskilder. Interessant er også, ikke mindst af metodiske grunde, at den ikke-oversatte kilde faktisk scorer dårligere end den, som er oversat fra russisk. Det kunne måske tyde på, at der ikke er noget afgørende problem forbundet med at bruge oversatte tekster (af en vis kvalitet) til undersøgelser af ordbogers dækningsgrad. Det er klart, at tallene kun giver en vis indikation om ordbogens evne til at give brugeren succesopslag. Men de tre meget forskellige kilder peger i samme retning.

Ordbogen “mangler” selvfølgelig nogle ord og ordforbindelser eller betydninger, herunder fx *sjöplan*. Det kan der være forskellige grunde til. Undersøger man ordbogen lidt nærmere, finder man faktisk *sjöflygplan*, som *sjöplan* må være en kortere variant til. Og

prioriteringen, hvis der ligger en sådan bag, ser ud til at være fornuftig, for den treleddede sammensætning har 2-3 gange højere forekomst på internettet. På den anden side er den korte form jo samtidig den mindst gennemsigtige af de to. Det peger på, at den nuværende løsning kunne suppleres med et henvisningsopslag.

Man kunne her – almenleksikografisk – spørge, om en god to-sprogsordbog skal levere den bedste ækvivalent i enhver kontekst, hvad vi tilsyneladende har forudsat ovenfor. Svaret må være, at den skal levere en ækvivalent for samtlige etablerede betydningsvarianter – sådan at læseren *forstår* alle disse. Men sprog er så kompliceret, at man ikke kan forlange, at der aldrig er et bedre ord som ækvivalent. En ordbog kan ikke tage hensyn til, at fx *småpiger* visse steder lyder bedre end *små piger* og omvendt.

Herudover må man sige, at fraværet af flere af de “manglende” ord og udtryk er velbegrundede, idet de tilsyneladende ikke er voldsomt udbredte i svensk.

3.2. Unødvendige opslag?

Når man i en anmeldelse af en trykt ordbog påpeger, at visse ord ikke kan slås op, er man næsten forpligtet til at anvise mulige kandidater til stryging. Og det er faktisk ingen vanskelig opgave her, selv om man ofte kan producere et modargument for alligevel at medtage et ord. Her er et par stykker: *landskamp*, *kaffer*, *landstorm*. Ordene er særdeles transparente for en dansktalende, og derudover er et par af dem endda historiske. Andre vel nok knap så oplagte opslag er fx *gerontologi* og *it*.

En klassisk pladsfylder er nationalitetsgloserne, der tilsyneladende ofte optages kritikløst i stedet for at vurdere dem enkeltvis ud fra opslagsbehov. Det er de vist også blevet her. Rariteter som *tuvulansk*, *botswansk* og *seychellisk* skal ydermere ganges med tre, fordi de er i selskab med 2 yderligere og lige så sjældne orddannelser på opslagsplads.

3.3. Sproglige eller encyklopædiske oplysninger?

Opslaget “*kajskjul*: havneskur” kunne også se ud, som om vi her befandt os i den mere marginale del af ordforrådet. Både lemmaet og ækvivalenten virker umiddelbart meget specielle. Vi blev derfor overraskede, da vi fandt mere end 6.000 hits på internettet for ækvivalenten, mens det svenske udgangspunkt optrådte omkring 30.000 gange. Helt sært eller sjældent er det altså ikke. Men konteksterne viser tydeligt, hvad problemet er:

- (a) Kajskjul 8 – Kända artister och dans på Packhuskajen
- (b) Kajskjul 107 Här finns bl a radiostationen Bandit
104 8 Göteborg
- (c) Kajskjul 207. Room for up to 400 guests

Der er åbenbart tale om en omfunktionalisering af pakhusområdet til en art forlystelses- og kulturområde, hvad der havde været relevant at forklare encyklopædisk. Det leksikalske indhold er her sekundært. For der er ikke tale om en ordinær ny betydningsvariant i svensk. *Kajskjul* har sin gamle betydning. Det er *referentens brug*, der er ændret. Jævnfør, at *brosten* ikke skifter betydning, selvom referenten bruges mod politiet. Noget andet er, at *havneskur* nok ikke dækker helt, hvis vi tager størrelsen i betragtning. Et havneskur er meget mindre end et *kajskjul*, som man kan forvisse sig om ved en billedresearch på nettet.

Vi har et lignende eksempel i opslaget *galärvarv* = værft for galejer. Igen er det en slags institutionalisering af ordet, for *Galärvarvet* har fået en helt anden funktion end den oprindelige. Det er et museums- og forlystelsesområde i Stockholm; men det fortæller artiklen intet om. Man kan sige, at ordbogen på dette punkt er lidt gammeldags, når den insisterer på kun at beskrive den rent sproglige betydning af det specifikke ord. Måske ud fra en forestilling om, at det encyklopædiske hører leksika til. Men ordbogen

indeholder masser af encyklopædisk stof i øvrigt, og desuden har bl.a. Århuskolen (jf. også Farø & Hyllested 2007) peget på, at en sådan opdeling i sprog og encyklopædi er unødvendig, ja ligefrem uhensigtsmæssig. For selvfølgelig er de encyklopædiske betegnelser *også* ord. Man skulle naturligvis have bragt propriet og overladt appellativet til leksika og evt. en svensk-latinsk ordbog.

4. Lumske ligheder (falske venner)

De lumske ligheder er vores smertensbarn. Ikke, at vi besværer os over, hvad der står i de mange bokse. Men først og fremmest, at de er placeret der. Det lumske er jo, at man ikke søger oplysning på alfabetisk plads. Man nødes til at tro, at fx svensk *korn* betyder ‘korn’ og slår det derfor ikke op. Hvis man skulle gardere sig mod både lumske ligheder og ukendte ord, måtte man slå alt i en tekst op. Den rette løsning er at gøre diskret opmærksom på lumskeheden *i* selve artiklen på alfabetisk plads, altså fx en sigel, der vækker den sløve læser: Du læste rigtigt: ‘byg’ og dermed ikke ‘korn’ – og så samle de 100 lumske i en særlig liste at lære udenad.

Ordbogen rummer mindst 1.500 lumske ligheder. Men ca. halvdelen er slet ikke lumske for danskere, kun for svenskere. Hvis to falske venner har 1 fælles betydning på det ene sprog, men 2 betydninger på det andet, så er dette kun lumsk for ham med 1 betydning; han risikerer at møde den falske ven i en ham ukendt betydning. Ham med 2 betydninger får altid serveret sin falske ven i en betydning, han kender. Svensk *karl* er dermed ikke lumsk for dansktalende, da vi har den svenske betydning, men dansk *karl* er lumsk for svenskere. Der er hundreder af den type.

Der næst må vi indvende, at ‘lighed’ tolkes alt for bredt. Svensk *gråtig* ‘pylret’ hævdes at være falsk ven med dansk *grådig*, men her er ligheden ikke slående som ved de “ægte” falske venner; man vil selvfølgelig få mistanke om ‘grådig’, men man er ikke prisgivet.

Lidt anderledes med svensk *hed* ‘hede(strækning)’, der minder slående om dansk ‘brændende varme’, men samtidig minder lige så meget om *hede* i den betydning. Et ord, der minder om et ord i et andet sprog, vil selvfølgelig minde lige så meget om dettes eventuelle homonym. Ikke lumsk.

For sig er også typen *bane*. Svensk *bane* ‘drabsmand’ hævdes at have en falsk ven i dansk *bane* ‘jernbane osv.’ Dette forudsætter, at danskeren mangler det bogstav-identiske *bane(mand)*. Men her ved sprænges det rimeligt klare begreb “lumske ligheder”, og disse bliver legio.

Toppen af overflødighedshornet er svensk *koral* og *korall*. De hævdes begge at være lumske, skønt dansk har yderst vellignende modsvarigheder til begge. Samtidig er betydningen så forskellig som den vel kan være for to substantiver. Dette sidste, at der i praksis ikke er forvekslingsmulighed, spiller for lille en rolle: Svensk *ocker* ‘åger’ kaldes lumsk, fordi det kan lede til det mildest talt betydningsfjerne ‘okker’.

5. Ækvivalenter: 6 skævheder

De danske ækvivalenter er ordbogens hovedområde. Vi har derfor undersøgt 55 tilfældige opslagsord med henblik på dem, og det må straks siges, at vi ingen deciderede fejl fandt, kun visse skævheder.

Brugen af semikolon og komma er ikke god. Forordet siger side 15, at tilnærmelsesvis synonymymer adskilles af komma, lidt større betydningsafstande af semikolon. Imidlertid er der alt for mange semikoloner. Der er også for mange ækvivalenter, skønt forordet bedyrer, at der er sparet på dem. Et godt eksempel er *løjlig*: “latterlig; komisk; naragtig; fjollet”. *Latterlig* er rigeligt – og stilistisk en smule bedre end de andre (der snarere passer til *komisk*, *narraktig*, *fånig*), og semikolon mellem dem er da helt skævt.

Husera bet. 2: “holde til; have til huse”. Det er rigtigt, at de to

danske udtryk ikke er helt synonyme: En bande *holder til* på et gadehjørne, men har ikke *til huse* dér. Men det svenske ord har én og samme betydning, den er overbegreb for de to danske, de *kompletterer* den svenske betydning. Og det er naturligvis det svenske ord, der bør være afgørende for eventuelle betydningsopdelinger. Tilsvarende i *knipsa*: “klippe; nippe;...”. SAOB regner ‘knipsa’ for én betydning: “...med fingrarna l. med en tång l. sax...”, og det har denne eminenteste ordbog ret i. På dansk signalerer verbet *klippe*, at der bruges et instrument, men intet om, at det sker i et snuptag. På dansk kan *nippe* godt bruges om både fingre og instrument, så dét bør stå alene. Men i hvert fald er der ikke to *svenske* betydningsnuancer.

Varsna: “opdage; få øje på; blive opmærksom på; mærke”. Alt for mange ækvivalenter, og *få øje på* er direkte forkert, da *varsna* intet siger om sansemodaliteten.

Kniptång: “slags knibtang med flade kæber”. Bedre: niptang, jf. SAOB.

Varsla: “1. varsle... 2. bebude; være et forvarsel om”. Her brydes forordets regel s. 15 om ikke at udtømme de svenske betydningsvarianter, når dansk har helt tilsvarende. Selvfølgelig er der forskel på dét at varsle bevidst og dét at være et forvarsel. Det er den berømte forskel mellem signaler og symptomer. Men dansk *varsle* dækker fint begge dele: *Stewardesserne varsler strejke / stewardessernes stress varsler samarbejdsproblemer*.

Under det rent formelle opslagsord *langen* finder vi udtrykket *langen går!*, der gengives med det besynderlige “så langer vi!; lad gå videre!”. Her kunne man med fordel have forklaret, hvad udtrykket bruges til, en oplysning, man kun får meget indirekte gennem eksemplet: “hon ropade ”langen går” och alla bildade kedja för att snabbt få ner utrustningen i grottan”. Her er nok lidt for meget fokus på funktionen (tekst)oversættelse i snæver forstand frem for på at fortælle brugeren noget om svensk sprog, der adskiller sig fra dansk, hvormed brugeren kunne blive klogere på fremmed-

sproget. I øvrigt mangler ordet en markør om dets (formentlig snarlige) uddøen.

Via *tunn* finder man flerordsforbindelsen *lova runt och hålla tunt* "love mere end man kan holde; loven er ærlig men holden besværlig". Ideen med ækvivalent nr. 2 er vel at give en oversættelse, der mere har tilbøjeligheds-betydning end kasuistisk gyldighed. Men valget er ikke heldigt. For et ordsprog er en afsluttet sætning med konjugation, pronominaludfyldning etc. og kan aldrig være ækvivalent til et idiom, der ikke er sætningsformet. På samme måde som *jeg spiser* er en ganske anden sproglig fremtrædelsesform end *spise*.

6. Varia

6.1. Gammeldags/arkaiserende

På vores vej gennem de 55 opslag blev det klart, at ordbogen ikke skelner mellem *arkaiserende* og *gammeldags*. Det gør både SAOB og ODS, og det har det ene medlem af anmelderkollektivet argumenteret for i Brink (2008a). De to kategorier har faktisk intet med hinanden at gøre: Den, der arkaiserer, efterligner bevidst forgangne tiders sprog, oftest på fuldt nutidig etableret vis (*en vælig gang* etc.). Den, der er gammeldags, har blot et sprogræk, der er ved at uddø (fx *probenreuter* 'fusentast, starut'). I modersmålsordbøger har det arkaiserende mest lingvistisk interesse, sprogbrugere er jo med hensyn til dette ikke i tvivl om deres eget sprog. Men her er det fatalt. En dansker kan ikke ane, om et *gl.* betyder det ene eller det andet. Det er fx umuligt at vide, om ældre svenskere siger *tjäll* 'hytte' stilneutralt, eller om alle arkaiserer.

6.2. Huller i ordforrådet

Ordbogen er ikke for alvor opmærksom på de højinteressante tilfælde af huller i det danske ordforråd, dvs. at dansk undertiden helt mangler en etableret ækvivalent. Samme anmelder (Brink 2008b) har viet sådanne huller en afhandling, men det mærkes ikke her. – Svensk *kusin* angives “fætter; kusine”. Med sit semikolon angiver redaktionen, at svensk *kusin* énten betyder det ene, eller det andet. Det er forkert. *Kusin* er *overbegreb* for ‘fætter’ og ‘kusine’, hvad dansk ikke *kan* udtrykke med et etableret ord. Ingen finder på at hævde, at *människa* stundom betyder ‘kvinde’. Hvis det fremgår af en tekst, at *människa* konkret møntes på en kvinde, så er det med garanti ikke overbegrebet *människa*, der signalerer det. Det er sandt, at en god oversættelse oftest vil udnytte en sådan oplysning i teksten, men det kan langt fra altid lade sig gøre: *Han skriver till sin kusin.* – *Hur många kusiner har hon?* – *Det var inte nån kusin.* Under *by* står der, at dansk *by* ikke hedder *by*, men *stad*. Galt. Dansk *by* er *overbegreb* for svensk ‘by’ og ‘stad’, og svensk har et *hul* her.

6.3. Den pålydende betydning

Vi har også slået til lyd (Brink 2010) for, at ordbøger viser ords synkroner pålydende-betydning. De store nationale ordbøger gør det i og med, at de analyserer ordene i deres morfemer, men igen: Det har langt større praktisk betydning i fremmedsprogsordbøger. At forstå et ord helt er også at forstå dets pålydende-betydning. En dansker slår måske *maskros* op, men får ikke fat i “orme-rose” (ligesom en svensker i sin dansk-svenske ikke får fat i “løve-tand”, men tror, det er noget med “löv”). Og de færreste danskere vil kunne gennemskue *rallarsving*, selvom ordbogen korrekt giver den faktiske betydning. Vi plæderer selvfølgelig hverken for, at de selvindlysende ord analyseres, eller at ord med rent historisk pålydende-betydning annoteres, såsom *bröllop* † “brude-løb”.

Til vores glæde ser vi i øvrigt, at ordbogen under *løjrom* giver pålydende-betydningen på smukkeste vis.

7. Udtale

Udtale kan ikke være en hovedsag i dette værk for danskere, der – i nabosprogssituationen – skal lære at *forstå*, ikke selv *tale*. Alligevel har mindst 5.000 opslag fået transskription, nemlig alle simplicia, der ikke skønnes helt selvforklarende. Der er altså anvendt stor møjse med lydskriften, hvorfor den også her skal tages alvorligt.

Lydskriften hævdes s. 11 at være IPA med tillem্পninger til svensk. Men reelt er det IPA's *inventar* (undtagen g, der i IPA er ɣ) med så mange tillem্পninger, at den ikke bør bære IPA's velrenommerede navn.

Vi nævner ikke dette blot for at bevare begrebet IPA rent og purt, men fordi disse tillem্পninger er en pædagogisk fejltagelse: Lydskriften bliver let at læse for svenskere, men umulig at udlæse korrekt for danskere, der vil give de uskyldige bogstavtegn danske værdier.

Tegnet <i> har således sine normale svenske værdier, idet svensk <i> modsvares to vidt forskellige lyde kort og lang; den korte er let åbnet, den lange er meget snæver og let diftongeret el. med ð-klang (Viby-i). Lydtavlen antyder intet om denne svenske afvigelse fra dansk.

Alle langvokaler undtagen [ɑ:] er diftongerede i svensk: [ˈb̥l̥oːɹ̥] *blå*, [ˈst̥eːɹ̥n̥] *sten* etc., men de tillem্পes til monoftonger – som er lette at læse for svenskere, men unddrager danskeren den særsvenske udtale. I særdeleshed ville korrekt IPA-gengivelse af langt <u> (*mulet*), nemlig [yβ] (hvor β dog er støjfri), have vist danskeren denne svære lyd så pædagogisk som muligt; den har et eksotisk tegn for en eksotisk lyd, men ordbogens [ʉ:], der er lige så eksotisk, ville man slippe for.

De svenske apikaler (undtagen r) er dentale, ikke som i dansk alveolære. De noteres i præcis IPA [d̥ t̥ n̥ l̥ s̥]. Det lille diakritikon distraherer naturligvis, men det skal det også; det viser, at der er noget særligt ved disse svenske lyde, som næsten ingen danskere udtaler rigtigt. Det er ingen subtilitet. Da den ene anmelder underviste i dansk (på dansk) og fx sagde *vane*, indfortolkede svenskerne et r og ledtes da til fx “värna”, fordi rn i svensk udtales retroflekst eller, let reduceret, som dansk alveolært n. De danske alveolarer vildleder svenskerne til r-tolkning.

Men endnu bedre end præcis IPA ville Dania have været. Dania er som bekendt skræddersyet til dansk, men det er et komplet system, der kan gengive alle sprog korrekt. De danske lyde har simple tegn, men de fremmede har eksotiske (fordi de simple er udtømt). Et eksotisk tegn viser dermed, at der er en særlig udansk lyd på spil.

Dermed kunne man have ledt brugeren på sporet af, at b d g i svensk er udanske, i og med at de er stemte.

Danias *α* ville vise, at det svenske korte a er specielt og ikke skifter som det danske.

Danias *å* i fx svensk *komma* ville bedre end *ɔ* vise, at det er en å-agtig vokal.

Svensk æ-lyd (især kort), fx *sända, läsa*, er klart åbnere end dansk og får let danskeren til at lede efter ord med dansk a-lyd. I ordbogen noteres den, med korrekt IPA, *ɛ*, men Danias *á* ville tydeligere vise lyden, idet *ɛ* umiddelbart peger på noget e-agtigt, især hvis ordet også staves med <e> som i *hem*.

Naturligvis har bogen en lydskrifttavle, der i nogen grad afbøder eventuelt manglende pædagogik. Her forklares hvert tegn – stort set godt og rigtigt. Og dét bør noteres. Vi har kun følgende mindre indvendinger:

Den ærgerligste er en trykfejl. Ordet *själ* skal være *sjæl*! Forklaringen er ellers glimrende: Svensk tj-lyd (*tjäna, kär*) lyder praktisk talt som dansk sj-lyd nu, hvor svensk klusil-udtale er stærkt på

retur. Dette er meget lumsk, fordi svensk også har en sj-lyd, der er helt anderledes (*själ*), skrevet *ſ* – vi ville have foretrukket IPA's *ʃ*, som netop er beregnet til den nu almindeligste svensk sj-variant.

Svensk [y(:)] siges at være som dansk *y*, men den er mere lukket og fremskudt, så at den ofte får en i-klang i danske ører. Det er en meget almindelig dansk fejl at udtale langt svensk *y* og *u* ens.

Endelig beskrives *œ* (*göra, mörk*) som en meget åben *ø*-lyd, skønt den er en åben midttungevokal hen ad *å*.

Men én ting er disse generelle skrift→lyd-korrespondenser, noget andet er, om det enkelte lemma har den rette udtale.

Vi bemærker først, at man har været for nær igang med udtaleangivelser. Manglende lydskrift skulle betyde, at ordet udtales helt simpelt efter de svenske skrift→lyd-regler (som brugeren ganske vist ikke kender). Det er dristigt at antage, at brugeren af sig selv "forlænger" trykstærk *V* i *bibringa, nymåne* etc., eller i *bidé* 'bidet'. – Har man snuset en smule til svensk, vil man indsætte *ç* i *bikini*, og har man snuset lidt mere, vil man være i tvivl. (Det må dog siges, at ordbogen ellers er påpasselig med *ç/k*). Brugeren kan heller ikke vide, om den danske "V-forkortelse" ved *-ini* også gælder i svensk, hvad den ikke gør. I *ej-* og *øj-*diftongerne forudsættes det, at brugeren overfører sin danske regel om vokal-korthed, samtidig med at han *ikke* overfører sin regel om *a-* hhv. *å-*udtale. Endnu mere krævende er det, at den danske bruger af sig selv skal vide, at der er langt mørkt *a* i *tavla* og *varna*, men kort lyst *a* i *varm* etc. Helt umuligt er det ham at kende den normale udtale af *varit*, nemlig [ˈvaʔ:], ordbogen giver kun den mere formelle skriftnære. Også de topfremvente tryksvage former af *ta, var, vad, ha, ska* på udlydende kort lyst *a* mangler; de er uforudsigelige.

Ét er mangler, noget andet fejl. Trods nidkær søgen har vi kun fundet to: Navnet *Olle* har [u], ikke [ɔ], og eden *fan* udtales (helt overvejende) med det fonotaks-stridige lange *lyse a*, der hænger sammen med ordets interjektionelle brug. Foruden dette kender vi kun krageskrigets *kra* med langt lyst *a*, men dette ord er udeladt.

Og så helligbrøden: Man har droppet de svenske ordtoner. Dem er der ellers skrevet mere om end om alle sproglydene tilsammen. At udelade ordtonerne er at udstille sit bogstavfikserede syn på sprog. Det er værre end, om man havde udeladt sproglyden f, dét ville kun have ramt en lille brøkdelt af ordene, men samtlige ord har en ordtone. Ordtonen præger markant hvert ord fra først til sidst og bestemmer endda trykket, idet accent 2 giver et bitrykslignende tryk på en posttonisk stavelse, mens accent 1 bevirker svagtryk, fx ²*kom,ma* 'komme' vs. ¹*kom,ma* 'komma' (el. accent 2). – De færreste interesserer sig for ordtoner, javel, men nogle gør, og hvis fx vi slår et ukendt ord op, vil vi gerne straks oplyses om accenten og ikke tvinges over i andre værker.

8. Konklusioner

Selv om vi her har begået den klassiske anmelderfejl at fokusere mere på det kritisable end det rosværdige, så lad det være sagt utvetydigt: Ordbogen repræsenterer et tigerspring for svensk leksikografi med dansk. Den er både omfattende, brugervenlig og indeholder utroligt meget nyt. Vores dækningsgradsundersøgelse – med de mangler, den nu opviser – dokumenterer, at ordbogen i meget høj grad lever op til sin hovedopgave: at levere svar på svenske ords og udtryks betydninger for danskere. Der er virkelig mange succesopslag, også hvad angår kollokationer og idiomer.

Vi har nogle kritikpunkter bl.a. vedrørende sprog vs. encyklopædi, overflødige ord og udtale. Men disse forbehold skal ikke skygge for det generelle indtryk af et solidt og vellykket leksikografisk arbejde, der er resultatet af en metodisk bevidst og erfaren redaktions indsats, der bl.a. hviler på skuldrene af andre storprojekter som Den Danske Ordbog og ODS-Supplementet.

Vi har ganske vist *arbejdet* med ordbogen, og den er et glimrende arbejdsredskab. Men vi har også nydt den guidede lystvan-

dring i moderne svensk og gang på gang undret os over, hvor mange store og små forskelle der er mellem de to så nært forbundne sprog. Det dokumenterer værket også på fornemste vis.

Litteratur

- Brink, Lars 2008a: Arkaiserende – Gammeldags – Historisk. Tre slags ælde. I: *Mål og Mæle* 3, 17-20.
- Brink, Lars 2008b: Leksikalske huller. I: *Nordiske Studier i Leksikografi 10. Rapport fra konference i Akureyri maj 2007*. NFL: Reykjavík, 61-67.
- Brink, Lars 2010: Pålydende-betydning i ordbøger. I: *Nordiske Studier i Leksikografi 11*. NFL: Tammerfors 2010, 113-117.
- Farø, Ken 2005: Fiktionymi i bilingval leksikografi. I: *Nordiske Studier i Leksikografi 8. Rapport frå Konferanse om leksikografi i Norden. Volda 20.-24. mai 2003*. Red.: Ruth Vatvedt Fjeld & Dagfinn Worren. NFL: Oslo, 120-132.
- Farø, Ken & Adam Hyllested 2007: *Ordbog, leksikon eller encyklopædi? Et koncept for et dansk lingvistisk opslagsværk*. I: *Lexico-Nordica* 14, 107-124.
- Kalle Ankas Pocket 326: Storm* (2004). Malmö: Kärnan.
- Sjöwall, Maj & Per Wahlöö 1970: *Polis, polis, potatismos*. Stockholm: Norstedts.
- Solsjenitsyn, Aleksander 1969/1974: *Den första kretsen*. Stockholm: W & W.

Lars Brink
professor i nordisk sprog og
litteratur
Elmevænget 30
DK-2880 Bagsværd
labri@webspeed.dk

Ken Farø
adjunkt i tysk sprog, ph.d.
Institut for Engelsk, Germansk og
Romansk
Københavns Universitet
Njalsgade 130
DK-2300 København S
kenfaroe@hum.ku.dk

Finska ord på finska för invandrare

Ilse Cantell & Riitta Korhonen

Saarikalle, Anne – Vilkuna, Johanna: *Suomen kielen sanakirja maa-hanmuuttajille*. Helsingfors: Gummerus 2010. 328 sidor, ca 3 500 uppslagsord. 22 EUR.

1. Invandrare i Finland

I Finland har vi inte haft en lika omfattande arbetskraftsinvandring som i t.ex. Sverige, dit ju invandringen började redan på 1960-talet. Vi har därför inte invandrare från t.ex. Grekland och Turkiet närapå i samma utsträckning som Sverige. För antalet talare av främmande språk i Finland ser tio i topp-listan år 2010 ut så här: 1. Ryska (54 559 personer¹; i Sverige på 21 plats²), 2. estniska (28 493; 18:e), 3. somaliska (12 985; 10:e), 4. engelska (12 855; 17:e), 5. arabiska (10 415; 3:e), 6. kurdiska (8 032; 9:e), 7. kinesiska (7 546; 40:e), 8. albanska (7 113; 11:e), 9. thai (5 722; 13:e), 10. vietnamesiska (5 637; 19:e). Den största invandrargruppen i Sverige är finländare. Också i Finland är det grannar som toppar invandrarstatistiken: ryssar och ester. Det kan antas att det för de språk som talas i grannländerna finns läromedel och ordböcker och att man därför inte behöver vara helt okunnig i språket då man flyttar till ett nytt land.

Det är dock inte fallet när det är fråga om flyktingar från länder som ligger långt borta från mottagarlandet. Språkstudier är kanske inte heller det första man tänker på då man söker överlev-

1 Befolkningsstruktur 2010, Statistikcentralen.

2 Parkvall (2010:142)

nadsmöjligheter på annat håll. Därför blir studier i det nya språket en del av integrationen i det nya landet, och mottagarlandet måste ordna undervisning.

2. Vilka behov ska en ordbok för invandrare fylla?

Förutom att ”vanliga” invandrare har ett behov av att lära sig det nya landets språk behövs det också ordböcker för tolkning och översättning vid kontakter mellan invandrare och t.ex. myndigheter, skolor, banker mm. i det nya landet.

Då det gäller ryska, estniska och engelska finns det ett gott utbud av tvåspråkiga ordböcker i Finland. För de andra språken på Finlands tio i topp-lista är utbudet knappare.

Under de sista åren har det i Finland kommit ut många läroböcker i finska som riktar sig till finskinlärare i olika åldrar och med olika språklig bakgrund. Vanligen innehåller läroböckerna en ordlista med finska förklaringar och ekvivalenter på det andra språket. Finska ord kan också sökas på Internet: det finns t.ex. ordlistor till stöd för yrkesinriktad undervisning. I dessa har målet varit att använda ett förklaringspråk som är tydlig och lättläst finska.

Det förefaller alltså behövas två slag av lexikografisk språkhjälp för invandrare, dels ordböcker för ett rätt elementärt språkinlärningsbehov, dels ordböcker och ordlistor för rätt komplicerade kontaktsituationer där användarna är professionella språkbrukare, t.ex. tolkar. Hur ska en ordbok fylla dessa behov? Är det över huvud taget möjligt att ta fram en ordbok för båda behoven? Enligt de erfarenheter man i Sverige haft med Lexin-projektets ordböcker ska det vara möjligt: Lexin-ordböckerna används både av ”vanliga” invandrare som lär sig svenska och av tolkar och översättare.³

3 Muntligt från Birgitta Romppanen.

3. Ordböcker i finska för invandrare

Det finns inte många egentliga inlärningsordböcker för personer som vill lära sig finska. Timo Nurmis *Suomen kielen sanakirja ulkomaalaisille* (Finsk ordbok för utlänningar) (Gummerus 1999) och en utvidgad version, *Suomen kielen sanakirja opiskelijoille ja ulkomaalaisille* (Finsk ordbok för studerande och utlänningar) (Gummerus 2009) baserar sig på en enspråkig ordbok som är avsedd för finskspråkiga användare. De är utan tvekan nyttiga för längre hunna studerande, men svåra för sådana som bara börjat lära sig finska. Också Vesa Jarvas och Timo Nurmis *Oikeeta suomee* (Finska på riktigt) (2006) med talspråkliga uttryck och fraser passar bättre för en längre hunnen språkinlärare. Ända fram till nu har vi i Finland helt och hållet saknat en enspråkig ordbok som är gjord enkom för inlärare som inte kan så mycket finska och i vilken förklaringarna ges på tillräckligt enkel finska. Anne Saarikalle och Johanna Vilkuna vill med sin *Suomen kielen sanakirja maahanmuuttajille* (Finsk ordbok för invandrare) (Gummerus 2010) fylla en uppenbar lucka.

Ordboken innehåller enligt baksidestexten det centrala ordförråd som behövs vid inläring av finska språkets grunder och ska bedömas som en sådan produkt. Frågan om det är möjligt att fylla ordboksbehoven både för språkinlärare och professionella användare förblir alltså än så länge obesvarad för finskans del.

4. Vad lovar den finska invandrarordboken?

I inledningen till *Suomen kielen sanakirja maahanmuuttajille* nämns det att ordboken är avsedd att fungera som en kompletterande kunskapskälla för inlärare av finska i alla åldrar, både för sådana som lär sig grunderna och för längre hunna språkinlärare.

Vidare nämns det att den är avsedd att användas både självständigt och i klass. Dessutom ska den vara lämplig också för andra som behöver en ordbok med lättläst språk; målgruppen definieras inte, men lättläst språk används bl.a. för kommunikation med personer som har svårigheter att förstå abstrakt språk. Ordbokens inledning och användarvägledning är dock skrivna på allmänfinska, inte med förenklat språk. Instruktionerna är kanske inte avsedda för ordbokens användare?

Ordboken är enspråkig. Enligt förordet passar den för alla invandrare oberoende av modersmål. Författarna är medvetna om att användarna kanske inte är vana vid att använda enspråkiga ordböcker, och därför kan den här ordboken enligt dem gott användas vid sidan av tvåspråkiga ordböcker.

Enligt inledningen utgör ordboken en för invandrarnas behov bearbetad version av barnordboken *Eka oikea suomen kielen sanakirja* (Den första riktiga finska ordboken) (Gummerus 2005). Bland uppslagsorden ingår förutom finskans centrala ordförråd också ord som är ”speciellt viktiga för invandrare”. Dessutom sägs det att uppslagsorden förklaras ”på ett enkelt sätt” och att ”bekanta ord” används.

Ordboken innehåller en inledning och en grafisk och en skriftlig användarvägledning. I slutet finns några bilagor: finskans alfabet med uttalet av bokstäverna, räkneord, väderstreck, veckodagar, månader och sist en förteckning över Europas stater, deras huvudstäder och invånarbetesbeteckningar.

En enspråkig ordbok som är avsedd för invandrare är per definition en inlärningsordbok, och av detta följer att användarna förväntar sig hjälp både med förståelse av det finska ordförrådet och vid produktion. Det förutsätter en tillräckligt åskådlig beskrivning av ordens betydelser och deras användning. Dessutom förväntas en inlärningsordbok erbjuda en del encyklopediska uppgifter för de omtalade företeelserna. Allt detta finns i den finska invandrarordboken, men helheten håller inte vad som lovats. Till och med

för en längre hunnen språkinlärare är det inte alltid möjligt att förstå alla förklaringar och exempel.

5. Ordboksartiklarnas innehåll

5.1. Mikrostrukturen

Den finska invandrarordboken är med sina ca 3 500 uppslagsord en mycket liten ordbok som inte kommer nära antalet uppslagsord för en normal tvåspråkig fickordbok.

Uppslagsorden är tryckta i fet stil och kommer tydligt fram på ordbokens sidor. Då ett ord har olika betydelser ges betydelserna i egna numrerade avsnitt. Efter betydelsebeskrivningen (förklaringen) kommer en eller flera kursiverade exempelfraser efter en triangel (►). Eventuella synonymer ges efter markören *myös* 'också' och motsatsord markeras med ordet *vastakohta* 'motsats'. Dessa tillägg är mycket nyttiga med tanke på inläring av finskans ordförråd. Synonymer och antonymer kunde ha getts i större utsträckning än vad nu är fallet. Nu förekommer de bara sporadiskt. Sist i ordboksartiklarna avskilda med markören *sanontoja* ('fraser') finns ett avsnitt med "uttryck där uppslagsorden inte har riktigt den betydelse som anges i förklaringen" (s. 8). I artiklarna ges betydelsebeskrivningen utan förtydligande radbyten och mellanrum mellan raderna. Detta gör att texten inte är så lättläst som förordet lovar. Ett exempel (svensk översättning här och i fortsättningen med tanke på icke-finskkunniga, ordboken är enspråkigt finsk):

avata 'öppna'

Kun avaat esimerkiksi oven tai laatikon, se ei ole enää kiinni. ► *Liisa avasi oven minulle. Avaa silmäsi ja kerro, mitä näet.* MYÖS aukaista VASTAKOHTA sulkea.

'Då du öppnar till exempel en dörr eller en låda, är den inte längre stängd. ► Liisa öppnade dörren åt mig. Öppna ögonen och berätta vad du ser.'

För uppslagsorden ges inga grammatiska upplysningar. Det finns heller inga upplysningar om de ändringar som sker i stammen då orden böjs. Dessa formväxlingar kommer fram bara i exemplen, t.ex. *vesi – vettä*:

vesi 'vatten'

Vesi on nestettä, jota on järvissä ja merissä. Vettä myös tip-puu maanpinnalle, kun sataa.

'Vatten är den vätska som finns i sjöarna och haven. Vatten faller också ner till jordytan då det regnar.'

De centrala temaformerna (t.ex. *vesi: veden, vettä, veteen*⁴) kunde ha placerats inom parentes efter uppslagsordet. På så vis skulle växlingarna i stammen komma fram på en gång.

Uppslagsorden är till största delen neutrala allmänspråkliga ord. Men det finns också vardagligare uttryck som inte markerats för sitt stilvärde. Till exempel orden *bileet* 'party, skiva, fest', *dek-kari* 'deckare' och *pussata* 'pussa' ger sken av att vara neutrala ord på samma sätt som alla andra uppslagsord. Ur inlärningssynvinkel skulle en stilistisk markör i samband med förklaringen ha varit på sin plats.

I en typisk inlärningsordbok finns också annan information än beskrivningar av ords betydelser och användning. Ofta är det nödvändigt att beskriva själva företeelserna för personer som har en annan bakgrund. I samband med ordet *armeija* 'armé' anger

4 Temaformerna av ordet *vesi* 'vatten' är nominativ, genitiv, partitiv och illativ. Dessa former ges i den enspråkiga finska ordboken *Kielitoimiston sanakirja* (2006).

ordboken exempelvis vem som ”går i armén”, dvs. gör militärtjänst i Finland och vad utbildningen går ut på. Förklaringar ges också om till exempel Finlands riksdag (*eduskunta*) och regering (*hallitus*) och vidare om julen (*joulu*), midsommaren (*juhannus*), gummistövlar (*kumisaappaat*) och bastuugnen (*kiuas*) samt hinduism (*hindulaisuus*), islam (*islam*) och kristendom (*kristinusko*).

5.2. Definitionsspråket

Enligt baksidestexten är uppslagsorden förklarade ”enkelt och lättförståeligt med hela meningar”. Man kan gissa att detta avser någon typ av definitioner i COBUILD-format⁵. Ett exempel på en sådan definition är artikeln för verbet *äänestää*:

äänestää ’rösta’

Kun ihminen *äänestää*, hän antaa *äänensä* vaaleissa parhaalle henkilölle. *Äänestäjä* kirjoittaa paperille numeron.

’Då en person röstar, ger han eller hon i val sin röst till den bästa personen. Väljaren⁶ skriver ett nummer på ett papper.’

Men definitionen blir i detta sammanhang närmast naiv och något missvisande. Det är ju den person väljaren *tycker är bäst* han eller hon röstar på.

Förklaringarna är konsekvent hela meningar och detta leder ibland till konstlade formuleringar. Adjektiven förklaras till exempel på minst tre olika sätt: 1) ”Då en människa/sak är liten, (är) han eller hon/den...”, 2) ”En människa/sak är liten, om...” och 3) ”En liten människa/sak är...”. Vi har inte hittat någon systematik i valet mellan de olika sätten, det verkar bara allmänt inkonsekvent.

5 Definitioner som är delade i två delar så att lemmat i den första delen ges i en typisk kontext och i den senare delen får en betydelsebeskrivning (NLO).

6 Finskans *äänestäjä* är en avledning av verbet *äänestää* ’rösta’, dvs. direkt översatt ’röstare’.

En del av de ord som används i förklaringarna saknas som uppslagsord i ordboken. Ett exempel:

ajankohtainen 'aktuell'

Kun asia on ajankohtainen, se on tärkeä juuri nyt. Ajankohtaisista asioista kirjoitetaan paljon esimerkiksi lehdissä.

'Då en sak är aktuell, är den viktig just nu. Om aktuella saker skrivs det mycket till exempel i tidningarna.'

Orden *tärkeä* 'viktig', *kirjoittaa* 'skriva' och *lehti* 'tidning' finns med som uppslagsord i ordboken, av homograferna *juuri* finns bara substantivet 'rot' men inte adverbet 'just, precis'. Orden *asia* 'sak; fråga; angelägenhet; ärende' m.fl. betydelser, *nyt* 'nu', *paljon* 'mycket', *esimerkiksi* 'till exempel' saknas helt och hållet, också i nominativ.

Substantiven förklaras antingen genom att ordet upprepas i början av förklaringen eller genom att det berättas att någon har saken i fråga eller så att tillverkningssättet nämns eller med hjälp av exempel. Som ett tillägg till förklaringen ges ofta encyklopediska upplysningar. Ett exempel:

ajokortti 'körkort'

Jos ihmisellä on ajokortti, hän saa ajaa autoa. Ajokortin saa sen jälkeen, kun on käynyt autokoulun ja opetellut ajamaan.

'Om en person har körkort, får han eller hon köra bil. Man får körkort efter att man gått i bilskola och lärt sig köra.'

Många substantiv får på samma sätt som adjektiven en onödigt besvärlig förklaring, och en del av de ord som ingår i definitionerna saknas som uppslagsord, t.ex.

haava 'sår'

Haava on kohta, josta iho menee rikki, kun loukkaat itsesi.

Haavasta tulee yleensä verta.

'Ett sår är ett ställe där huden går sönder då du skadar dig.

Vanligen blöder ett sår.'

I ordboken saknas orden *kohta* 'ställe', *rikki* 'trasig, sönder', *yleensä* 'i allmänhet, vanligen' och *itse* 'själv'. Ordformerna *loukkaat* (2 person singularis) och *verta* (partitiv singularis) måste användaren kunna söka under uppslagsorden *loukata* ('skada, sår') och *veri* ('blod').

På några ställen är förklaringarna utformade så att de fäster läsarens uppmärksamhet på någonting som inte är väsentligt eller rentav är missledande med tanke på ordet som förklaras. På basis av förklaringen och exempen som valts för artikeln *tapa* kunde man få för sig att ordet främst har negativ betydelse, dvs. 'ovana':

tapa 'vana'

Jos sinulla on tapa esimerkiksi syödä kynsiäsi, olet tottunut syömään niitä.

'Om du brukar till exempel bita på naglarna, har du för vana att bita på dem.'

► *Ricolla on tapana juoda teetä maidon kanssa. Onko sinulla pahoja tapoja?*

'Rico brukar dricka te med mjölk. Har du några dåliga vanor?'

Förklaringarna av verben är de mest oenhetliga. De kan bl.a. ibland stå i andra person singularis, ibland i passiv. Växlingen förklaras inte på något sätt i inledningen och man kan inte heller sluta sig till varför det ena eller det andra alternativet valts utgående från artiklarna. Valet mellan de olika alternativen är inte helt lätt. Fin-

skans andra person singularis kan nämligen ofta verka naivt eller onaturligt i text, och i passiv är tolkningen av subjektet alltid öppen.

syödä 'äta'

Kun syöt, ruokaillet tai popsit esimerkiksi karkkia.

'Då du äter intar du en måltid eller mumsar i dig till exempel godis.'

► *Lounas syödään kello 12. Kissani tykkää syödä kalaa. Mira söi ravintolassa ystäviensä kanssa.*

'Vi äter lunch klockan 12. Min katt tycker om att äta fisk. Mira åt på restaurang tillsammans med sina vänner.'

juhlia 'fira; festa'

Kun juhliitaan, vietetään jotakin juhlaa tai pidetään muuten hauskaa. Erilaisia juhlapäiviä vietetään eri tavoilla.

'Då man festar har man en fest eller har trevligt på annat sätt. Olika högtidsdagar firas på olika sätt.'

► *Jouluna juhliitaan Jeesuksen syntymää. Naapurissa juhliitiin eilen illalla.*

'Julen firas till åminnelse av Kristi födelse. Det festades hos grannen i går kväll.'

I baksidestexten och i inledningen nämns att ordboken innehåller "ord som är speciellt viktiga för invandrare". Sådana är enligt författarna tydligen ord som anknyter till kontakter med myndigheterna. Men den här typen av ord har inte fått bättre förklaringar än de andra uppslagsorden, något användarna säkert hade förväntat sig. Att förstå en del av förklaringarna är en utmaning t.o.m. för en modersmålstalande, t.ex.

paluumuuttaja 'återinvandrare'

Paluumuuttaja on entinen tai nykyinen Suomens kansalainen, joka har asunut toisessa maassa ja palaa Suomeen. Myös inkerinsuomalaiset maahanmuuttajat ovat paluumuuttajia.

'En återinvandrare är en tidigare eller nuvarande finsk medborgare som bott i ett annat land och återvänder till Finland. Också finskingermanländska invandrare är återinvandrare.'

Orden *asua* 'bo' ('då du bor på ett ställe, är ditt hem där'), *maa* 'land', *palata* ('återvända', grundformen av presens 3. personsformen *palaa*) och *maahanmuuttaja* ('invandrare') finns på annat ställe i ordboken, men inte de lika viktiga orden *entinen* 'tidigare', *nykyinen* 'nuvarande', *toinen* 'annan', *inkerinsuomalainen* 'finskingermanländsk' eller *kansalainen* 'medborgare' (ordet *kansalaisuus* 'medborgarskap' finns visserligen med, men man hade i första hand väntat sig att det som tas upp som lemma är grundordet och inte avledningen).

Enligt användarhandledningen återfinns det i slutet av ordboksartiklarna under rubriken **sanontoja** ('fraser') "uttryck där uppslagsordet inte har exakt den betydelse som anges i förklaringen". Under rubriken finns t.ex. överförda betydelser och i de fallen förklaras bara den överförda betydelsen av själva uppslagsordet:

sammua 'slockna': Mika sammui [= nukahhti] sohvalle.
'Mika slocknade [= somnade] på soffan.'

kiitos 'tack': Kiitos lumisateen [= koska satoi lunta] lapset pystyivät tekemään lumiukon. 'Tack vare att det snöade [= eftersom det snöade] kunde barnen göra en snögubbe.'

Men under samma rubrik finns också idiom som får en utförligare förklaring:

nappi -- jos kaikki menee nappiin (= onnistuu täydellisesti), talo on valmis kesällä. 'Om allt går väl, är huset färdigt i sommar.'

6. Vem är ordboken avsedd för?

Suomen sanakirja maahanmuuttajille kan säkert vara till hjälp för många som redan kan rätt mycket finska. För att förstå betydelseangivelserna måste användaren till exempel känna finskans passivböjning och finskans olika tempusformer. Under handledning av en lärare på en lektion kan ordboken nog också konsulteras av en nybörjare.

Om man i den recenserade ordboken hade utnyttjat illustrationerna ur utgångsordboken *Eka oikea suomen kielen sanakirja*, som alltså är riktad till barn, hade helheten varit något klarare också för en inlärare som inte kan så mycket ännu. Men inte ens bilder hade kunnat rädda en del av de klumpiga förklaringarna. För många av de förklaringar som är barnvänliga i en barnordbok verkar bara naiva i en ordbok för vuxna.

Vi recensenter förundrar oss också över de talrika namnen som används i förklaringar och exempel. Det kan vara svårt för en främmandespråkstalare att uppfatta att det är fråga om namn och inte om ord. Sådana fall är t.ex. *Pihla* och *Veikka*. Det hade kanske varit mer lyckat att i stället för den onödiga mångfalden av namn använda t.ex. *Matti* och *Maija* genomgående i artiklarna. Många av namnen i ordboken är på något sätt speciella och fäster läsarens uppmärksamhet vid oväsentligheter.

7. Sammanfattning

Trots all kritik vi kommit med är Saarikalles och Vilkunas ordbok ett försök i rätt riktning och den gör sitt för att fylla ett behov. Ordboken visar dessutom att det av en författare till en inlärningsordbok krävs rätt mycket annat än lärarerfarenhet i språket i fråga.

Den lilla ordboken täcker en mycket liten del av det finska ordförrådet. Med tanke på att användarna kanske ändå måste kunna finska i någon mån, hade det varit bra om det belysta ordförrådet varit ganska mycket större.

Framför allt saknar vi bland uppslagsorden ett stort antal ord som ingår i definitionerna och förklaringarna. Frånvaron av dem är ett problem och det ringa omfånget är ingen ursäkt. Avsaknaden av definitionsorden verkar närmast vara en följd av att lemmalistan inte gått igenom ordentligt och med eftertanke. På grund av denna brist måste användaren antingen kunna finska rätt bra eller ha tillgång till en annan, gärna tvåspråkig ordbok. Men det finns inte så många och framför allt inte många gedigna ordböcker mellan finskan och de flesta invandarspråken i Finland.

Litteratur

- Befolkningsstruktur 2010. Statistikcentralen. www.stat.fi/til/vaerak/tau.html, avläst 7.4.2011.
- Jarva, Vesa & Timo Nurmi 2006: *Oikeeta suomee. Suomen puhekielen sanakirja* (Finska på riktigt. Ordbok i finskt talspråk). Jyväskylä: Gummerus.
- Kielitoimiston sanakirja* 1–3 (Finska språkbyråns ordbok). Huvudredaktör: Eija-Riitta Grönros. *Kotimaisten kielten tutkimuskeskuksen julkaisuja* 140. Helsinki: Kotimaisten kielten tutkimuskeskus 2006.

- NLO = H. Bergenholtz, I. Cantell, R.V. Fjeld, D. Gundersen, J.H. Jónsson & B. Svensén: *Nordisk leksikografisk ordbok*. Oslo: Universitetsforlaget AS 1997.
- Nurmi, Timo 1999: *Suomen kielen sanakirja ulkomaalaisille* (Finsk ordbok för utlänningar). Jyväskylä: Gummerus.
- Nurmi, Timo 2009: *Suomen kielen sanakirja opiskelijoille ja ulkomaalaisille* (Finsk ordbok för studerande och utlänningar). Jyväskylä: Gummerus.
- Parkvall, Mikael 2010: *Invandrarspråk*. I: Östen Dahl & Lars-Erik Edlund (red.): *Språken i Sverige*. Stockholm: Norstedts Förlagsgrupp AB, 142–147.
- Saarikalle, Anne & Johanna Vilkuna 2005: *Eka oikea suomen kielen sanakirja* (Den första riktiga finska ordboken). Jyväskylä: Gummerus.
- Saarikalle, Anne & Johanna Vilkuna 2010: *Suomen kielen sanakirja maahanmuuttajille* (Finsk ordbok för invandrare). Jyväskylä: Gummerus.

Ilse Cantell
forskare, fil. lic
Forskningscentralen för de
inhemska språken
Berggatan 24
FI-00100 Helsingfors
ilse.cantell@focis.fi

Riitta Korhonen
specialforskare, fil. dr
Forskningscentralen för de
inhemska språken
Berggatan 24
FI-00100 Helsingfors
riitta.korhonen@kotus.fi

Norstedts ordboksappar

Louise Holmer

Norstedts engelska ordbok Pro, version 1.3

Systemkrav: iOS 3.2 eller senare, kompatibel med iPhone, iPad och iPod touch. Storlek: 14,3 MB. Pris: 349 SEK. Utvecklad av isolve AB 2010. Norstedts Förlagsgrupp AB, Stockholm 2010.

Norstedts svenska fickordbok, version 1.0

Systemkrav: iOS 3.1 eller senare, kompatibel med iPhone, iPad och iPod touch. Storlek: 1,8 MB. Pris: 55 SEK (men har sålts till bokreapriset 7 SEK). Utvecklad av isolve AB 2010. Norstedts Förlagsgrupp AB, Stockholm 2010.

En liten ordbok för dig som är Kär & Galen

Systemkrav: iOS 3.1 eller senare, kompatibel med iPhone, iPad och iPod touch. Storlek: 0,9 MB. Gratis. Utvecklad i samarbete med isolve AB. Norstedts Förlagsgrupp AB, Stockholm 2011.

1. Inledning

I den här recensionen behandlas några nytillskott på den svenska ordboksmarknaden, nämligen tre av Norstedts ordboksappar. Vad är då en app? Själva ordet har använts i svenskan i ett antal år men har inte blivit vanligt i allmänspråket förrän under senare tid. Ordet *app* fanns med i Språkrådets lista över nyord 2010 och definitionen lyder ”program som laddas ned från nätet till en mobiltelefon [efter engelskan, av *application*]” (Språkrådet 2010). Ordet finns inte med i *Svensk ordbok utgiven av Svenska Akademien* (SO 2009) men är nyinlagt i den kommande upplagan av *Svenska Akademiens ordlista* (SAOL) som är under utarbetning. Språkrådets definition skulle dock kunna utökas något eftersom appar

även kan laddas ner till exempelvis surfplattor, som inte primärt är gjorda för att man ska kunna ringa med dem. Surfplattor och s.k. smarta telefoner, alltså mobiltelefoner med fler datorliknande egenskaper än traditionella mobiltelefoner, har ökat i antal framför allt de senaste åren. Med de nya produkterna kommer även en mängd olika tilläggsfunktioner och marknaden för olika betalprogram till mobiler och surfplattor har ökat enormt. Som exempel kan nämnas att företaget Apple har ca 350 000 tillgängliga appar och det Google-ägda Android ca 310 000 (uppgifter från april 2011, se vidare Wikipedias artikel ”List of digital distribution platforms for mobile devices”).

Många appar kräver internetåtkomst för att kunna fungera, som t.ex. sådana som är gps-baserade, medan andra inte gör det. De ordboksappar som är aktuella i den här recensionen fungerar så att de laddas ner och lagras lokalt i telefonen. Sedan kan man använda sin ordboksapp i mobilen utan att telefonen har internetåtkomst. Därmed har ordboksapparna en given plats på marknaden för dem som befinner sig på resande fot och inte vill eller kan använda telefonens internetmöjligheter.

En app kan vara gratis eller kosta allt mellan några få svenska kronor och flera hundra dollar, även om det senare inte är särskilt vanligt för de populäraste apparna. De ordboksappar som finns på marknaden är i regel betalappar även om det också finns ett flertal gratisvarianter. Priset för en ordboksapp ligger ofta relativt högt; de flesta kostar mellan 50 och 200 SEK. Som jämförelse kan sägas att de vanligaste spelen för mobiltelefoner och surfplattor kostar ca 7 SEK och lite mer avancerade program ca 15 SEK. Vanligt är också att ordboksförlagen har någon eller några gratisappar för att locka kunder och några betalappar med fler och bättre funktioner än gratisvarianterna.

Apparna utgör alltså ännu ett sätt att profilera sig för ordboksförlagen på den digitala marknaden, både i Sverige och internationellt, när pappersupplagornas försäljning minskar.

2. Ordboksappar i Sverige, Danmark och Norge

Internationellt sett finns det, och har funnits sedan ett tag tillbaka, ett flertal ordboksappar, särskilt tvåspråkiga. För den svenska ordboksmarknaden är det dock relativt nytt med appar, och de flesta som finns på marknaden utgörs av Norstedts ordböcker anpassade för Apples produkter iPhone, iPad och iPod touch. De appar som finns att köpa eller ladda ner gratis i nuläget (april 2011) är *Norstedts engelska ordbok Pro* (svenska/engelska), övriga stora tvåspråkiga ordböcker (svenska/spanska, svenska/franska, svenska/tyska och svenska/italienska), fickordböcker (svenska, svenska/engelska, svenska/spanska, svenska/franska, svenska/tyska och svenska/italienska) och ett antal ordquiz på de olika språken. Dessutom finns även *En liten ordbok för dig som är Kär & Galen*, som lanserades till Alla hjärtans dag 2011.

Förutom de ovanstående befintliga apparna har förlaget också annonserat att kommande lansering innehåller appar för operativsystemet Android och att även *Norstedts stora svenska ordbok* är under utarbetande.

Den trettonde och senaste upplagan av SAOL finns som app till Android-telefoner. Appen är utvecklad av en privatperson och utgivningen är inte sanktionerad av Svenska Akademien. Den appen kommer inte heller att behandlas vidare här.

Beträffande den övriga skandinaviska marknaden har t.ex. Gyldendal i Danmark släppt sina ordboksappar, hittills sju stycken, för iPhone och Android. Gyldendal har ett antal tvåspråkiga ordböcker men också *Dansk Ordbog* (en app jag själv använder i Android-version) som innehåller *Retskrivningsordbog*, *Fremmedordbog* och *Synonymordbog* i en och samma app. Norska Kunnskapsforlaget har t.ex. *Engelsk blå ordbok* och ett antal tvåspråkiga ordböcker för Apples produkter. Prismässigt ligger de danska och norska apparna lägre än vad de svenska gör – Gyldendals *Engelsk Ordbog Large* kostar 249 SEK i App Store (Apple 2011a) och Kunn-

skapsforlagets *Engelsk blå ordbok* kostar 199 SEK (Apple 2011b), jämfört med *Norstedts engelska ordbok Pro* som alltså kostar 349 SEK (Apple 2011c).

Många av de stora engelska enspråkiga ordböckerna har funnits som appar sedan år 2008. En jämförelse mellan de tre stora ordböckerna och deras appvarianter *American Heritage Dictionary*, *The Concise Oxford Dictionary* och *Webster's Third New International Dictionary* finns att läsa i Boychuck (2008). Boychuck jämför de tre och kommer fram till att *American Heritage Dictionary*, utvecklad av Enfour, är den ledande appen när det gäller grafik och funktionalitet.

En engelskspråkig ordboksapp jag själv har använt mig av är *Merriam-Webster's Dictionary* för Android, en gratisapp som passar mina behov väl – slå upp ett ord och lyssna på det engelska uttalet.

Här bör nämnas att utbudet av appar ändras otroligt snabbt och att det troligen kommer att finnas en mängd nya ordboksappar på den svenska marknaden inom en snar framtid. Det är mycket möjligt att t.ex. funktioner som saknas i de befintliga apparna kommer att finnas i senare versioner som släpps innan denna recension har publicerats, eller att priser har ändrats.

3. Befintliga funktioner hos de undersökta apparna

De appar jag har testat inför den här recensionen är *Norstedts engelska ordbok Pro*, *Norstedts svenska fickordbok* och *En liten ordbok för dig som är Kär & Galen*, samtliga i både iPhone och iPad.

En iPhone är alltså Apples smarta telefon som nu finns i version 4. Skärmen är en pekskärm och har storleken 3,5 tum. iPad är en surfplatta från Apple, även den med pekskärm. Skärmen är 9,7 tum, betydligt större än på iPhone. Båda produkterna har så-

dana funktioner att skärminnehållet ”följer med” när man vrider på telefonen. Man kan alltså ändra sin telefon eller platta mellan stående och liggande läge och få innehållet på skärmen att orientera sig därefter.

Apparna är testade på den första versionen av iPad. Ingen av de ovan nämnda apparna är testad på iPod touch, även om de fungerar där också. (iPod touch är en mediaspelare från Apple, också den med den mindre pekskärmen.)

I recensionen uppehåller jag mig vid funktionaliteten hos apparna, alltså formatanpassningen, snarare än att göra en utvärdering av ursprungsprodukten med avseende på lemmaurval och liknande.

3.1. Norstedts engelska ordbok Pro, version 1.3

Ordboken innehåller 276 000 ord, fraser och översättningar (Apple 2011c). Den är därmed till innehållet mycket lik den tryckta varianten *Norstedts engelska ordbok professionell* (2010) som kostar mellan 700 och 800 SEK i nätbokhandlarna, d.v.s. det dubbla jämfört med appversionen.

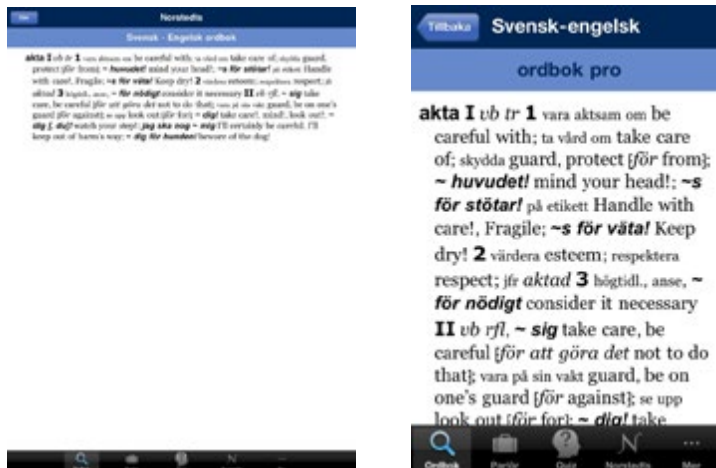
Utseendet är mycket avskalad, i både iPhone och iPad. Den första söksidan som öppnas i iPad är helt vit med endast en söknapp uppe i vänstra hörnet. För att börja söka behöver man klicka på Sök så att en sökruta och en rullmeny med lemmarna kommer fram. Där kan man välja vilket språk man vill söka i (Svenska – Engelska eller Engelska – Svenska). Längst ner på sidan finns fem små ikoner för ”Ordbok”, ”Parlör”, ”Quiz”, ”Norstedts” och ”Mer”. Som default är inställningen ”Ordbok” medan man får klicka på skärmen för att komma vidare till de andra funktionerna. När man klickar på ”Parlör” förblir sidan åter helt vit tills man klickar på ”Sök”. Då rullas en meny upp där man väljer mellan kategorier som ”Artighetsfraser” m.m., ”Vanliga ord och fraser” osv.

För att komma till det eftersökta ordet måste man markera det

i rullistan; det räcker alltså inte att bara skriva in det i sökrutan och klicka på ”Sök”. När man får upp artikeltexten försvinner sökrutan och för att göra en ny sökning måste man klicka på ”Sök” (i iPad) eller ”Tillbaka” (i iPhone). Se figur 1 för exempel på gränssnitt och artikelvisning i både iPad och iPhone.

För varje uppslagsord får man en förkortad ordklassuppgift. Här hade jag gärna sett att man hade expanderat förkortningarna – eftersom endast en artikel kommer upp i taget finns det gott och väl utrymme för mer text, och ordklassuppgiften är knappast den som tar mest plats.

Skillnaden mot Norstedts engelsk/svenska och svensk/engelska ordbok som finns på internet (www.ord.se) är relativt stor. På nätversionen visas en förkortad form av varje artikel som man sedan kan välja att expandera genom att klicka på plussymbolen. Appens sätt att presentera artiklarna följer däremot den tryckta versionen av boken snarare än nätversionen, eftersom man inte kan välja hur mycket information som artiklarna ska visa. Se vidare Trap-Jensen (2010) för en redogörelse av för- och nackdelar med att expandera och minimera artiklar.



Figur 1: Skärmdump från *Norstedts engelska ordbok Pro*. Den vänstra bilden föreställer uppslagsordet *akta* i iPad och den högra bilden samma uppslagsord i iPhone.

Eftersom både iPad, iPhone och iPod touch (såväl som övriga smarta telefoner och surfplattor) är gjorda med tanke på att man ska kunna spela upp filmer och musik finns goda möjligheter för apputvecklarna att lägga till uppläst uttal. Detta är dock något som saknas i samtliga av de svensktillverkade ordboksapparna, vilket är synd. Den dansk-engelska och norsk-engelska appversionen har uppläst uttal på åtminstone ca 50 000 av lemman (Apple 2011a, Apple 2011b). De engelska enspråkiga apparna som finns erbjuder också i de flesta fall uppläst uttal som man får genom att klicka på en högtalarikon vid uppslagsordet. En möjlig nackdel är att appen då blir mer utrymmeskrävande, men Norstedts-apparna är ganska små i sina nuvarande versioner (Kunnskapsforlagets *Engelsk Blå Ordbok* är tio gånger så stor som Norstedts med 108 MB mot Norstedts 14,3 MB), så den utökade storleken skulle knappast vara ett problem. Internetversionen av *Norstedts engelska ordbok* erbjuder dock uppläst uttal på både engelska och svenska.

Norstedts engelska ordbok Pro är relativt avskalad till utseendet och lättanvänd. Styrkan ligger framför allt i den rena designen, men det är också den som gör att utseendet känns lite väl enkelt och faktiskt tråkigt. På iPad blir det helt enkelt alltför mycket vit-yta. Ytan skulle kunna utnyttjas till att visa fler artiklar än bara den uppslagna, som till exempel de följande fem artiklarna.

Artikeltexten är svart mot vit bakgrund. En utveckling av detta skulle kunna vara att olika färgmarkeringar används för att göra informationskategorierna tydligare för användaren.

En självklar funktion som saknas är en tillbaka-knapp i iPad-versionen. För att göra en ny sökning klickar man helt enkelt på "Sök igen" när man befinner sig i ordboksläget eller på "Mer" när man befinner sig i "Mer"-avdelningen ("Karta", "Ordbokens uppbyggnad" m.m.) i stället för att klicka på en tillbakapil eller en tillbakaknapp. Som jämförelse kan nämnas att *Dansk Ordbog* för Android har framåt- och bakåtpil såväl som upp- och nerpil som gör att man enkelt klickar sig vidare i lemmalistan.

En ren bugg är att programmet för den engelska ordboken inte hänger med i rörelse mellan lodrät och horisontell förflyttning. Om man har haft sin iPad i stående läge och slår upp ett uppslagsord med lång artikeltext, avbryts texten i vänstermarginalen när man vrider plattan till liggande läge i stället för att ”brytas om” och passa det liggande formatet. Detta är dock inget problem i iPhone eftersom appen bara kan orienteras i stående läge.

3.2. Norstedts svenska fickordbok, version 1.0

Den svenska fickordboken innehåller 18 000 ord och fraser och är tänkt att ge ett svenskt basordförråd (Apple 2011d). Definitionerna är av Cobuild-typ och består alltså av hela meningar. Man kan välja mellan avdelningarna ”Ordbok”, ”Quiz”, ”Norstedts” och ”Mer”. Under ”Mer” finns bland annat information om ordboken, ett kortare avsnitt om svensk grammatik, svenska helger, svenska landskap och en digital karta. Dessutom finns en tabell över världens länder med uppgifter hämtade från Utrikesdepartementet.

Norstedts svenska fickordbok ser i appversionen ut ungefär som *Engelsk ordbok Pro* sett till gränssnitt och sökmöjligheter. När man slår upp ett ord rullas själva uppslagsorden fram och sedan får man klicka på det ord man vill läsa mer om (se figur 2). Själva laddningen går snabbt. Om man sedan går vidare i lemmalistan genom att svepa med fingret över skärmen tills man kommer till ytterligare ett ord man vill slå upp och sedan klickar på ”Tillbaka” kommer man tillbaka till ordet man skrev in först, i stället för det senast markerade ordet. Det borde vara självklart att ”Tillbaka” betyder att man går tillbaka till den senaste delen i lemmalistan där man klickade och inte till en ännu tidigare uppslagning.



Figur 2: Skärmdump från *Norstedts svenska fickordbok*. Man söker genom att skriva in sökordet i sökrutan och sedan markera det i listan, vilket visas i den vänstra bilden. Den högra bilden visar sökresultatet för uppslagsordet *cello*.

Den svenska fickordboken har i iPhone också en knapp för ”Avbryt” – det förefaller onödigt eftersom det ändå finns ett kryss man kan klicka på som raderar allt man har skrivit om man råkar skriva fel. Det utrymmet skulle hellre få lämnas tomt eller länka till information om ordboken eller liknande.

En bugg i uttalsangivelsen är dock att det fonetiska tecknet för vokalen *e* har förvandlats till en fyrkant, vilket också syns i figur 2. (Detta är dock inte fallet med den tvåspråkiga ordboken eftersom den inte har uttal angivet på de svenska orden.)

Allt som allt är den svenska fickordboken enkel och lättanvänd och har ett trevligt utseende. Det som kan diskuteras är huruvida lemmauppsättningen utgör ett svenskt basordförråd eller ej, men det är egentligen en annan fråga. Däremot skulle förlaget tydligare kunna klargöra ordboksappens målgrupp, eftersom de användare

som har kommenterat den är mycket negativa till just lemmauppsättningen och definitionerna (se nedan).

Där den engelska ordboken har storleksanpassats till den större skärmen i iPad, har man med den svenska fickordboken bara möjlighet att se innehållet i 2 x standardstorleken. När man förstorar till dubbla storleken blir texten pixlig (dvs. inte anpassad för skärmupplösningen, vilket gör att pixlarna syns tydligt). Skillnaden mot utseendet i iPhone är enorm. I iPhone kan man inte förstora texten, men skärpan är av mycket god kvalitet vilket gör att texten verkligen blir knivskarp.

3.3. Kär & Galen

Appen *Kär & Galen* släpptes till Alla hjärtans dag 2011. Den är rosa med uppslagsorden i svart och pilar i form av röda hjärtan som ligger ner. När man startar appen står det ”Laddar fina ord ...”, en liten detalj som harmonierar väl med temat. Lemmalistan utgörs av ca 150 ord på svenska som har översättningar till sju språk: latin, engelska, tyska, franska, spanska, italienska och finska. Enligt förlaget är appens tänkta målgrupp de som vill sätta ord på sina känslor, och med den här appen får åtta språk att göra det på (Apple 2011e).

Man kan inte själv skriva in något sökord eftersom det inte finns någon sökruta, utan man skrollar sig fram i lemmalistan och väljer bland de ord som finns där. Varje artikel illustreras av en liten bild längst ner i högra hörnet (se figur 3). Den lilla illustrationen slumpas fram mellan, vad det verkar, knappt ett tiotal bilder.

Som idé är den här appen oemotståndlig – vem har inte haft *syd-änsuru* eller känt ett styng av *gelosia* någon gång (’hjärtesorg’ på finska, ’svartsjuka’ på italienska)? Lemmalistan bjuder på både det ena och det andra. Givna ord som har med kärlek att göra, som *brudpar*, *drömprins* och *kärleksförklaring* samsas med något överraskande ord som *separera*, *toffelhjälte* och *vänsterprassel*. Vidare

finns ett antal fraser eller flerordslemman som *får jag lov?*, *vara gubbsjuk* och *venerisk sjukdom*.



Figur 3: Skärmdump från *En liten ordbok för dig som är Kär & Galen*. Den vänstra bilden föreställer en del av lemmalistan och den högra bilden artikeltexten för uppslagsordet *bitterljuv*.

Den här appen skiljer sig från Norstedts övriga ordboksappar genom att den bara har målspråksordet som informationskategori. Det ges alltså ingen annan information såsom böjning, uttal eller liknande. Appen är en digital fortsättning på Norstedts tidigare utgivna onomasiologiska pappersordböcker *En liten ordbok för dig som är glad/arg/hungrig eller törstig*.

Kär & Galen är en rolig och mycket lättanvänd app och eftersom den är gratis är det svårt att ha några som helst invändningar mot den. Det troliga är också att den är gratis för att locka kunder att sedan också köpa en av betalapparna från Norstedts.

En utveckling av konceptet skulle kunna vara att utnyttja det elektroniska formatet till att ange uttal för orden eftersom det

handlar om så pass många olika språk som det svenska ordet översätts till.

3.4. Användarkommentarer

En intressant aspekt av användarfeedback i elektroniska sammanhang är möjligheten för användarna att betygsätta och kommentera de produkter de använder. Appar betygsätts också och det vanliga är att man kan ge ett antal stjärnor i betyg beroende på vad man tycker om appen. Till sitt betyg fogar också många användare en kort kommentar. Betyg och kommentarer finns att läsa på appens produktinformationssida i iTunes och i App Store. Där kan man också se ungefärligt antal nerladdningar av appen. De kommentarer som finns till de undersökta apparna ovan är av skiftande karaktär, allt ifrån mycket negativa till mycket positiva (här bör tilläggas att antalet betyg och kommentarer fortfarande är relativt få för alla de undersökta apparna). En person som kommenterar *Engelsk ordbok Pro* nämner bl.a. priset som en negativ faktor och efterlyser funktioner som sökhistorik och favoritmarkering. Kommentarererna till fickordboken är mycket hårdare – lemmauppsättningen diskuteras ("var är alla ord?") och priset är för högt, tycker den användare som betalade 55 SEK, medan användaren som köpte den till bokreapriset 7 SEK är nöjdare. De tre kommentarer som finns till *Kär & Galen* är genomgående positiva. Och det är troligtvis lättare att vara positiv till en trevlig gratisapp med bara "bonusfunktioner" än till en dyr app som man har högre förväntningar på.

Genomgående för de kommentarer som finns är att en app helst inte får kosta för mycket. Priset 349 SEK kommenteras negativt och det är, i jämförelse med andra ordböcker av ungefär samma omfattning, helt klart mycket högt.

Inför kommande uppdateringar av funktionerna tycker jag att apputvecklarna och förlaget ska ta en titt på användarkommenta-

rerna för att apparna dels ska kunna få fler eftersökta funktioner, dels marknadsföras bättre genom att målgruppen tydliggörs.

3.5. Anpassning till formatet och önskade funktioner för framtiden

Alla testade appar tar fram de eftersökta orden snabbt, och det är bland den viktigaste funktionen, som jag ser det. De större ordboksapparna har en sökruta där man kan skriva in det ord man vill slå upp samtidigt som lemmalistan rullas fram vartefter man skriver in ordet. Dessa två möjligheter är helt självklara och en service till användaren som på så sätt får hjälp i sin sökning. Man behöver helt enkelt inte skriva in hela ordet utan kan skriva halva och sedan klicka på det ord man söker tack vare prediktionslistan.

En funktion som inte finns i någon av apparna är möjlighet till stavningshjälp. Så länge man stavar rätt rullas det eftersökta ordet fram i listan, men om man stavar fel och klickar på ”Sök” händer ingenting, inte heller om man söker efter ett ord som inte finns. Lemmalistan blir helt enkelt ”bara stående” utan att användaren får någon indikation på felstavning eller att ordet saknas i ordboken. Här ser jag en stor förbättringspotential.

Ytterligare en detalj som gärna får inkluderas i kommande upplagor är det upplästa uttalet. Eftersom smarta telefoner, surfplattor och mediaspelare är gjorda för film och ljud borde denna funktion kunna ha en given plats.

En bra funktion skulle också vara länkning inom ordboken, t.ex. för att användaren snabbt ska kunna slå upp ett ord som används i definitionen, men det är i nuläget inte heller möjligt (denna funktion finns hos exempelvis appen *Dansk Ordbog* för Android). Önskelistan på funktioner som är anpassade till formatet är därmed relativt lång.

4. Sammanfattning och diskussion

De svenska ordboksapparna utgör ett välkommet nytillskott på den digitala ordboksmarknaden. De appar som har testats inför denna recension får sägas utgöra första generationens svenskproducerade appar med bara de mest basala funktionerna. Det går bra att söka på lemmat, men inte t.ex. på delar av ord, med hjälp av wildcards eller med stavningshjälp. Själva priset avskräcker antagligen många potentiella användare – för 349 SEK, som Norstedts dyraste app kostar, får man numera ganska avancerade spel och program, så att lägga den summan på en relativt enkelt utformad app tror jag knappast många kommer att göra på sikt. Gyldendals och Kunnskapsforlagets ordboksappar, som är jämförbara med Norstedts, har, som nämnts tidigare, lägre priser.

Jag efterlyser också utveckling för både Apples och Androids (m.fl.) produkter samtidigt, även om jag förstår att många apputvecklare lanserar sina produkter via Apple först. Androids marknadsandelar ökar dock snabbt och det bör man ta hänsyn till när man utvecklar apparna.

Sedan kan man diskutera hur mycket man använder sin app när man väl har installerat den. Personligen använder jag mina ordboksappar ytterst sällan och använder mycket hellre telefonens webbläsare eller en tryckt ordbok om jag inte har datortillgång, men skulle troligtvis använda ordboksapparna mer om jag befann mig på resa utanför Sverige. Frågan är också om man vill att kunden ska nöja sig med att köpa appen eller om man vill att kunden faktiskt ska använda den också. Beträffande apparna finns det alltså goda möjligheter för apputvecklarna och lexikograferna att göra användarstudier så att det blir tydligt vad mobilanvändarna är ute efter när de söker en ordboksapp och hur de sedan använder appen när de väl har installerat den.

För lexikograferna och systemerarna ställer apparna nya krav.

Tidigare har det räckt att anpassa ordböckerna enbart till det tryckta formatet. Sedan blev man tvungen att även ha nätpublicering i åtanke. Nu har det alltså kommit ytterligare ett format att anpassa sig till, och jag tror att det kommer att ta några år innan samspelet mellan lexikografer, systemutvecklare, apputvecklare och användare har stabiliserat sig och apparna optimerats.

Referenser

- Apple 2011a: *Gyldendals Engelsk Ordbog – Large*. <http://itunes.apple.com/dk/app/gyldendals-engelsk-ordbog/id338831838?mt=8> (juni 2011).
- Apple 2011b: *Kunnskapsforlagets Engelsk Blå Ordbok*. <http://itunes.apple.com/sv/app/kunnskapsforlagets-engelsk/id388381155?mt=8> (juni 2011).
- Apple 2011c: *Norstedts engelska ordbok Pro*. <http://itunes.apple.com/se/app/norstedts-engelska-ordbok/id382586485?mt=8> (juni 2011).
- Apple 2011d: *Norstedts svenska fickordbok*. <http://itunes.apple.com/se/app/norstedts-svenska-fickordbok/id400444149?mt=8> (juni 2011).
- Apple 2011e: *Kär & Galen*. <http://itunes.apple.com/se/app/kar-galen/id416998560?mt=8#> (juni 2011).
- Boychuck, Ben 2008: In battle of the Big Three dictionaries, American Heritage stands out on iPhone. I: <http://www.macworld.com/appguide/article.html?article=137137> (april 2011)
- Språkrådet 2010: *Språkrådets nyordslista 2010*. <http://www.sprakradet.se/9287> (april 2011)
- Svensk ordbok utgiven av Svenska Akademien* 2009. Stockholm: Norstedts i distribution.
- Trap-Jensen, Lars 2010: One, Two, Many. Customization and User Profiles in Internet Dictionaries. I: Anne Dykstra & Tanneke

ANMELDELSER

Schoonheim (eds.): *Proceedings of the XIV Euralex International Congress*. Ljouwert: Fryske Akademy, 1133–1143.

Wikipedia 2011. List of digital distribution platforms for mobile devices. http://en.wikipedia.org/wiki/List_of_digital_distribution_platforms_for_mobile_devices (april 2011).

Louise Holmer
doktorand och biträdande forskare vid Lexikaliska institutet
Institutionen för svenska språket
Göteborgs universitet
Box 200
S-405 30 Göteborg
louise.holmer@svenska.gu.se

En modern ordbok för alla som läser och forskar i Kalevala¹

Kalevi Koukkunen

Raimo Jussila: *Kalevalan sanakirja* [Ordbok över Kalevala]. Helsinki/Keuruu: Otava 2009. 520 s. Pris 40,30 EUR.

1. Bakgrund – om eposet Kalevala

Kalevala är det finska nationaleposet på samma vis som *Eddan* är det nordiska. Den kände folkdiktsforskaren och -utgivaren Elias Lönnrot (1802–1884) samlade in dikter bland folkets breda lager och sammanställde ett enhetligt verk. Kalevala representerar visserligen folkdiktning men till tema och komposition är verket också skönlitterärt. Lönnrot publicerade två versioner: Gamla Kalevala kom ut 1835–1836 och Nya Kalevala 1849.

I anslutning till Kalevaladagen år 2009² publicerades en ny ordbok över Kalevala, redigerad av lexikografen Raimo Jussila. Titeln är kort och gott *Kalevalan sanakirja* ('Ordbok över Kalevala', här efter JKS).

I Finland har årskullar av skolelever bekantat sig med Nya Kalevala och kunnat konstatera att verket vimlar av ord med dunkel betydelse. Många kommer länge efter skoltiden ihåg citat och till och med hela strofer. Kalevala är visserligen ett finskt verk men det in-

1 Denna recension är en omarbetning av ett bidrag som ingått i tidskriften *Bibliophilos* 3/2009 s. 53–57 med titeln *Loistava analyysi kansalliseepoksemme sanastosta* 'En förstklassig analys av ordförrådet i vårt nationalepos'.

2 Kalevaladagen och den finska kulturens dag firas till åminnelse av Kalevala; förordet till Gamla Kalevala har dateringen 28 februari 1835.

nehåller en så stor mängd ord som är främmande för dagens finnar att det har översatts till nufinska (Aulis Rintala 2006, Risto Puttonen 2008). Kalevala eller delar av verket finns översatt till 60 språk³.

Dagens finländare kan utan större besvär förstå texter från 1840-talet, men Kalevala innehåller en mängd dialektala och kvadialektala drag och därtill poetiska tillfällighetsord som förmodligen var lika svårförståeliga för folk på 1800-talet som i våra dagar.

Finska Litteratursällskapet har ett informationspaket om Kalevala på sin webbplats (<http://www.finlit.fi/kalevala>) där varje kapitel får en egen beskrivning (också på svenska).

2. Vem riktar sig en Kalevalaordbok till? – JKS och föregångarna

JKS är inte den första Kalevalaordbok som kommer ut. År 1949 publicerades *Kalevalan sanakirja* av Aimo Turunen (i fortsättningen TKS) och den har använts i decennier av studenter och forskare. Turunen kompletterade senare ordboksmaterialet, och år 1979 utkom det omfattande verket *Kalevalan sanat ja niiden taustat* 'Orden i Kalevala och deras bakgrund' (härefter benämnt TKST). Nya utgåvor av Kalevalaordböcker har alltså kommit vart 30:e år (1949, 1979 och 2009).

Turunen uppger att samtliga de 7 830 olika ord (lemman) som ingår i Kalevala presenteras i TKST, medan Jussila anger antalet ord till 8 447. Skillnaden i antal beror främst på ordböckernas makrostruktur. TKST anför vissa typer av ord som sublemman och dessa inbegrips inte i det angivna antalet. I artikeln *painaa*

3 Se <http://www.finlit.fi/kalevala>. Svenska översättningar av Kalevala har publicerats 1841, 1864–1868, 1944, 1948 och 1999, en dansk översättning kom ut 1908, en isländsk 1957 och en norsk 1967. För den som inte orkar läsa hela Kalevala finns Anders Larssons *Kalevala för lata* (i sin tur översatt till finska 2010: *Laiskojen Kalevala*).

'trycka'⁴ behandlas exempelvis nomen agentis -avledningen *painaja* 'tryckare, den som trycker' som sublemma liksom olika slag av verbavledningar: *painella*, *painaltaa* och *painatella*. I JKS anförs alla uppslagsord som huvudlemman. I Förordet anger Jussila att det totala antalet löpande ord i Kalevala är 67 210.

Turunens ordböcker är språkvetenskapliga. Även om formuleringen *ja niiden taustat* 'och deras bakgrund' kom in i titeln först 1979 finns det etymologiska uppgifter redan i TSK från 1949. I JKS förekommer inga etymologiska förklaringar utan det är betydelsebeskrivningen som står i fokus.

I fråga om typ kan Kalevalaordböckerna räknas till flera kategorier. De är ordlistor och författarordböcker, och till en del kan de också anses vara dialektordböcker och givetvis ordböcker över en äldre språkform. Jussila själv definierar sin ordbok som tolkordbok och citatordbok och som deskriptiv ordbok (JKS s. 5). Ingen av Kalevalaordböckerna är konkordanser utan förklarande ordböcker som ska hjälpa användarna att förstå Kalevalatexter.

Alla tre är enspråkiga finska ordböcker. Folkdiktningen representerar en så pass speciell språkform att en folklorist måste kunna finska för att kunna bedriva forskning inom området. Det hade inneburit ett stort arbete att lägga in till exempel engelsk- eller tyskspråkiga motsvarigheter i betydelseangivelserna.

De tilltänkta målgrupperna för ordböckerna vittnar om att tiderna förändras. Aimo Turunen klagade över att det trots att vetenskapen breddat och fördjupat kunskaperna är tydligt att "det moderna teknologiska tidevarvets människor" (detta var år 1940) står främmande inför den värld som beskrivs i Kalevala. Nuvarande och kommande generationer behövde en bro till nationalepots rika andliga tradition (TKS, s. I). Målgruppen specificeras på omslaget: de som idogt arbetar för den nationella kulturen, de som

4 Ordöversättningarna i denna recension har tillhandahållits av artikel-författaren själv. De är hämtade ur *Suomi–ruotsi-suusanakirja* (2004) och Lönnrots *Suomalais-Ruotsalainen Sanakirja* (1866–1880), och därtill har Huldéns *Kalevala*-översättning kontrollerats.

intresserar sig för området och i synnerhet modersmållärarna.

Raimo Jussila ser behovet av en ny Kalevalaordbok ur en europeisk synvinkel. Hans syfte har varit att överbrygga gapet mellan språket i Kalevala och i det moderna Finland. Ordboken är tänkt som ett hjälpmedel för vanliga finländare som bläddrar och läser i Kalevala (JKS s. 5). Till detta vill recensenten gärna lägga en förhoppning om att ordlistan ska bidra till att författare, teaterfolk, bildkonstnärer och kompositörer via ordboken låter sig inspireras av Kalevala i sitt skapande.

I JKS ansätts samtliga ord ur Kalevala som uppslagsord. Det är alltså inte bara för moderna läsare obekanta ord som behandlas. Jussila har således språkvetenskapligt analyserat också ord som *ei*, *ja*, *kun*, *niin*, *oi* och *vaan* ('inte', 'och', 'då', 'så', 'oj' och 'utan'). Basen för ordboken har Jussila sammanställt manuellt genom att lemmatisera varje ord och skapa en konkordans i form av en Excel-tabell med kapitel- och versadsnummer för varje förekomst (JKS s. 8). Manuell konkordanssammanställning kan komma till heders också i datorns tidevarv.

3. Artikelutformning och frekvensanalys

Ordboksartiklarna i JKS är strukturellt åskådliga. Direkt efter varje uppslagsord nämns antalet förekomster i Kalevala inom parentes, t.ex. *lyödä* (41) ['slå']. För exempelvis verbet *lyödä* anges nio numrerade betydelser och varje betydelse belyses med ett eller flera exempel ur verket (vilket alltså gör att ordboken får drag av en citatordbok). Det första betydelsemomentet i artikeln *lyödä* innehåller tre exempel med angivelse om nummer på kapitel och versad för citatet: *elä anna anopin lyödä eläkä apen torua* (24:201–202) [i Huldéns översättning *inga slag från hennes svärmor, inget grumsande av svärfar*]. Efter betydelsemomenten följer – med fortsatt löpande numrering – grupper innehållande uttryck och vänd-

ningar. I artikel *lyödä* utgör dessa momenten 10–15 med bl.a. uttrycket *lyödä löylyä* 'hälla vatten på stenarna i bastuugnen'. Jussila har alltså stannat för att ha grundord som lemman och har inte fört upp fraserna som egna uppslagsord på vederbörlig plats i alfabetet.

I slutet av många artiklar återfinns en lista på sammansättningar eller därmed jämförbara ord. Ordet *leipä* 'bröd' förekommer 33 gånger i Kalevala, och sist i *leipä*-artikeln finns sammansättningarna *mesileipä* 'honingskaka', *olkileipä* 'halmbröd' och *vehnäleipä* 'vetebröd'. Därtill finns det hänvisningar till *petäjäinen* 'barkbröd' och *silkko* 'rent barkbröd'. I artikeln *karhu* 'björn' finns det tio-tals hänvisningar till sammanhörande ord, dvs. till omskrivningar som i detta fall utgörs av hypokorismer, t.ex. *aika poika* 'vuxen pojke'⁵, *karvalallinen* 'hårtotte, luddelulla' och till och med *vanha mies* 'gammal man, gammelfar'. I artikeln *puu* 'träd' finns det 31 sammansättningsreferenser, från *emäpuu*⁶ 'stomme' till *viipsinpuu* 'härvel'. Dessa ord återfinns också som egna lemman, men sammansättningslistan ger ordboksanvändaren en god bild av ordfamiljen. Homonymer separeras med index: *kuu*¹ 'måne', *kuu*² 'ister'.

JKS anger antalet förekomster i Kalevala för varje ord, vilket kan ge intressanta inblickar i Kalevalavokabulären. Belysande är exempelvis en jämförelse av släktskapstermer (tabell 1):

<i>poika</i> 'pojke; son'	316	<i>taatto</i> 'far'	30
<i>emo</i> 'mor'	224	<i>poiut</i> 'liten pojke'	27
<i>iso</i> 'far'	107	<i>maammo</i> 'mor'	25
<i>tyttö</i> 'flicka'	71	<i>tytti</i> 'flicka'	21
<i>tytär</i> 'dotter'	61	<i>isä</i> 'far'	8
<i>äiti</i> 'mor'	57		

Tabell 1: Frekvenser för släktskapsord.

5 I Huldéns översättning (46:273): *karlakarli, berömda hjälte*.

6 Den gängse betydelsen hos *emäpuu* är 'kölstock' men i Kalevala avser ordet den stomme som strängarna fästs i på en kantele.

Kan man utifrån frekvenserna dra slutsatsen att Kalevala mer är sönernas och mödrarnas epos än flickornas och fädernas (*poiuta* 'son' totalt 343 förekomster och *emo* + *äiti* + *maammo* 'mor' totalt 306 jämfört med *tyttö* + *tytär* + *tytti* 'dotter' totalt 153 och *iso* + *taatto* + *isä* 145). En viss antydning ger siffrorna, men påpekas bör att frekvenserna gäller form och inte betydelse. Av de 224 förekomsterna av *emo* avser en del djurhonor och en del urmödrar. Å andra sidan uppträder *mies* 'man' 325 gånger (därtill i tolv olika sammansättningar från *ampumies* 'skytt; trollkarl' till *tietomies* 'trollkunnig man, kännare'), *uros* 'hjärte' 143 gånger, *nainen* 'kvinna' 97 gånger och *vaimo* 'hustru' 75 gånger. Förekomstssiffrorna för personliga pronomen är: *minä* 'jag' 269, *sinä* 'du' 148, *hän* 'han; hon' 39, *me* 'vi' 54, *te* 'ni' 8 och *he* 'de' 3. Det stora antalet *jag*-belägg beror på att Kalevala-epoet är berättarcentrerat. I slutet av ordboken finns en tabell över ordförekomsterna, och det kan noteras att det finns nästan 4 000 ord som förekommer en enda gång (JKS s. 519: Antalet sällan förekommande ord i Kalevala).

Bäst framgår det vilket gediget lexikografiskt arbete JKS är om man jämför de tre Kalevalaordböckerna. Uppslagsordet *oma* 'egen' saknas helt och hållet i TKS, och i TKST finns bara en kortfattad betydelseangivelse *jnk omistama, jllek kuuluva* 'som ägs av ngn, som hör till ngn'. I JKS finns hela fjorton betydelser angivna. Huvudbetydelse är numrerade och därtill förekommer betydelsen angivna med bokstäver i flera av momenten. Artikeln *ruveta* med grundbetydelsen 'börja' har följande uppställning (tabell 2):

1. toiminnan aloittamisesta, ryhtymisestä	1. om att inleda en verksamhet
a. ”ruveta + 3. infinitiivin illatiivi”	a. ruveta + infinitiv 3 illativ [<i>ruveta tekemään</i> = börja göra]
b. ”ruveta jollekin”	b. ’börja med ngt’ [<i>ruveta juttusille</i> = börja samspråka]
c. ”ruveta joksikin, jonkinlaiseksi”	c. ’bli något’ [<i>ruveta opettajaksi</i> = bli lärare] [<i>ruveta hankalaksi</i> = bli besvärlig]
d. ”ruveta johonkin” = tulla joksikin, jonkinlaiseksi	d. ”ruveta johonkin” = ’bli något’ [sällsynt betydelse: Taivas reikihin rupesi (Huldén: det blev hål på hål i himlen)] [<i>taivas</i> (NOM) ’himmel’, <i>reikihin</i> (ILL.PLUR) ’hål’]
2. koskea, kajota	2. ’beröra, vidröra’
3. tarttua, liittyä kiinni	3. ’fastna’

Tabell 2: Betydelseangivelserna i artikeln *ruveta* ’börja’.

Exemplet visar hur ingående betydelsebeskrivningen är i JKS. I sin närläsning har Jussila verkligen funnit Kalevalaordförrådets rikedom och han presenterar det på ett mycket lockande sätt för moderna finländare. När man läser Kalevala måste man slå upp ord i en ordbok, men JKS fungerar också tvärtom: genom att läsa i JKS får man en god uppfattning om Kalevalas vokabulär. Betydelseanalyserna åskådliggör de finaste nyanser; man inser hur semantiskt skickliga Kalevaladiktarna har varit. Av tabell 3 nedan framgår vilka ord som har de utförligaste betydelsebeskrivningar-

na i ordboken. För antalet betydelseangivelser har nummer- och bokstavsangivna betydelser räknats samman.

<i>saada</i> 'få' 29	<i>kulta</i> 'guld' 20	<i>tulla</i> 'komma' 16	<i>oma</i> 'egen' 14
<i>pitää</i> 'hålla; måste' 26	<i>suuri</i> 'stor' 20	<i>lyödä</i> 'slå' 15	<i>Manala</i> 'underjorden' 13
<i>ottaa</i> 'ta' 25	<i>toinen</i> 'andra; annan' 20	<i>päivä</i> 'dag' 15	<i>mieli</i> 'sinne' 13
<i>panna</i> 'sätta' 25	<i>maa</i> 'jord' 19	<i>tehdä</i> 'göra' 15	<i>sija</i> 'ställe' 13
<i>se</i> 'den, det' 24	<i>tie</i> 'väg' 18	<i>alla</i> 'nedanför; under' 14	<i>sormi</i> 'finger' 13
<i>käydä</i> 'gå' 23	<i>tuli</i> 'eld' 18	<i>antaa</i> 'ge' 14	<i>taivas</i> 13
<i>laskea</i> 'lägga ned' 23	<i>ilma</i> 'luft' 17	<i>laulaa</i> 'sjunga' 14	<i>vetää</i> 'dra' 13
<i>paha</i> 'ond' 23	<i>pää</i> 'huvud' 17	<i>lintu</i> 'fågel' 14	<i>yksi</i> 'en; ett' 13
<i>vesi</i> 'vatten' 21	<i>suu</i> 'mun' 17	<i>lähteä</i> 'gå' 14	
<i>aika</i> 'tid' 20	<i>käsi</i> 'hand' 16	<i>nousta</i> 'stiga upp' 14	

Tabell 3: Artiklar med stort antal betydelseangivelser.

Av de 38 ord som har flest betydelser angivna är 14 (nästan 37%) verb. Verben utgör ju kärnan i en sats. Syntagmatisk information⁷ ges via substantiv, och antalet substantiv på listan är 17 (nästan 45%). Att verb är polysema är förståeligt, men att exempelvis ordet *guld* har 20 betydelser förvånar. Angivna betydelser är t.ex. 'älskling', 'far', 'bror', 'vän', 'björn', 'häst', 'gök', 'ekorre', 'måne', 'sol' och 'kantele'. Bland de högfrekventa substantiven noterar man många kroppsdelbenämningar: hand, huvud, finger och mun.

⁷ Se NLO s.v. *syntagmatisk information*.

4. Önskemål visavi användarvägledningen

JKS innehåller en mycket kort användarvägledning omfattande en sida. Då ordboksförfattaren i förordet ger en intressant redogörelse för Kalevalaeposets uppkomst och struktur (JKS s. 5–8), skulle jag gärna ha sett att han också hade beskrivit sina lexikografiska mål och metoder närmare. Å andra sidan ”talar verket för sig själv” som man brukar säga. Inledningen kunde också ha innehållit bl.a. ett omnämnande om att skildringarna i Kalevala ofta är hyperboliska. Ett exempel på det kunde ha varit dikten om Väinämöinsens och Joukahainens sångarstrid (3:295–300):

*Lauloi vanha Väinämöinen:
järvet läikkyi, maa järisi,
vuoret vaskiset vapisi,
paaet vahvat paukahteli,
kalliot kaheksi lenti,
kivet rannoilla rakoili.*

I Huldéns översättning:

Gamle Väinö, vad han kunde!
Sjöar sjöd och jorden skälvde,
kopparbergen gick i vågor,
fasta hållar for i stycken,
klippor brast med brak i bitar,
stenar sprack på alla stränder.

I en allmän presentation av Kalevala kunde också de finurliga rimmen ha nämnts: *ihveniä ahvenia, tuimienia, taimenia*⁸ eller *Enkä lähe Inkerelle, penkerelle, pänkerelle*. Kalevala innehåller alltså också nonsensord som ger framställningen rytm och för den framåt.

8 I Huldéns fyndiga översättning (48:99): *abborrar och andra borrar, öringar och andra ringar*.

En del språkvetenskapliga arkaismer och särdrag kunde ha tagits upp. Kalevala innehåller t.ex. karitiva former som är främmande för nufinskt språkbruk: *armotuus* (nufinska *armottomuus*) 'obarmhärtighet', *kalatuus* (*kalattomuus*) 'avsaknad av fisk', *lemmetyys* (*lemmettömyys*) 'kärlekslöshet' m.fl. I artikeln *lapsepana* 'i barndomen, som yngre'⁹ [till *lapsi* 'barn' med komparativändelsen *-mpi* och essivändelsen *-na*] kunde möjligheten till adjektivisk böjning för vissa substantiv ha nämnts för moderna användare.

Kalevala innehåller ett flertal ordformer som anger att något går genom eller via något: *nenitse* 'via näsan', *tanteritse* 'landvägen', *teitse* 'längs (lands)vägar'. Prolativformen kan fortfarande marginellt vara produktiv; den bäst kända moderna bildningen är *sähköpostitse* 'via e-post'.

5. Slutord

Efter ett par års flitigt användande av JKS kan recensenten konstatera att ordboken är ett högklassigt och innehållsrikt verk som ger inspiration att läsa Kalevala. Till dess främsta kvaliteter hör en enhetlig och logisk lexikografisk struktur. Ytterst förtjänstfull är också den semantiska analysen och tolkningen av ordförrådet i Kalevala: eposet får liv. Man kunde tillämpa reklamtexten för TKS och säga att JKS är "ett oundgängligt hjälpmedel för dem som arbetar med nationell kultur, för studenter och för lärare i finska".

Men ordboken i sig kan också vara en läsebok. Den guidar sina användare och tillhandahåller en mängd intressanta fakta om det finska språket och den finska folktraditionen och kulturen. JKS är resultatet av ett grundligt arbete och utgör därmed också en förträfflig grundlag för semantisk forskning kring Kalevala. Själv blev jag inspirerad att läsa de svenska översättningarna av eposet och att analysera det ur translatalogisk synvinkel.

9 I Huldéns översättning (21:332): *när jag var yngre*.

Litteratur

- Kalevala – Finlands nationalepos*. Översättning Lars Huldén och Mats Huldén. Fjärde upplagan. Atlantis: Stockholm 2008.
- Larsson, Anders 2000: *Kalevala för lata*. Alfabet: Stockholm.
- Larsson, Anders 2010: *Laiskojen Kalevala* (alkuteos: *Kalevala för lata*). Suomentanut Liisa Ryömä. Schildts: Helsinki.
- Lönnrot, Elias 1958: *Suomalais-Ruotsalainen Sanakirja – Finskt-Svenskt Lexikon*. Osat 1 (A–M) ja 2 (N–Ö). Ilmestynyt alun perin Suomalaisen Kirjallisuuden Seuran kustannuksella 1866–1880. Käytetty editio: Kolmas, manul-menetelmällä jäljennetty painos. Werner Söderström Osakeyhtiö: Porvoo/Helsinki.
- NLO = *Nordisk leksikografisk ordbok*. Toimittaneet Henning Bergholtz, Ilse Cantell, Ruth Vatvedt Fjeld, Dag Gundersen, Jón Hilmar Jónsson & Bo Svensén. Skrifter utgitt av Nordisk forening for leksikografi. Skrift nr. 4. Universitetsforlaget AS: Oslo 1997.
- Pottonen, Risto 2008: *Kalevala suomeksi*. Books on Demand GmbH: Helsinki/Norderstedt.
- Rintala, Aulis 2006: *Kalevala nykysuomeksi*. Pilot-kustannus Oy: Tampere.
- Suomi–ruotsi-suursanakirja – Stora finsk-svenska ordboken*. Päätöimittaja – huvudredaktör Birgitta Romppanen, toimitussihteeri – redaktionssekreterare Nina Martola, toimittajat – redaktörer Ilse Cantell, Mats-Peter Sundström. 3., tarkistettu ja päivitetty painos, 2. laitos. Werner Söderström Osakeyhtiö: Helsinki/Juva 2004.
- TKS = Turunen, Aimo 1949: *Kalevalan sanakirja*. Suomalaisen Kirjallisuuden Seuran toimituksia 235. Suomalaisen Kirjallisuuden Seura: Helsinki.
- TKST = Turunen, Aimo 1981: *Kalevalan sanat ja niiden taustat*. 2. painos. Karjalaisen Kulttuurin Edistämissäätiö: [Joensuu/Lappeenranta].

ANMELDELSER

Kalevi Koukkunen
filosofie licentiat
Sockkärrsvägen 5 B 33
FI-02760 Esbo
kalevi.koukkunen@gmail.com

Översättning från finska: Nina Martola

En formidabel metaordbok

Sven-Göran Malmgren

Loránd-Levente Pálfi: *Leksikon over ordbøger og leksika*. Under medverkan av Johnny Finnsson Lindholm. 2. udgave. København: Frydenlund 2011. 640 sidor.

Låt det bli sagt från början: *Leksikon over ordbøger og leksika* (LOL) är ett beundransvärt verk med mycket få föregångare i Norden. Det förtecknar (med all säkerhet) en stor majoritet av alla danska ordböcker som utgivits mellan 1990 och 2010. Det har ett beskrivningsdjup som knappast tidigare uppnåtts i liknande lexikon i Norden. Pálfi nämner några danska föregångare, men som icke-dansk tänker man gärna på Eva Haugens *A Bibliography of Scandinavian Dictionaries* (Haugen 1984; i fortsättningen BSD), en på många sätt förnämlig bibliografi över ca 2.500 skandinaviska ordböcker från 1500-talet till ca 1980. I ett nordiskt forum som LexicoNordica kan det vara motiverat att inledningsvis kasta en blick på BSD, under viss jämförelse med LOL.

Värdet av Eva Haugens arbete ökas väsentligt av att en historisk översikt över nordisk lexikografi av maken Einar Haugen ingår. Samtidigt är verket behäftat med betydande brister; det är långtifrån fullständigt och innehåller många felaktigheter. Det är uppenbart att Eva Haugen inte har lyckats leva upp till bibliografins ”gyllene regel”, nämligen att bibliografen bör ha haft varje förtecknat verk i handen. En värdefull, men otruckyt komplettering till Haugens arbete är Hannesdóttir (1987). Här förtecknas (nästan) samtliga svenska ordböcker utgivna före 1850; många av dem är inte upptagna i BSD.

I rättvisans namn bör det sägas, dels att Pálfi förfogat över

hjälpmedel som inte var tillgängliga för Eva Haugen, dels att föremålen för hans bibliografi – ordböcker på ett och samma språk under en begränsad period i nutiden – är betydligt mer lätthanterliga och sammantaget mer överskådliga än materialet för Haugens bibliografi. Man studsar faktiskt en aning när Pálfi skriver att BSD kan användas som ett ”supplement” till LOL (s. 43). Mer adekvat – med tanke på att BSD, låt vara ofullständigt, täcker nära 500 år mot LOL:s 20 – skulle det väl vara att se LOL som ett supplement till BSD, vad danska ordböcker beträffar.

När det är sagt, bör man dock notera att LOL förtecknar fler danska ordböcker än BSD, trots att den täcker en period som bara är 1/25 så lång. Det är en fin illustration till klichén att våra moderna samhällen är informationssamhällen på ett helt annat sätt än tidigare samhällen. Och det kan tilläggas, att det stora antalet titlar i LOL i sig är ett bevis på den höga ordbokskulturen i nutidens Danmark. (Jfr Humbley 2010:241–242, som betvivlar att någon skulle ha åtagit sig uppgiften att sammanställa en bibliografi över de senaste 20 årens franska ordböcker.)

Kärnan i LOL liksom i BSD är en lista över ett par tusen ordböcker, tillsammans med vederbörliga bibliografiska data. Medan BSD har en tematisk disposition av bibliografin (allmänordböcker för sig och fackordböcker för sig, varje språk för sig, för varje språk enspråkiga ordböcker för sig, tvåspråkiga för sig etc.), har Pálfi valt en rent alfabetisk disposition. I LOL kompletteras därför bibliografin med ca 700 översiktsartiklar med uppslagsord som *engelsk* (om dansk-engelska och engelsk-danska ordböcker), *geologi* (om geologiska fackordböcker) och även mer udda uppslagsord som *okkultisme* och *tinsoldater*. Pálfis metod är onekligen platskrävande (minst 200 av knappt 500 sidor hade kunnat sparas med Haugens ”minimalistiska” lösning), men många av översiktsartiklarna är så värdefulla att de är väl värda det extra utrymme som krävs. I själva verket är det inte minst uppläggningsen med översiktsartiklar som komplement till bibliografin som gör LOL principiellt

överblicksartikeln BSD. Som exempel på en översiktsartikel återges här den relativt korta artikeln *cykling*; längre artiklar kan omfatta flera sidor:

cykling

Et forholdsvis omfattende og ikke uambitiøst værk er ■ *Cykelleksikon* (1994), som med sine 1.700 artikler (450 omhandler personer) omhandler alt vedrørende cykler, cykling, cyklens, cyklingens historie, rekorder, cykelsport, cykelryttere, mesterskaber, producenter, tekniske landvindinger, udstyr mv. både nationalt og internationalt. || I flersprogs-kategorien foreligger ■ *Ligfiet.net's Dictionary* (2009), som er en onlineordbog for termer med tilknytning til cykelsporten. Brugeren kan vælge mellem følgende sprog: dansk, engelsk, fransk, hollandsk, norsk, spansk, svensk, tysk. Søgemaskinen vil søge på det sprog, man har aktiveret. Resultatet gives i en tabel, der viser alle sprogene samtidigt, ordnet efter det målsprog, man har valgt. Man kan også vælge et bogstav og se alle ord under dette bogstav. Det kan også til-/fravælges, hvor mange sprog der skal vises. ►► sport

Som läsare får man här mycket mer information om de båda moderna danska lexikonerna om cykling än i bibliografin (och detta även om denna som i BSD skulle ha varit tematiskt disponerad). För pappersordboken får man exempelvis reda på antalet artiklar och vilka delområden av ämnesområdet cykling de behandlar. För nätordboken beskrivs olika sökrutiner och sökmöjligheter. Ett svagt positivt omdöme ("ikke uambitiøst") avges om pappersordboken. Om en ordbok har recenserats i en lexikografisk tidskrift, så anges det, och då får läsaren ofta en utförligare kvalitetsbedömning. Vi återkommer till kvalitetsomdömena lite senare.

Artiklarnas överskådlighet främjas av fyra olika strukturindi-

katorer, av vilka tre ses i artikeln *cykling*. Varje ny ordbokstitel signaleras med indikatorn ■, ett nytt avsnitt i artikeln, t.ex. om en ny ordbokstyp, med indikatorn ||, och hänvisningsformlerna ”se” och ”se även” symboliseras av tecknen ► resp. ►►.

Trots att texten i översiktsartiklarna med nödvändighet är komprimerad, gör den sällan ett katalogartat intryck. Ganska ofta är den faktiskt njutbar, särskilt i längre artiklar där Pálfi ibland frångår den bakre gränsen 1990 och t.ex. kan ge en minihistorik för en viss typ av ordbok. Sammantaget innehåller översiktsartiklarna oerhört mycket information, ovärderlig t.ex. för läsare som är i valet och kvalet om vilken ordbok de i första hand ska köpa eller beställa fram på biblioteket.

Också de rent bibliografiska uppgifterna i LOL är mycket mer ingående än BSD:s. De är också mer uttömmande, och enligt Pálfis egen uppgift, som det inte finns någon anledning att betvivla, inte sällan mer korrekta än de uppgifter som återfinns i den stora danska nät databasen www.bibliotek.dk (Pálfi ger flera exempel på rättelser, s. 27). En post i LOL:s bibliografi kan se ut så här:

1782. *Symboler og ideogrammer* / Carl G. Liungman / oversættelse: Hanne Steinicke / forlagsredaktion: Henrik Ægidius / 1. udgave, 1. oplag / København: Aschehoug 2003 [oversat efter Carl G. Liungmans *Ideogram och symboler : en uppslagsbog [sic!]*, 1999; værket er også blevet oversat til engelsk, og den engelske udgave foreligger tillige online på: <http://www.symbols.com>]

Här är LOL (jämfört med www.bibliotek.dk) ensam om uppgiften om förlagsredaktör liksom om uppgiften att lexikonet blivit översatt till engelska och att den engelska versionen finns på nätet.

Inte desto mindre har den danska nät databasen spelat en avgörande roll för Pálfis projekt. Det var den som möjliggjorde den första stora datafångsten, upprättandet av en preliminär grupp av

kandidater till lexikonbibliografin. Om detta kan man läsa i LOL:s ganska långa inledande text (s. 30ff.). En sökning gjordes på alla titlar i www.bibliotek.dk som innehöll vissa nyckelord som *ordbog*, *leksikon*, *encyklopædi*, *database*, *opslagsbog*, *grundordbog*, *miniordbog* etc., och som var utgivna 1990 eller senare. Därefter gjordes en bedömning av varje verks ”ordboksmässighet”; många kandidater befanns inte vara riktiga ordböcker och togs inte med i LOL. Den sålunda reducerade listan kompletterades ”med eget kendskab til det leksikografiske landskab i Danmark og ved mere eller mindre tilfældigt fremfundne dokumenter” – som alltså inte hade något av nyckelorden i sina titlar. Ett exempel är det just nämnda lexikonet *Symboler og ideogrammer*.

Här inser man att ett verk som LOL aldrig kan uppnå fullständighet, något som också framhålls i en efterskrift (s. 626). Även om de allra flesta språkordböcker innehåller nyckelord som *ordbog*, *leksikon* e.d. i titeln, är detta långtifrån alltid fallet med fackordböcker. Pálfi har en dokumenterat god överblick över det danska fackordboksbeståndet, bl.a. genom sin flitiga recensionsverksamhet, och han har uppenbart goda kontakter i förlags- och biblioteksvärlden. Troligen skulle ingen annan kunna kartlägga den danska ordboksutgivningen under den senaste 20-årsperioden på ett mer uttömmande sätt än han. Ändå hade man gärna sett att han hade diskuterat fullständighetsproblemet (eller med en målande svensk statistikterm: mörkertalet) lite mer ingående, och formuleringen ”tilfældigt fremfundne dokumenter” känns en smula bekymmersam.

Den inledande texten i LOL, som omfattar nära 50 sidor, innehåller också bl.a. en användarvägledning och en intressant teoretisk diskussion kring avgränsningsproblemen (vad är en ordbok, en encyklopedi etc.). Texten är förhållandevis oteknisk och bör kunna läsas med behållning av en bred grupp användare.

Efter översiktsartiklarna (s. 47–304) och huvudbibliografin (s. 305–532) följer en bibliografi över ca 250 relevanta artiklar från

lexikografiska tidskrifter, huvudsakligen recensioner av de aktuella ordböckerna. Därefter följer ett namnregister, en kort efterskrift och allra sist, något otippat, en artikel av Henrik Lorentzen om webbsajten ordnet.dk. Man får inte veta om artikeln är skriven speciellt för LOL (i så fall borde författarens namn kanske ha angetts på titelsidan), eller om den publicerats på annat håll.



Inte ens det mest imponerande lexikon är fullkomligt. Här följer till sist några tankar om hur LOL ev. kunde förbättras ytterligare. Huvudsakligen handlar det om småsaker.

Först en kommentar till bokens titel, som är överraskande allmän. Borde det inte åtminstone i en underrubrik anges, att det är *danska* ordböcker som bibliograferas?

Låt oss så se lite närmare på värdeomdömena som tillhandahålls om vissa ordböcker. De kan, som vi sett, vara svagt positiva, eller svagt negativa, men de kan också uttrycka översvallande beröm resp. hård kritik. Utan att direkt ange vilka ordböcker det gäller kan vi citera följande omdömen: ”en formidabel citatordbog” (s. 51), ”hører til det ypperligste i dansk fagleksikografis historie” (s. 80), ”en ganske glimrende onlineordbog” (s. 113), ”et flot og seriøst værk” (s. 141), ”En meget ambitiøs, flot ordbog” (s. 159), ”en ualmindeligt godt gennemtænkt og gennemarbejdet ordbog” (s. 184), ”et udmærket lille værk” (s. 233); resp. ”værket gør sig stadig skyldig i underlødigt stofbehandling” (s. 207), ”en meget lille, og i øvrigt primitivt udarbejdet ordliste” (s. 249), ”fremstår mere end underlødigt og derfor komplet ligegyldigt” (s. 276), ”Disse /.../ er dog ikke at regne for noget i forhold til de svenske pendants” (s. 285). Den stora majoriteten av ordböckerna registreras dock utan något värdeomdöme. Det är inget att säga om det; i de flesta fall är det svårt att hitta recensioner, och det vore för mycket begärt att Pálfi skulle bekanta sig närmare med

mer än 1000 ordböcker. Som läsare skulle man inte vilja vara utan de kvalitetsbedömningar som ändå ges och som gör ett tillförlitligt intryck, men det finns en liten risk att läsaren tolkar frånvaro av värdeomdöme ungefär som ”godkänd, men inte mer”, när det i själva verket kan vara fråga om allt från en ”glimrende” till en ”underlödlig” ordbok. Detta är dock mer en allmän kommentar än en egentligen kritisk anmärkning.

Så ett par kommentarer som relaterar sig till bibliografin över (huvudsakligen) recensionsartiklar. Det är utomordentligt att en sådan bibliografi har tagits med (något motsvarande finns inte i BSD), även om den i stort sett inte omfattar recensioner av fackordböcker i facktidsskrifter; att spåra sådana recensioner hade varit ett för stort arbete. Det förefaller dock som om många poster i bibliografin är onödigt utförliga. Det finns knappast någon anledning att upprepa alla uppgifter från ordboksbibliografin om de recenserade ordböckerna – exempelvis återkommer samma 12 rader med alla bibliografiska upplysningar om *Den Danske Ordbog* på tre ställen i recensionsbibliografin. I ordboksbibliografin uppträder *Den Danske Ordbog* som nummer 471; varför inte nöja sig med författare, ”rec. av 471” samt källa i recensionsbibliografin, så skulle den kunna förkortas med 50–75%? (Detta gäller självfallet inte för de recensionsartiklar som har egen rubrik; i sådana fall händer det å andra sidan att läsaren inte får veta vilken ordbok recensionen handlar om.)

Det är lätt att navigera mellan översiktsartiklarna och recensionsbibliografin, men däremot finns ingen direktförbindelse mellan ordboksbibliografin och recensionsbibliografin. Läsaren måste i stället gå omvägen över en lämplig översiktsartikel, som kanske inte alltid är självklar, eller i andra fall mycket lång. Låt oss säga att en läsare har slagit upp *Juridisk ordbog tysk-dansk* och vill veta om det finns en recension av den. Med lite förtrogenhet med ordboken kan han räkna ut att han ska slå upp översiktsartikeln *Jura*, som dock är fyra sidor lång och tar upp en mängd enspråkiga och

tvåspråkiga juridiska ordböcker. Till sist hittar han nog rätt referens, Simonnæs (1996), men visst hade det varit tidsbesparande om den upplysningen också hade funnits i ordboksbibliografin. I andra fall hade det varit mer eller mindre nödvändigt – vilken är den rätta översiktsartikeln till *Skandinavisk ordbok*?

Till sist ett par innehållsliga detaljer. I LOL räknas serbiska, kroatiska och bosniska som (varianter av) ett och samma språk, serbokroatiska, något som i varje fall varit kontroversiellt i Sverige sedan 1990-talet och med all säkerhet är i strid med den officiella ståndpunkten i de tre berörda länderna. Som bekant har serbiska och kroatiska olika skriftspråk, en uppgift som faktiskt inte ges i LOL. Och slutligen vill man som svensk hoppas och tro att Pálfi har fel när han i översiktsartikeln *skånsk* skriver att "[n]utidens skånske skolebørn skal /.../ fejre svenske konger, der var ansvarlige for masse mord på Skånes befolkning, mens de ikke lærer om Skåne-landenes danske historie og kultur." (Förmodligen besöker många skånska barn under sin skoltid det utomordentliga museet i Malmöhus slott där de kan få en icke-nationalistisk framställning av Malmö och Skånes historia.)



LOL är ett storverk – men ett storverk som snabbt åldras. I och med den föreliggande upplagan utkommer LOL enligt Pálfi "for anden og sidste gang" (s. 629). Om något eller några år måste LOL "suppleras", i praktiken förmodligen av nät databasen www.bibliotek.dk. I det läget kanske många nöjer sig med att konsultera databasen, trots de många tilläggsupplysningarna i LOL.

Ändå är det föreliggande resultatet värt den uppenbarligen mycket stora arbetsinsatsen. Med LOL har en svåröverträffad standard för framtida lexikonbibliografier etablerats. Lexikografer och andra har fått en samlad överblick över den danska ordboksutgivningen under två decennier. Och många av översiktsartiklarna har ett värde i sig som kommer att bestå länge.

Litteratur

- BSD = Haugen, Eva L. 1984: *A Bibliography of Scandinavian Dictionaries*. With an Introduction by Einar Haugen. White Plains, New York: Kraus International Publications.
- Hannesdóttir, Anna H. 1987: *Förteckning över äldre svenska ordböcker. Tiden fram till 1850*. Göteborgs universitet: Institutionen för nordiska språk. [Dupl.]
- Humbley, John 2010: [Recension av LOL] I: *Cahiers de Lexicologie* 97, 240–242.

Sven-Göran Malmgren
professor
Lexikaliska institutet
Institutionen för svenska språket
Göteborgs universitet
Box 200
SE-405 30 Göteborg
malmgren@svenska.gu.se

En etterlengtet ordbok

Ole Michael Selberg

M. J. Filipek, M. Iwanow, R. Iwanow: *Norsk-polsk polsk-norsk juridisk ordbok*. Oslo: Cappelen Akademisk Forlag 2010. 320 sider. Pris 398 NOK.

1. Innledning

I betraktning av det store antall polakker som nå oppholder seg i Norge – dels midlertidig, dels mer permanent – torde behovet for en ordbok som denne være innlysende. Målgruppen den henvender seg til, har hittil måttet hjelpe seg med det lille som fantes av leksikalske broer mellom de to språkene – det vil i praksis si FAGORDBOK 1994, som inneholder ca. 850 juridiske termer, og Soleng og Soleng (2008), som er en allmennspråklig ordbok med i alt ca. 60 000 oppslagsord.

Den nye ordboken omfatter ifølge forordet ca. 18 000 norske og polske lemmaer (opplagsord), og skal dekke «sentrale fagspråklige felt innenfor norsk juridisk terminologi», med hovedvekt på strafferett og privatrett. Målgruppen er først og fremst «jurister, forretningsmenn, studenter og andre som har behov for å overføre juridisk terminologi til polsk». Det er forfatterens håp at boken vil være «et nyttig hjelpemiddel for bedrifter, domstoler og andre statlige institusjoner».

2. Typografi

Bare to typer skrift er brukt – halvfet for kildespråket og normal for målspråket, som i «alderspensjon – emerytura, renta».¹ Artikler der lemmaet følges av en henvisning til et annet lemma, skiller seg ikke fra artikler der det angis én eller flere målspråksekvivalenter: «alderstrygd – alderspensjon». En irriterende detalj i artikler med nummererte målspråksekvivalenter er manglende mellomrom etter nummer fulgt av punktum: «anbefaling – 1.opinia, 2.list polecający».

3. Lemmaene

Lemmaene er alfabetisert etter ord-for-ord-metoden, hvilket innebærer at ordmellomrom gjelder som første tegn i alfabetet, slik at f.eks. *felles foreldreansvar* kommer foran *fellesanlegg*. Den eneste alfabetiseringsfeilen jeg har oppdaget, opptrer under bokstaven S, hvor det mellom *straffeformildende omstendighet* og *straffegebyr* har kilt seg inn elleve feilalfabetiserte lemmaer som begynner på *straffri-* og *straffull-*. Feilen har antagelig sammenheng med at det i usus er vakling mellom *straff-* og *straffe-* som første sammensetningsledd.

Siden det som i andre ordbøker antagelig ville blitt behandlet som eksempler, gjennomgående (men ikke alltid) er skilt ut som egne artikler, består lemmaene for en stor del av syntagmer. I tillegg til lemmaet *påstand* opererer ordboken således med hele fem flerordslemmaer der *påstand* er første ledd (*påstand i ankesak, påstand mot påstand, påstand om erstatning, påstand om frifinnelse, påstand om full frifinnelse*), samt med fire lemmaer der *påstand* er objekt for et verb: *fremsette påstand, legge ned påstand (om)*,

1 Tegnsetning og mellomrom i eksemplene er som i originalen.

*nedlegge påstand*², *nedlegge påstand om (...) års fengsel*. Oppføringen av eksempler som egne oppslag er imidlertid ikke konsekvent gjennomført – under *hevd* finner man f.eks. *ha hevd på å bruke en vei* og to andre eksempler.

I en del tilfeller etterfølges et lemma av nøyaktig samme lemma, men med andre ekvivalenter. Det gjelder f.eks. *fritt rettsråd*, *kapitalinnskudd*, *prinsipal påstand*, *privat straffesak* og *voldsoffererstatning*. De to lemmaene *la saken avgjøre ved votering* og *la saken avgjøres ved avstemning* kan, til tross for at de er ulike på to punkter, også plasseres i samme kategori.

Bokens preg av å være mer en opplisting av ord og ordforbindelser enn en strukturert ordbok gjør det dessverre ofte ganske vanskelig å finne det brukeren leter etter. *Grunnbokshjemmel*, *inhabilitetsinnsigelse* og *omsetningsverdi* forekommer f.eks. ikke som egne lemmaer, men må søkes under henholdsvis *ha grunnbokshjemmel*, *gjøre inhabilitetsinnsigelse gjeldende mot dommer* og *fastslå omsetningsverdien*. I andre tilfeller er ordboken svært raus med oppføringer. Som egne lemmaer opptrer således ikke bare *partsevne*, *prosessuell handleevne*, *rettsbruddsevne* og *rettsevne*, men også, nokså unødvendig, syntagmene *ha partsevne*, *ha prosessuell handleevne* osv.³ Ved enkelte lemmaer kan man lure på om det i det hele tatt finnes brukere som vil falle på å slå opp på dem. Det gjelder f.eks. *hindring av noens straffellelse* (dvs. *straff-fellelse*; *straff-felle* og *straffellelse* finnes ikke som egne oppslag), *få en måneds oppsigelse* og *meddelelse om å ta begjæring med tvangsfullbyrdelse til følge*. Det sistnevnte eksempelet er attpåtil uforståelig, for det

2 Fra *nedlegge påstand* henvises det til *fremsette påstand*, til tross for at *fremsette påstand om* som juridisk term med betydningen 'si hva man mener rettens dom bør gå ut på' er langt sjeldnere enn *nedlegge påstand om* og *legge ned påstand om*; *fremsette* brukes først og fremst i allmennspråklige uttrykk som *fremsette en påstand*, dvs. 'hevde noe (som sant)'.

3 Syntagmet *ha rettsevne* står også som eksempel under *rettsevne*, og er der korrekt oversatt med polsk infinitiv + objekt; i artikkelen *ha rettsevne* er det feilaktig gjengitt med presens partisipp + objekt.

må hete *begjæring om tvangsfullbyrdelse*, og mens *beskjed/befaling/ordre/pålegg o.l. om å gjøre noe* er mulige konstruksjoner, kan det samme neppe sies om *meddelelse om å gjøre noe* – det kan bare hete *meddelelse om at noe skal/må gjøres*.

Av lemmaer jeg savner, kan jeg nevne: *annuitetslån* (*serielån* finnes), *bankremisse*, *beføyelse*, *boliglån* (*erhvervslån* finnes), *fedrekvote*, *foreldrepenger*, *graviditet* (*gravid* finnes), *gårdsnummer* og *bruksnummer* (*matrikelnummer* finnes), *mangelsbeføyelse*, *opprette* i betydningen 'etablere, stifte, grunnlegge' (nå finnes bare oppslag der *opprette* betyr 'sette opp (et dokument)'), *salær*, *stebarn*, *stedatter*, *stesønn* (*stefar* og *stemor* finnes).

Den mest påfallende utelatelsen er kanskje ordet *tingrett*, som i 2002 ble innført som navn på domstoler i første instans. Ordboken oppfører kun de eldre betegnelsene *byrett* og *herredsrett*. I begge artiklene blir det i ekvivalentfeltet gitt en opplysning i parentes om at det aktuelle navnet på vedkommende domstol er *tingrett*, men bare i artikkelen *herredsrett* angis en generell og brukbar polsk ekvivalent, nemlig *sąd pierwszej instancji* {= *domstol i første instans*}. Denne ekvivalenten, samt tilhørende forklaring i parentes, figurerer så i den polsk-norske delen som lemmaet *sąd pierwszej instancji* (*obecna nazwa: tingrett*) {= *domstol i første instans* (*nåværende navn: tingrett*)}. En norsk bruker som slår opp på den sentrale termen *tingrett*, finner altså ingenting, og et alternativt søk på *domstol i første instans* er like forgjeves, for dette opptrer hverken som lemma eller som eksempel under *domstol* og *instans*.

De polske motstykkene til *tingrett* og *lagmannsrett* som domstoler i hhv. første og annen instans er *sąd rejonowy* og *sąd okręgowy*. Begge disse finnes som lemmaer i den polsk-norske delen, med hhv. *distriktsdomstol* og *områdedomstol* som foreslåtte ekvivalenter. Det kan være grunner som taler for ikke å oversette *sąd rejonowy* med *tingrett* og *sąd okręgowy* med *lagmannsrett*, men når man lager to nye termer, slik forfatterne har gjort i dette tilfellet, burde det vært tilføyd en opplysning både om hvilket nivå

i domstolssystemet det er tale om, og hva en norsk domstol på tilsvarende nivå heter.

4. De polske ekvivalentene

Den norsk-polske delen er åpenbart den primære; den polsk-norske gir inntrykk av å være laget ved nokså mekanisk å «vrenge» den norsk-polske. Et nokså klart vitnesbyrd om dette er det faktum at når et norsk lemma har flere polske ekvivalenter, eventuelt plassert under ulike betydningsnumre, kan man finne hver ekvivalenttrekke oppført som ett lemma i den polsk-norske delen. Til den norske artikkelen «vedtatt – 1.przyjęty, uchwalony, 2.zaakceptowany, zaaprobowany, uzgodniony» svarer således de to polske artiklene «*przyjęty, uchwalony* – vedtatt» og «*zaakceptowany, zaaprobowany, uzgodniony* – vedtatt». Fremgangsmåten blokkerer i praksis inngangen til lemmaer som *uchwalony* og *uzgodniony*.

Norske brukere får stort sett ingen hjelp til å velge når ordboken oppfører flere ekvivalenter. Slår man f.eks. opp på *forlover*, finner man to nummererte ekvivalenter, hvorav den ene betyr 'vitne ved vielse' og den andre 'kausjonist', men ingen ekvivalentdifferensierende informasjon. Dersom det norske lemmaet har flere betydninger og den polske ekvivalenten bare dekker én av disse, som ved *handle*, er man også helt overlatt til seg selv.

Ved flerordslemmaer er det ofte ikke ordklassesamsvar mellom lemma og ekvivalent. Til norsk verb svarer polsk substantiv i artiklene *forevise veksel, ha grunnbokshjemma, omlevere, reise sak om barnets ektefødsel*. I andre tilfeller er et norsk verb gjengitt med et polsk adjektiv. Det gjelder f.eks. artiklene *ha partsevne, ha prosessuell handleevne, ha rettsbruddsevne, være prosessdyktig*.

I enkelte tilfeller er tilstandsverb gjengitt med overgangsverb, som i ekvivalenten til eksempelet *ha hevd på å bruke en vei* under lemmaet *hevd*, eller overgangsverb med tilstandsverb, som i ek-

sempelet *bli separert etter to års ekteskap* under lemmaet *separert*.

Ordbokens forfattere har en tendens til å sky korte, innarbeidede termer og erstatte dem med mer ordrike varianter. Som ekvivalent til *trafikkforseelse* oppfører de *wykroczenie przeciwko przepisom drogowym*, som gir 0 treff ved et Google-søk, i stedet for det normale *wykroczenie drogowe*, som gir over 50 000 Google-treff. Den vanligste norske betegnelsen på fenomenet, nemlig *brudd på trafikkreglene*, finnes ikke som oppslag. Slår man opp på *anbudskonkurranse*, finner man som ekvivalent ikke *przetarg*, men konstruksjonen *konkurs ofert przetargowych*, som åpenbart er kalkert over den norske termen (et Google-søk på *konkurs ofert przetargowych* gir 0 treff; et søk på *przetarg* gir over fem millioner treff).

I enkelte tilfeller er det som står på målspråkssiden, mer forklaringer enn håndterbare ekvivalenter. Til *rettslig forundersøkelse* angis som polsk term *postępowanie przygotowawcze prowadzone przez sędzię śledczego mające na celu zebranie dowodów i rozstrzygnięcie czy zostanie przedstawiony akt oskarżenia*. Her burde *postępowanie przygotowawcze* stått som foreslått ekvivalent, og resten vært skilt ut som en utfyllende forklaring. I andre tilfeller presenterer ordboken én eller flere ekvivalenter samt en definisjon, uten å gjøre det klart hva som er hva: «**prejudikat** – prejudykat, precedens, uprzedni wyrok innego sądu będący wytyczną dla wyrokowania w podobnych sprawach w przyszłości» (‘prejudikat, presedens, en annen domstols tidligere dom som tjener som rettesnor for pådømmelse av lignende saker i fremtiden’). Den lange definisjonen virker overflødig, all den stund det på polsk finnes nesten likelydende ekvivalenter, men dersom den allikevel skal være med, må det iallfall markeres at den ikke er tenkt som en ekvivalent.

Ikke sjelden er ekvivalentene som foreslås, upresise. Den polske motsvarigheten til *parkeringsbot* er ikke *mandat drogowy*, som betyr *trafikkbot*, men *mandat za (nieprawidłowe) parkowanie* (hverken *trafikkbot* eller *fartsbot* finnes forøvrig som lemmaer), og de to ekvivalentene under *angrefrist* – *termin reklamacji* og *termin w któ-*

rym należy dokonać wymiany – er begge gale, for den første betyr *reklamasjonsfrist* og den andre *byttefrist*. Ved *abort* gis ekvivalentene *aborcja, przerwanie ciąży*, som betyr henholdsvis '(kunstig) abort' og 'svangerskapsavbrudd', men ikke ekvivalenten *poronienie*, som betyr 'naturlig, spontan abort'. Helt galt går det så i neste lemma, *abortere*, der man som polske ekvivalenter finner *dokonać aborcji, przerwać ciążę*, dvs. *foreta/ta abort, avbryte svangerskapet*, til tross for at *abortere* på norsk betyr 'føde et ikke levedyktig foster', og på polsk må gjengis med *poronić*.

Ordboken inneholder i det hele tatt forbausende mange grove feil i overføringen av norske lemmaer og brukseksempler til polsk. Noen eksempler (oversettelsen av de polske ekvivalentene og eventuelle bemerkninger tilføyd av meg står i sløyfeparentes og innledes med likhetstegn):

arvestykke – udział spadkowy {= *arvelodd*}

blodprøve – 1.pobranie krwi {= *blodprøvetaking*},
2.grupa krwi {= *blodgruppe!*}

kabinettsspørsmål – votum zaufania {= *tillitsvotum*}

leie 1 s [...] **bo til leie** – mieszkanie do wynajęcia {= *bo-
lig til leie*}

løsøreregisteret – rejestr prawnych stosunków majątkowych pomiędzy małżonkami {= *register over ektefellers innbyrdes formuesforhold*}; løsøreregisteret er et tinglysningsregister for rettigheter og pant i løsøre (f.eks. motorkjøretøyer), og omfatter ikke bare ektepakter}

mindreåriges rett til å slutte arbeidskontrakt – prawo osoby niepełnoletniej do rozwiązania stosunku pracy {= *mindreårig persons rett til å avslutte arbeidsforhold*}; *slutte* betyr her ikke 'avslutte', men 'inngå'}

nåde 2 v – okazać łaskę, ułaskawić {= *vise nåde, benåde*}; som verb finnes *nåde* bare i uttrykk som (*gud*) *nåde deg hvis ...*}

omgjøre fra dollar til kroner – wymieniać dolary na korony {= *veksle dollar til kroner*}

prokurator – oskarżyciel, prokurator {= *aktor, statsadvokat; prokurator* er på norsk en nedsettende betegnelse på en advokat}

pådra seg ansvar for (noe) – wziąć (coś) na swoją odpowiedzialność {= *påta seg ansvar for noe*}

rettledning – pouczenie o prawach {= *belæring om rettigheter; rettledning* er det samme som *veiledning*}

rettvinklet – zawiły prawnie {= *juridisk innviklet!*}

rådslå – doradzać, udzielać porady {= *råde, gi råd*}

sakfører – powód, skarżący, strona wnosząca powództwo {= *saksøker, klager; sakfører* er en eldre betegnelse på en advokat}

straffebestemmelse – ustawowy wymiar kary {= *strafferamme*}

vedta en dom – wydawać wyrok {= *avsi en dom; å vedta en dom* vil si å akseptere den}

5. Omtekst

Ordboken har ingen liste over anvendte forkortelser. Det er uansett ikke mange – stort sett bare s (substantiv) og v (verb), brukt for å skjelne mellom homografer. Dessuten finner man et par steder «(pot.)», som åpenbart står for *potocznie* (*familiært, dagligtale*). Den siste forkortelsen synes å være brukt helt tilfeldig – den mangler f.eks. ved *på fersken* (en familiær variant av det stilistisk nøytrale *på fersk gjerning*).

6. Samme ordbok i to utgaver

Ordboken foreligger også i en polsk utgave, SŁOWNIK 2010. Denne omfatter ifølge forordet ca. 16 000 lemmer (2 000 mindre enn den norske), og inneholder som et tillegg en samling skjemaer og formularer som det ofte er aktuelt å oversette (erklæring fra brudefolkene før prøving av ekteskapsvilkårene, forlovererklæring, søknad om separasjon, samboerkontrakt, gjeldsbrev o.l.).

Satsen i den polske utgaven er noen hakk bedre enn i den norske. Den mangler den overflødige tankestreken mellom lemma og ekvivalent, og i tillegg til halvfet for kildepråket og normal skrift for målpråket bruker den kursiv ved henvisning fra ett lemma til et annet: «**alderstrygd** zob. *alderspension*». Når lemmaet er et intetkjønnsstavn, er genus vist ved en etterstilt **n**. Eksempler er dessuten i noe større grad innarbeidet i artiklene i stedet for å stå som egne lemmer. Til de to artiklene *allmenne hensyn* og *allmenne hensyn krever påtale* i den norske utgaven svarer således artikkelen «**allmenne hensyn** interes publiczny (społeczny); ~ **krever påtale** interes publiczny wymaga wszczęcia postępowania karnego» i den polske. Et pluss for brukere med godt syn er at de slipper å irritere seg over en typografisk besynderlighet i den norske satsen. Halen (*ogonek*) under den polske bokstaven **ą** er nemlig galt plassert. I stedet for å henge helt til høyre, som den skal, er den forskjøvet et langt stykke mot venstre.

Uansett utgave vil ordboken, alle feil og ufullkommenheter til tross, utvilsomt kunne være nyttig for brukere som behersker begge språk godt. De bør imidlertid vokte seg for å ta alt som står i den, for god fisk. Er de for godtroende, kan de komme ille ut.

Litteratur

FAGORDBOK 1994 = *Fagordbok for tolker. Norsk-polsk. Terminologimedarbeidere for polsk*: Jadwiga Teresa Purzycka Kvadsheim og Ole Michael Selberg. Oslo: Folkeuniversitetets Forlag.

SŁOWNIK 2010 = M. J. Filipek, M. Iwanow, R. Iwanow: *Słownik terminologii prawniczej norwesko-polski polsko-norweski*. Warszawa: C. H. Beck.

Soleng, Żanetta Wawrzyniak og Harald H. Soleng: *Polsk-norsk, norsk-polsk*. Oslo: Ad Infinitum 2008.

Ole Michael Selberg
Tåsenveien 135A
NO-0880 Oslo
o.m.selberg@c2i.net

MELDINGER

Rapport frå den 11. konferansen om leksikografi i Norden, Lund 24.–27. mai 2011

Oddrun Grønvik

2011 var konferanseår for Nordisk foreining for leksikografi. Konferansen markerte samstundes eit jubileum, ettersom det er 20 år sidan den første nordiske leksikografikonferansen gjekk av stabelen, og organisasjonen NFL vart skipa. Vertskap denne gongen var Svenska Akademiens ordboksredaktion, med Anki Mattisson og Birgit Eaker i spissen, med kyndig bistand frå Språkrådet i Noreg. Konferansestaden var universitetsbyen Lund, der SAOB blir redigert frå Ordbokens hus, ein vakker funkisvilla med stor hage rundt (teikna av arkitekt Sten Samuelson, mannen attom Nya Ullevi). Dei faglege arrangementa var samla i bygningen Palaestra, som ligg innanfor det sentrale universitetsområdet – opphavleg gymnastikksal, i dag ein svært funksjonell møtestad for ein konferanse som denne. Det er alltid ei utfordring å halda tre sesjonar gåande i parallell – men det gjekk fint, avstandane mellom romma var små og mellom benkeradene stor nok.

Den 11. konferansen vart også den største i NFLs historie, både når det gjeld deltaking (meir enn 130 frammøtte) og føredrag (meir enn 60 utanom plenarføredraga, avvikla i tre parallelle sesjonar). Ettersom somme av føredraga hadde to eller tre forfatarar, og somme var med på meir enn eitt innlegg, må meir enn halvparten av deltakarane ha vori i elden i løpet av tre ganske tett-pakka dagar. Det er med andre ord ei forsamling med høgt aktivitetsnivå som møtest annakvart år for å snakka om leksikografiens hovudtendensar og irrgangar, presenterte i føredrag, demonstrasjonar av programvare, posterutstillingar og utstillingar av bøker og publikasjonar. Konferansen fekk også god framhjelp av det

strenge (men venlege og rettferdige) korpset av sesjonsstyrarar – det var ikkje mogleg å gå over tida.

Deltakinga var overveldande nordisk, som rimeleg er, med nokre ikkje-nordiske deltakarar. Dei to mest prominente var John Simpson, hovudredaktør i *Oxford English Dictionary* (OED), og Godelieve Laureys, professor i skandinavistikk ved universitetet i Gent, som båe heldt plenarføredrag. Ei påfallande endring i deltakinga om ein ser det nordiske leksikografmiljøet over tid, er elles at det sterke forlagsnærveret frå dei tidlege konferansane langt på veg mangla denne gongen. Dette er vel ei følgje av at ordboksproduksjon ikkje lenger er ei sikker kommersiell satsing. Kva dette kan få å seia for det leksikografiske fagmiljøet i Norden, er det vanskeleg å spå om no. Men møtelyden vart siste dagen sterkt oppmoda til å tenkja over kva nye medium vil innebera for faget, i innlegget til Jens Erlandsen.

NFLs konferansar har sine plenarføredrag, denne gongen var det tre av dei. Dei var svært ulike, men kvar for seg representative for faglege interesser som kom tydeleg fram under sesjonsføredraga. I det første tok Bo Ralph (professor i nordiske språk, Göteborg) oss med på ei erkjenningstrekk i historia og drivkreftene attom leksikografien som disiplin, der spørsmålet ikkje var korleis ein lagar ordbøker, men kvifor. Kva er det å forstå, avgrensa og skildra omgrep – og kva har viljen til å driva med ord, tyding og definerings hatt å seia for framståande svenske språkbrukarar som ikkje primært var leksikografar – med Strindberg som den best kjende? Det kan henda at Bo Ralphs emne, den menneskelege trongen til å fanga omgrep i ord, slik at tanken står støtt, vil bli ei endå tydelegare drivkraft i framtidig leksikografi, og særleg i morsmålsleksikografien, dersom lønsemda i ordboksproduksjon blir svekt. For kulturkodifisering gjennom leksikografisk arbeid vil nok halda fram likevel.

Dei to andre plenarførelisarane syntte, ut frå svært ulike prosjekt, at leksikografisk arbeid og leksikografiske produkt i dag

knapt kan tenkjast utan bruk av IKT, anten det gjeld materialanalyse, arbeidsmetode, produktutforming eller samanlenking av tilfang og produkt. John Simpson drøfta korleis storverket OED på nett blir lagt til rettes for nye og større brukargrupper. Godelieve Laureys drøfta føremonar og ulempar ved å bruka IKT til å snu ei tospråkleg ordbok, slik at målspråk blir kjeldespråk og omvendt, og kva for utfordringar dette inneber for leksikografen. Både emne, med vinkling, er sentrale i dagens leksikografiske arbeid og fekk lyset mot seg også i fleire sesjonsføredrag, og både førelesarar synta at leksikografen aldri blir arbeidslaus, same kva verkty ein får i hende.

Den 11. konferansen om leksikografi i Norden hadde som uttrykt mål «med det vetenskapliga programmet ... att ge en så bred bild av den nordiska lexikografien som möjligt». I innhald spente sesjonsforedraga over leksikografi i stort og lite format, eittspråkleg og tospråkleg, nytt og gammalt, med utgangspunkt i skrift, tale-mål eller både delar. Ein kan grovinnndela føredraga etter emne og vinkling – det er gjort nedanfor. Men samstundes sit ein att med ei sterk kjensle av at alt heng saman med alt, og at dette ytrar seg i felles haldningar og interesser for emne som kategorisering, tilhøvet mellom rådata og produkt, systematisk innhaldsformidling, og alle metodar og verkty som kan sikra kvalitet og konsistens på desse områda.

Ettersom det er uråd å gjera rett mot alle i referatet frå ein så mangslungen konferanse, ber eg om orsaking her og no til alle som ser seg unemnde, og dreg fram nokre grupper av føredrag som fall meg i augo.

Tenkjer ein sjanger, så baud programmet på innlegg om heile spennet av leksikografiske produkt, frå dei store (og gamle) historiske ordboksverka, til små ordboksdatabasar for barn og ungdom. Det som var sams for dei, var at Internett var publiseringsmediet. Anna Helga Hannesdóttir, Margunn Rauset og Aldís Sigurðardóttir presenterte til dømes utfordringane i ISLEX-programmet, eit ærgjerrig opplegg for å laga nettbaserte tospråksordbøker mel-

lom islandsk og alle dei andre nordiske språka. Her ligg det under ein premiss om at tilhøvet skal vera så godt som identisk mellom islandsk og det andre språket i kvart av språkpara, til dømes i tydingsinndeling og illustrerande døme. Slik er ikkje røynda, og dermed må leksikografane hala og dra i det formatet dei har fått utlevert for å få til adekvate løysingar. Åse Wetås og Sturla Berg-Olsen kunne i innlegget sitt om ein nettpresentasjon av *Norsk Ordbok 2014* syna fram det motsette problemet: Artikkelformatet i denne store ordboka er finkategorisert og svært rikt. Korleis skal ein best forenkla nettpresentasjonen, slik at nettordboka blir nyttig og interessant for vanlege brukarar?

Også den leksikografiske tidsaksen for Norden vart dekt over heile spennet frå 1500-talet til i dag. Eit tital sesjonsforedrag tok for seg anten tidlegare leksikografiske arbeid, eller eldre tilfang som har leksikografisk relevans. Såleis fekk vi høyra om dei første islandske ordbøkene (Gunnlaugur Ingólfsson), hapax i Olof Verelius' *Index lingvæ veteris scytho-scandicæ sive gothicæ* (Anton Lundqvist), Ivar Aasen og svensk leksikografi (Dagfinn Worren og Kjetil Gundersen) og nye kjelder til den danske leksikografis historie (Simon Skovgaard Boeck). Foredrag av denne typen – alltid lærde, ofte morosame – er vel noko av desserten ved dei leksikografiske konferansane, og tener ofte som påminning om at det finst mindre nytt under sola av faglege problemstillingar enn ein somtid har trudd. Den diakrone vinklinga kjem også fram når ein jamfører eldre og nyare utgåver (Sven-Göran Malmgren: *Från Svenska ord 3 till Svenska ord 4*), eller legg inn mange utgåver av same ordlista i ein database (Louise Holmer om SAOL).

Eit IKT-bidrag til både forskning og folkedaning er elles at stadig meir av det eldre tilfanget kan gjerast tilgjengeleg på nettet, anten det er tekst som syner språket i bruk, leksikografiske arbeid eller anna tilfang som har leksikografisk relevans. Døme på dette er digitaliseringa av karttilfanget med samletittelen *Norsk dialekt-atlas*, som vart presentert av Jon Holmen. Dette tilfanget utgjer i

seg sjølv eit målepunkt for norske – og somme svenske – dialektar i første halvdel av 1900-talet. Men teknologien attom kombinasjonen av lingvistiske data og kart kjenner ingen grenser i høve til geografi eller tid, og dermed opnar det seg vyar for automatisert framsyning av geografisk utbreiing for språklege data, som kan koma til nytte for heile fagfeltet.

Ein leksikografisk spesialdistanse er kategorisering av data og organisering av desse kategoriane i faste format. Emnet er like gammalt som leksikografien sjølv, men IKT-bruk driv fram strengare krav til konsistens og redaksjonell sjølvdisiplin enn papir kan gjera. Ei ganske stor gruppe sesjonsforedrag handla om kategorisering, frå val av ordklassenemning (Jón Hilmar Jónsson, Leiv Inge Aa) til full analyse og konvertering av artikkelformat i ei akademisk ordbok (Ilse Cantell og Caroline Sandström om dei første banda av *Ordbok över Finlands svenska folkmål*).

Generering av ordnett frå ordbøker og korpus er en grein av leksikografien som blir stadig viktigare, og to foredrag presenterte slike prosjekt (Ruth Vatvedt Fjeld, Rune Lain Knudsen og Julie M. Torjusen frå Noreg og Kristín Bjarnadóttir frå Island), likeins teknikkar for å isolera det sentrale ordtilfanget i eit språk frå korpus (ofte med tanke på innlæring) (Sofie Johansson Kokkinakis og Elena Volodina; Daniel Ridings). Dette er ikkje særnordiske interessefelt, men det er viktig og nyttig at ulike teoriar og metodar knytt til desse fagfelta blir testa ut på nordiske språk. At sambandet mellom ordnett og ontologiar er eit gammalt interessefelt for leksikografar fekk Lars Vikør synt fram i innlegget sitt om to norske tesaurusar, laga med om lag 140 års fråstand.

Etter den 10. konferansen vart det oppsummert at konferansen alt i alt hadde stått i «i sprogets tegn» – og referenten (Ptaszynski 2010) etterlyste bidrag som kunne syna at leksikografi er mykje meir enn språk. Ein kan vel ikkje seia at han fekk ønsket sitt innfridd denne gongen – det er framleis språket som ligg i sentrum for NFL-konferansen. Om ein ønskjer andre leksikografiske ovrin-

gar enn dei språkbundne (for også *Norsk dialektatlas* må seiast å vera språkbunde) må ein nok gå aktivt ut og rekruttera. Dette kan jo komiteen for den 12. konferansen, som skal vera i Oslo i 2013, ta for seg til nøyare vurdering.

Konferansens sosiale program dekte dei tre kveldane deltakarane fekk i Lund. Under velkomstmottakinga fekk ein sjå seg om i Ordbokens hus og hage, og SAOB-redaksjonen fekk synt at dei kan meir enn å skriva ordbok. Dei hadde laga mat og den var veldig god, det same var stemninga. Ei uvanleg kraftig haglbye vart berre eit pikant innslag på veg ut til partytellet. Kvelden etter var det byvandring i Lund med omvisarar som kunne sine ting. Men Lund domkyrkje fekk vi ikkje vitja, for der heldt ein på og førebudde Doktorpromosjonen (sjå nedanfor). Til gjengjeld var vi atter bedne til Ordbokens hus, der det vart gjevi ein demonstrasjon av arbeidet med SAOB – Nordens suverent største ordboksverk, og det eldste som er under arbeid. Torsdag kveld var det konferansemiddag på Svaneholms slott, utanfor Skurup i søre Skåne, ein times veg med buss frå Lund. Her ville lagnaden at forsamlinga delte seg i to. To tredelar reiste beint til Svaneholm og fekk god tid med glas og musikk. Ein tredel stod att ganske lenge og venta på den siste bussen. Til gjengjeld fekk dei ein bussjåfør som underheldt med lokalhistorie og Svaneholm-anekdotar heile vegen ut. Takk vere denne sterke introduksjonen vart det, etter ein ypparleg middag, museumsvitjing og ein gløtt inn i ei svunnen tid i Skåne. Konferansens siste dag var også dagen for Doktorpromosjonen ved Lunds universitet i 2011. Doktorpromosjonen er den største akademiske høgtidsdagen, og har så vore sidan 1670. Dermed fekk konferansedeltakarane stetta turisttrongen den dagen også – promosjonen var eit imponerande syn – og somme kom seg etterpå endeleg inn i Lunds domkyrkje!

Som vanleg vart NFL-konferansen avslutta med generalforsamling. Her vart det lagt fast at neste NFL-konferanse blir i Noreg i 2013 (sjå Jan Hoels bidrag i dette bandet).

Litteratur

Ptaszynski, Marcin Overgaard 2010: Rapport fra den 10. Konference om Leksikografi i Norden, Tammerfors, 3.–5. juni 2009. I: *LexicoNordica* 17, 407–413.

Oddrun Grønvik
hovudredaktør, Dr. Litt. H. C. University of Zimbabwe
Norsk Ordbok 2014
Institutt for lingvistiske og nordiske studium
Universitetet i Oslo
Postboks 1021 Blindern
NO-0315 Oslo
oddrun.gronvik@iln.uio.no

Invitasjon til den 15. Euralex-konferansen i Oslo 7.-11. august 2012

Ruth Vatvedt Fjeld

Den 15. europeiske konferansen i leksikografi holdes ved Det humanistiske fakultet ved Universitetet i Oslo. Det er første gang denne konferansen arrangeres i Norge, tidligere Euralex-konferanser i Norden har vært i Tampere i Finland (1992), i Göteborg i Sverige (1996) og i København i Danmark (2002). At den europeiske konferansen i leksikografi holdes i et nordisk land, er en særlig begivenhet for medlemmene i Nordisk forening for leksikografi, og normalt deltar det flere nordiske medlemmer av foreningen enn vanlig.

Hovedtema på konferansen skal være Leksikografi og nasjonal identitet, og det planlegges et førseminar (pretutorial) om temaet Leksikografi og urbefolkningspråk, der vi fra norsk side bl.a. vil legge vekt på samisk språk og ordbøker over de samiske varietetene i Norden. Det kommer svært mange ledende eksperter på språkvitenskap og leksikografi fra hele Europa til denne konferansen, og vi oppfordrer selvsagt alle medlemmer av Nordisk forening for leksikografi om å delta og helst bidra med faglige innlegg. Vi ønsker også ellers å vektlegge det nordiske aspektet i leksikografien. Følgende sesjoner er satt opp for konferansen, i tillegg til de inviterte plenarforelesningene:

1. Lexicography and nationalism
2. Reports on Lexicographical projects
3. Lexicography in Language technology
4. Bilingual Lexicography
5. Terminology, LSP and Lexicography

6. Lexicography and Semantic Theory
7. Corpus-driven lexicography
8. Indigenous languages and lexicography
9. Other topics

Det vil også bli arrangert emnebaserte workshops, software-demonstrasjoner og en postersesjon.

Som mange andre foreninger har også Euralex problemer med synkende medlemsmasse, og vi vil oppfordre alle medlemmer av Nordisk forening for leksikografi til å utvide sine faglige perspektiver og sitt faglige nettverk ved å melde seg inn i Euralex. Blant annet gir det noe lavere kontingent når man deltar på Euralex-konferansene, og man er stemmeberettiget ved generalforsamlingene. Innmelding kan gjøres på Euralex' hjemmeside <http://www.euralex.org>.

Man kan melde seg på konferansen i Oslo her: <http://www.hf.uio.no/euralex>.

Vi ønsker alle LN-lesere velkommen til faglig interessante og utfordrende dager i Oslo i august 2012!

Ruth Vatvedt Fjeld
leder av arrangementskomitéen
Institutt for lingvistiske og nordiske studier
Universitetet i Oslo
Postboks 1001 Blindern
NO-315 Oslo
r.e.v.fjeld@iln.uio.no

Nytt fra styret for Nordisk forening for leksikografi

Jan Hoel

Etter foreningens generalforsamling i Lund 24. mai 2011 konstituerte det nye styret seg på denne måten:

Jan Hoel, Norge, leder 2011–2013 (innvalgt 2009)

Caroline Sandström, Finland, nestleder 2011–2013 (innvalgt 2009)

Zakaris Svabo Hansen, Færøyene, kasserer (2011–2015)

Henrik Hovmark, Danmark, sekretær (2011–2015)

Anki Mattisson, Sverige (2011–2013)

Varamedlemmer:

Ilse Cantell, Finland (2011–2013)

Åse Wetås, Norge (innvalgt 2009)

Neste NFL-konferanse vil bli avholdt i Oslo i **midten av august 2013**, nærmere bestemt onsdag 14., torsdag 15. og fredag 16. august, med uformell mottakelse og registrering tirsdag 13. august om kvelden. Det er opprettet en arrangementskomité og en programkomité med medlemmer fra Institutt for lingvistiske og nordiske studier ved Universitetet i Oslo og Språkrådet i Norge, som i fellesskap er ansvarlige for arrangementet.

Foreningens nettsider er flyttet over til nettstedet til Nordisk språkkoordinasjon i København (<http://nordisksprogkoordination.org/nfl>). Det nye styret vil vurdere om det er behov for å innføre en styrefunksjon som nettsideansvarlig. Inntil videre har sekretær Henrik Hovmark sagt seg villig til å ivareta oppgaven.

Gjennom Nordisk ministerråds program *Nordplus språk og*

MELDINGER

kultur er det bevilget 15 000 EUR til en digital publiseringsplattform for språkressurser i Norden. ”Digitaliseringsprosjektet” er igangsatt og omfatter blant annet NFLs konferanserapporter og tidsskriftet *LexicoNordica*, så langt midlene strekker til.

Jan Hoel
seniorrådgiver
Språkrådet
Postboks 8107 Dep
NO-0032 Oslo
jan.hoel@sprakradet.no

REDAKSJONELT

1. LexicoNordica **udkommer** hvert år i november. Tidsskriftet indeholder leksikografiske bidrag som er skrevet på et af følgende nordiske sprog: dansk, finsk, færøsk, islandsk, norsk (bokmål eller nynorsk), svensk. Bidrag på engelsk, fransk eller tysk kan også optages hvis særlige forhold taler for det.
2. **Bidrag** sendes til det medlem af redaktionskomitéen som repræsenterer bidragerens land:

Sturla Berg-Olsen, Norsk Ordbok 2014, Institutt for lingvistiske og nordiske studier, Universitetet i Oslo, Postboks 1021 Blindern, NO-0315 Oslo. <sturla.berg-olsen@iln.uio.no>

Ken Farø, Københavns Universitet, Institut for Engelsk, Germanisk og Romansk, Njalsgade 128, DK-2300 København S. <kenfaroe@hum.ku.dk>

Annika Karlholm, Institutet för språk och folkminnen, Dialekt- och folkminnesarkivet i Uppsala, Box 135, SE-751 04 Uppsala. <annika.karlholm@sofi.se>

Veturliði G. Óskarsson, Institutionen för nordiska språk, Uppsala universitet, Box 527, SE-751 20 Uppsala. <veturlidi.oskarsson@nordiska.uu.se>

Mariann Skog-Södersved, Vasa universitet, Tyska språket och litteraturen, PB 700, FI-65101 Vaasa. <mss@uwasa.fi>

Seneste tidspunkt for aflevering af bidrag er **den 1. april** hvis artiklen skal kunne trykkes i det nummer af tidsskriftet som udkommer i november samme år. Bidraget indleveres i elektronisk form, fx i Word-, Open Office- eller RTF-format.

3. **Illustrationer** der skal medtages i artiklen, indsættes i manuskriptet og vedlægges som separate grafikfiler.

4. **Manuskriptet:** Bidraget forfattes i LexicoNordicas stilark, der kan rekvireres ved henvendelse til redaktionen. Manuskriptet **indledes** med forfatternavn og titel på artiklen. For tematiske og ikke-tematiske bidrag følger et **abstract** på engelsk på ca. ti linjer og dernæst selve artiklen, som opdeles i kapitler. Bidraget afsluttes med angivelse af forfatterens navn, titel samt post- og e-mail-adresse. Bidrag kan normalt have et omfang på højst 20 sider.
5. **Citater:** kortere citater (op til 3 linjer) bringes som en del af teksten med dobbelte anførselstegn omkring, mens længere citater eller fremhævelser af større vigtighed bør gives i et afsnit for sig selv **uden** anførselstegn.
6. Vi anbefaler en tilbageholdende brug af **fodnoter**. Evt. nødvendige noter gennemnummereres i teksten med højtstillet angivelse uden parentes.
7. **Litteraturhenvisninger** foretages i teksten på følgende måde:

Herbst (2009:158) eller Herbst (2009).

I den løbende tekst angives ikke hele internetadresser, men et forfatternavn eller en angivelse af titlen på internetbidraget, som bruges i litteraturlisten. Her angives internetadresser uden understregning, men omgivet af <>.

8. Særlige angivelser: **leksikografiske termer** kan, når de indføres, fremhæves med fed; **sprogksempler** markeres med kursiv, fx: ordet *ungkarl* har synonymet *alenemand*; **betydninger** af sproglige enheder angives ved hjælp af enkelte anførselstegn, fx: 'en ugift mand'; dobbelte anførselstegn bruges ved **citater**

eller **forbehold**, fx: De er vokset op i de “glade” tressere. Tegnsætningsreglerne, bl.a. for brug af komma, bindestreg og tankestreg, er forskellige i de nordiske lande, og forfatterne bør følge reglerne for det sprog som bruges i artiklen.

9. Eksempler på litteraturangivelser

Ordbøger

ALD (1948) = A.S. Hornby, E.V. Gatenby & H. Wakefield: *A Learner's Dictionary of Current English*. London: Oxford University Press.

COBUILD (1987) = *Collins COBUILD English Language Dictionary*. Editor in Chief: John Sinclair, Managing Editor: Patrick Hanks. London/Glasgow: Collins.

ELEXIKO = Annette Klosa m.fl. (red.): *elexiko*. Mannheim: Institut für Deutsche Sprache. <www.elexiko.de> (maj 2008)

Hällström, Charlotta af & Mikael Reuter (2001): *Finlandssvensk ordbok*. (1. upplagan). Helsingfors: Schildts.

Norstedts stora engelska ordbok (2000). Stockholm.

Oxford-Hachette French Dictionary (1994). Oxford: Oxford University Press.

STO = Anna Braasch m.fl. (red.): *Sprogteknologisk Ordbase*. København: Center for Sprogteknologi. <www.cst.dk/cgi-bin/defisto> (april 2007)

Anden litteratur

Haiman, John (1980): Dictionaries and Encyclopedias. I: *Lingua* 50, 329–357.

Lakoff, George & Mark Johnson (1980): *Metaphors we live by*. Chicago and London: The University of Chicago Press.

Mugdan, Joachim (1985a): Grammatik im Wörterbuch: Wortbildung. I: Herbert Ernst Wiegand (udg.): *Studien zur neuhochdeutschen Lexikographie IV*. Hildesheim/Zürich/New York: Olms, 237–308.

NRG = Jan Terje Faarlund, Kjell Ivar Vannebo & Svein Lie (1997): *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.

Svenonius, Peter (1996): The Verb-Particle Alternation in the Scandinavian Languages. <<http://ling.auf.net/lingBuzz>> (september 2011)

Zgusta, Ladislav (1971): *Manual of lexicography*. The Hague: Mouton.

10. LexicoNordica udkommer både som trykt tidsskrift og i en **internetudgave**. Ved indsendelse af et bidrag til redaktionen erklærer forfatterne sig derfor indforstået med både trykt og elektronisk publicering.