

MEIJERBERGS ARKIV

FÖR

SVENSK ORDFORSKNING

UTGIVET
AV
STYRELSEN FÖR MEIJERBERGS INSTITUT
VID GÖTEBORGS UNIVERSITET
GENOM
BO RALPH

27

GÖTEBORG 2001

MEYERBERG'S ALKYL

WATER SOLUBLE

THE
MAY 1957

1957

NORDISKA STUDIER I LEXIKOGRAFI

5

Rapport från Konferens om lexikografi i Norden
Göteborg 26–29 maj 1999

Redigerad av
Martin Gellerstam, Kristinn Jóhannesson,
Bo Ralph och Lena Rogström

Skifter utgivna av Nordiska föreningen för lexikografi
Skift nr 6
I samarbete med Nordiska språkrådet och Meijerbergs institut

Ingår som del 27 i Meijerbergs arkiv för svensk ordforskning

© Nordiska föreningen för lexikografi och
Meijerbergs institut för svensk etymologisk forskning
Göteborgs universitet
Box 200
SE-405 30 Göteborg

ISSN 0348-7741
ISBN 91-631-0854-2

Novum Grafiska AB, Göteborg 2001

Innehåll

Förord	1
<i>Anna Helga Hannesdóttir</i> Funktioner som försvann: det kontrastiva perspektivet i äldre svensk lexikografi	3
<i>Maria Arnstad</i> Idiombruk i talspråk	18
<i>Antti Arpe</i> Lärdomar från utveckling av inflekterande synonymordböcker	31
<i>Tove Bjørneset</i> Introduksjon till ordboksprojektet NORDLEXIN-N	44
<i>Erica Brewitz & Anna-Lena Bucher</i> TNCs arbete med Eurodicautom, Europeiska kommissionens termdatabas	54
<i>Yvonne Cederholm & Lena Rogström</i> Gamla ord blir som nya. OSA-databasen som källa och resurs	66
<i>Ulla Clausén</i> Funktion och attityd i Svensk konstruktionsordbok	80
<i>Elżbieta Dura</i> Automatisk extrahering av ordleder	94

<i>Øystein Eek</i> Marknadsorientert lemmaseleksjon: kriterier og prioriteringar	111
<i>Guðrún Kvaran</i> Tanker om en islandsk fremmedordbog	125
<i>Åsa Holmér</i> Vad ska <i>chatta</i> heta på finska? Efcot-projektet, att bygga upp ett europeiskt forum för dataterminologi	137
<i>Dimitrios Kokkinakis & Sofie Johansson Kokkinakis</i> Sense-Tagging at the Cycle-Level Using GLDB	146
<i>Ulrika Kvist Darnell</i> Rörande polysem belägenhet	168
<i>Godelieve Laureys & Gudrun Rawoens</i> Bilingval leksikografi. Nederländsk–dansk ordbogsprojekt	185
<i>Henrik Lorentzen</i> Jagten på det gode citat. Om vanskelighederne ved at finde egnede ordbogseksempler i et korpus	202
<i>Nina Martola</i> Översättningsmodeller i en ordbok? En liten användarundersökning	217
<i>Torbjørn Nordgård</i> Valensinformasjon i NorKompLeks	236
<i>Christian-Emil Ore</i> Metaordboken – et rammeverk for Norsk Ordbok?	250

<i>Riitta Palkki</i> Lagring av ordbok över äldre finskt skriftspråk i strukturerad form	271
<i>Bo Ralph</i> Orden i ordning. Den historiska framväxten av en lexikografisk tradition i Sverige (plenarföredrag)	282
<i>John Sinclair</i> Lexical Grammar (plenarföredrag)	323
<i>Emma Sköldbberg</i> Lever idiomen i svenska ordböcker sitt eget liv?	344
<i>Bo Svensén</i> Markeringsangivelser i ordböcker: några strukturella aspekter	357
<i>Lars Törnqvist</i> Från engelsk-svensk namnordbok till svensk-engelsk	376
<i>Ruth Vatvedt Fjeld</i> Om leksikalisk ubestämthet	386
<i>Fredrik Westerlund, Sjur Nørstebø Moshagen, Eva Orava & Juhani Birn</i> Morfologisk analys och disambiguering som stöd i skapandet av frekvensordlistor	407
<i>Lars Vikør</i> Purisme på hell? Nye tendensar i nynorsk ordforrådsnormering	425
<i>Matti Vilppula</i> Är eller var? Tempusproblem i ordboksdefinitioner	437

Jonas Winnerlöv

Länk och ekvivalens i en mångspråkig
lexikografisk relationsdatabas 449

Författarna 465

Förord

Den femte konferensen om Lexikografi i Norden hölls i Göteborg den 26–29 maj 1999. Arrangörer var Institutionen för svenska språket vid Göteborgs universitet i samarbete med Nordiska föreningen för lexikografi och Nordiska språkrådet. Den lokala arrangörskommittén bestod av Martin Gellerstam (ordf.), Sven-Göran Malmgren och Lena Rogström.

Konferensen samlade 110 deltagare och 33 föredragshållare fördelade på två sektioner. Arrangörerna inbjöd till bidrag inom hela det lexikografiska fältet, utan att ange ett specifikt tema. I denna volym publiceras 29 av föredragen. När det gäller de icke publicerade föredragen så har författarna av tidsbrist eller andra skäl avstått från att lämna in en slutgiltig version. Föredragen publiceras som vanligt på svenska, danska eller norska. I två fall (John Sinclair och Sofie Johansson Kokkinakis & Dimitrios Kokkinakis) är föredragen på engelska. Konferensens två plenarföredrag hölls av John Sinclair och Bo Ralph. På programmet stod också en paneldiskussion kring *Nordisk leksikografisk ordbok* med författarna som deltagare i panelen.

Vid sidan av programmet presenterade nordiska ordboksförlag en bokutställning.

I konferensen ingick också en utflykt till Ostindiefararen i Göteborgs hamn och en mottagning på Börsen, arrangerad av Göteborgs stad.

Konferensen stöddes ekonomiskt av Humanistisk-Samhällsvetenskapliga forskningsrådet, Svenska Akademien och Kungl. Gustav Adolfs Akademien. Meijerbergs institut vid Göteborgs universitet har publicerat konferensrapporten som ett nummer i serien *Meijerbergs arkiv för svensk ordforskning*. Humanistiska

fakulteten vid Göteborgs universitet ställde lokaler till förfogande för konferensen.

Tidigare konferensvolymerna har kompletterats med ett ordregister som för varje volym har fått en alltmer ambitiös utformning. Vi har med viss tvekan brutit denna praxis, dels för att inte ytterligare försena en redan sen tryckning, dels med bitanken att göra ett kumulerat ordregister för samtliga volymer som kan läggas på nätet.

Till alla som på olika sätt bidragit till konferensen och till denna konferensrapport riktas ett hjärtligt tack.

Göteborg i december 2000

Martin Gellerstam
Kristinn Jóhannesson
Bo Ralph
Lena Rogström

Anna Helga Hannesdóttir

Funktioner som försvann: det kontrastiva perspektivet i äldre svensk lexikografi

This article deals with the contrastive perspective in Swedish lexicography in the 19th century. The first complete monolingual dictionary of Swedish, published 1850–1855, depends largely on the bilingual tradition that had been developed during the 18th and early 19th century. In spite of this, it does not exhibit any traces of a contrastive perspective. In the two monolingual fragments, Kindblad 1840[–41] and Almqvist 1842[–44], on the other hand, a contrastive analysis is central to the lexical description of the headwords. In fact, the fragments demonstrate a very different and theoretically advanced lexical analysis compared to the succeeding monolingual dictionaries of Swedish.

Äntligen intresserar sig forskarna för den tvåspråkiga lexikografen!

När den första nordiska konferensen i lexikografi hölls 1991 ägnades en viss uppmärksamhet åt den då pågående diskussionen om huruvida lexikografen egentligen var en vetenskaplig verksamhet eller om den bara var "en ren ordinnsamling av ateoretisk karakter" (Fjeld 1992:1). Då konferensen var avslutad var nog de flesta av deltagarna överens om att visst var lexikografen – om inte redan så i alla fall på god väg att bli – en egen vetenskap.

De senaste åren har intresset ökat markant också för den tvåspråkiga lexikografen. I stället för att uppfattas som en verksamhet som i allt väsentligt varit sekundär till den enspråkiga lexikografen – och till och med som en ganska trivial verksamhet – har den kommit att uppmärksammas som ett eget, självständigt område inom disciplinen, ett område med egna syften och metoder som också kräver egen teoribildning. Detta är glädjande på många sätt.

I Sverige har den enspråkiga lexikografen så länge den har funnits haft en betydligt högre status än den tvåspråkiga. Detta

beror möjligen delvis på att den enspråkiga svenska lexikografins etablering sammanföll i tiden med ett idéklimat som dominerades av nationalromantiken. Det upplevdes som besvärande – eller rent av pinsamt – att det vid 1800-talets mitt ännu inte fanns någon ordbok över det svenska språket. Detta påpekas av Anders Fredrik Dalin i förordet till den första enspråkiga svenska ordboken, *Ordbok öfver svenska språket*. Den började ges ut 1850 och förelåg färdig 1855. Han kommenterar där den ringa aktiviteten inom den svenska språkforskningen rent allmänt, och han yttrar sig särskilt om läget på ordboksfronten. Enstaka insatser har visserligen gjorts, säger han,

men inga mera omfattande, djupgående forskningar, åsyftande att i hela dess vidd utreda sambandet emellan språkets nutid och forntid, hafva försports*); och – hvad som är ännu märkvärdigare – ett folk, som i nästan alla bildningens riktningar höjt sig till jemnbredd med jordens mest hyfsade nationer, saknar ännu en ordbok öfver sitt eget språk, något som de flesta Europas folkstammar äga, till och med de ofria. (Dalin 1850:71)

I den not som följer det citerade stället undantar han dock Johan Rydqvists *Svenska språkets lagar*, vars första del då precis hade kommit ut. Det skulle dröja ytterligare fem delar och 33 år innan dennes arbete förelåg komplett.

Liksom språkvetenskapen i allmänhet kom också den svenska lexikografen att präglas av de nationalromantiska idéerna. Från det att den enspråkiga lexikografen väl var etablerad kom den lexikografiska verksamheten som var akademiskt baserad helt att domineras av enspråkiga, historiskt inriktade projekt eller också av uppteckning av dialekterna och i förlängningen av det utgivningen av dialektordböcker. Dessutom naturligtvis av Svenska Akademiens ständigt aktuella ordboksprojekt – som till slut kom igång. Med tanke på att lexikografen vid sekelskiftet 1900 av självaste Noreen karakteriserades som en ”osystematisk och såtillvida ovetenskaplig uppställning af grammatikens innehåll” (Noreen 1903:49), är det inte särskilt anmärkningsvärt att tvåspråkiga ordböcker och tvåspråkig lexikografi inte tilldrog sig någon större uppmärksamhet från forskarsamhället.

Dalins *Ordbok öfver svenska språket* har framstått som ett pionjärarbete, märkligt bland annat genom att det tog så kort tid att fullborda det, och även genom att den enspråkiga lexikografin där framträdde för första gången och ändå så färdig i sin form. Vad hade då Dalin att bygga på? Han hade åtskilligt. För det första hade han den färdiga modell för den lexikaliska beskrivningen av svenskan som hade utvecklats i 1700-talets och det tidiga 1800-talets tvåspråkiga ordböcker. Den utvecklingen har jag beskrivit i annat sammanhang (Hannesdóttir 1998). För det andra hade han de två fragment till enspråkiga ordböcker som i 40-talets början publicerades av Karl Eduard Kindblad respektive Carl Jonas Love Almqvist. Sist men inte minst var han inte någon duvunge själv; han började sin lexikografiska verksamhet redan i slutet av 30-talet och gav bl.a. ut en stor fransk-svensk ordbok 1842–43.

Idag betraktas som sagt den enspråkiga och den tvåspråkiga lexikografin som två helt separata områden inom den lexikografiska disciplinen. Den enspråkiga lexikografins syfte kan ju formuleras så, att den bidrar till beskrivningen av det aktuella språket samt att den ger anvisningar om de olika ordens egenskaper och användning, medan den tvåspråkiga lexikografins syfte i första hand har varit pedagogiskt eller att framställa praktiska redskap för att underlätta olika slag av kommunikation mellan talare av olika språk. Det är lätt att hålla med Kromann, Riiber och Rosbach när de påpekar att det på alla plan är "easy to demonstrate the empirical dependence of the large bilingual dictionaries on the major national monolingual dictionaries" (Kromann, Riiber & Rosbach 1991:2714).

Detta skulle innebära att den tvåspråkiga lexikografin är logiskt sekundär i förhållande till den enspråkiga. Samtidigt kan man inte komma ifrån att den tvåspråkiga lexikografin har en betydligt äldre historia än den enspråkiga. Samma forskare har alltså rätt också när de påpekar att "it seems reasonable to say that bilingual or multilingual lexicography was the natural predecessor of monolingual lexicography" (Kromann, Riiber & Rosbach 1991:2712). Här förutsätts en lexikografisk utveckling som inte bara har inneburit att en ny typ av ordböcker har

utvecklats, dvs. de enspråkiga. Rimligen fanns det också en kvalitativ utveckling, eller som de också konstaterar: "there was a natural evolution from a primitive lexicography of translation to a highly-developed and sophisticated monolingual lexicography" (Kromann, Riiber & Rosbach 1991:2712).

Frågan är då: hur primitiv var den tvåspråkiga svenska lexikografin vid 1800-talets mitt – eller, hur långt hade lexikografins kvalitativa utveckling nått innan den enspråkiga lexikografin kom igång. Jag har i min tidigare nämnda avhandling (Hannesdóttir 1998) hävdad att svenskans lexikaliska beskrivning i allt väsentligt hade funnit sin form redan innan Dalins ordbok gavs ut. De stora, tvåspråkiga ordböcker som hade svenska som källspråk utgjorde kanske de viktigaste redskapen i den språkplaneringsprocess som för svenskans del kom igång på allvar under 1700-talet. I dessa ordböcker etablerades ett ordförråd som kunde sägas vara ett sanktionerat svenskt ordförråd, ett ordförråd som i stor utsträckning övertogs av Dalin. Där etablerades också en modell för beskrivningen av enskilda svenska ords grammatiska, syntaktiska och pragmatiska egenskaper.

Inom den tvåspråkiga lexikografin utfördes och utvecklades sålunda de båda delprocesser inom språkplaneringen som Einar Haugen har kallat *leksisering* respektive *grammatisering* (Haugen 1990:54; se också Hannesdóttir 2000). I de tvåspråkiga ordböckerna lämnades anvisningar och rekommendationer beträffande ordens användning. Ortografiska rekommendationer förekommer, både direkta, som när Abraham Sahlstedt i sin svensk-latinska *Swensk ordbok* (1773) förser uppslagsordet *Feck* med upplysningen "*perperam pro Fick*", och indirekta, som när han hänvisar läsaren från en form av ett uppslagsord till en annan form av ordet, t.ex. "**Fjäl**, *vide Fjärd*" och "**Fäja**, *vide Feja*" (Sahlstedt 1773, s.v. *feck, fjäl, fäja*). Också böjningsanvisningar är vanliga; såsom framgår av exempel (1) och (2) nedan kan de olika formerna redovisas noggrant. Även brukskommentarer var naturliga inslag i de stora tvåspråkiga ordböckerna (exempel 3).

- (1) Få, v. *anom.* jag får, vi få, j fån, de få, *pr.* jag fick, vi finge, de fingo, *imp.* fått, *sup.* (Möller 1790, s.v. *få*)
- (2) Gå, v. n. & a. 4. (*impf. indic.* Gick, *pl.* Gingo, *impf. conj.* Ginge; *pres. conj.* Gånge; *imper. sing.* Gack el. Gå, *pl.* Gån; *sup.* Gått; *p. p.* Gången) [---]. (*Parallèle des langues Française & Suédoise* (=Weste), del 3, 1807, s.v. *gå*)
- (3) GA`DDA sig tillsammans, bättre Sammangadda sig. GA`NTAS, v. d. I. (*pop. och mind. brukl.*) *Folåtrer*; *elj.* Galnas. Se äfv. Hångla. (*Parallèle* [...] del 3, 1807, s.v. *gadda, gantas*)

Vissa av de funktioner som idag enbart förknippas med den enspråkiga ordboken har således tidigare uppburits av de tvåspråkiga ordböckerna. Tämmligen kompletta, formaliserade grammatiska uppgifter om ordklassstillhörighet, substantivens deklinationstillhörighet, verbens konjugationstillhörighet samt möjligheten att använda dem i passiv form finns redan hos Sahlstedt (exempel 4).

- (4) **Bi**, N. 4. [...] *Def.* Biet. *Plur.* Bien. *Def.* Binen, *conf.* Humla.
Brand, M. 2. Titio. *Plur.* Brandar non Bränder, *conf.* Eldebrand.
Bringa, F. I.
Bringa, V. A. 3. *Ir.* [...] *Pr.* Bringar. *Ipf.* Bragte, *Pf.* Bragt, Bragt.
Brinna, V. N. 4. [...] *Pr.* Brinner. *Ipf.* Brann. *Pf.* Brunnit. Brunnen., *p. p.* *Def. & Plur.* Brundne. (Sahlstedt 1773, s.v. *bi, brand, bringa* (f.), *bringa* (v.), *brinna*)

Den tyske lexikografen Möller är den förste som i någon större utsträckning indelar de svenska uppslagsordens betydelser i numrerade moment (exempel 5). Det är också i en tvåspråkig ordbok som systematiska anvisningar för hur de svenska uppslagsorden skulle uttalas först förekommer. I Westes svensk-franska från 1807 markeras accenten för alla två- och flerstaviga ord, och också andra uttalsdrag kommenteras (exempel 6).

- (5) Kall, adj. kallt, adv. 1) kalt, der Wärme beraubt. 2) fig. aller lebhaften Empfindungen beraubt, 3) kaltsinnig.
 Panna, s.f. I. 1) eine Pfanne [...] 2) die Stirne.
 (Möller 1790, s.v. *kall, panna*)
- (6) E`NKEL (l. ängkel) adj. [...]
 E`NSAM, (sl[utet] e) adj. [...]
 (*Parallèle* [...] del 3, 1807, s.v. *enkel, ensam*)

I sin ordbok utnyttjar också Dalin i allt väsentligt den redan utvecklade modellen för att beskriva uppslagsordens formella egenskaper.

Det drag som av naturliga skäl var minst utvecklat i de tvåspråkiga ordböcker där svenskan är källspråk var betydelsebeskrivningen, dvs. det för den enspråkiga lexikografin specifika draget att beskriva uppslagsordens betydelse i form av definitioner eller parafrafer i stället för att bara redovisa en eller flera ekvivalenter, det senare en karakteristisk egenskap som räknas som konstituerande för den tvåspråkiga lexikografin. Denna fördelning av olika typer av semantisk specifikation på de olika lexikografiska grenarna är emellertid långt ifrån någon fullständig dikotomi. Såväl klassiska aristoteliska definitioner som parafrastiska förklaringar är ganska vanliga i tvåspråkiga ordböcker (exempel 7). På samma sätt har närsynonymer och ibland räckor av närsynonymer varit ett vanligt sätt att behandla uppslagsordets innehållssida också i enspråkiga ordböcker, här illustrerat av artiklarna *kitslig* och *geschäft* (exempel 8) i *Illustrerad svensk ordbok* (1985). På denna punkt är det emellertid inte självklart att Dahlin har haft samma hjälp av de tvåspråkiga föregångarna som då det gällt beskrivningen av uppslagsordens formella egenskaper.

- (7) **gubbe** *-en -ar* 1 pers. old man
lagom **I** *adv* just right; nog just enough; tillräckligt sufficiently; måttligt in moderation, moderately. (*Svensk-engelsk ordbok*, 1968, s.v. *gubbe*, *lagom*)
- (8) **kitslig** *-t, -are* snarstucken, lättretad, retsam, förtretlig, förarglig, nedrig, småaktig; svår, besvärlig *det var en k. historia*. **kitslighet** *-en, -er* snarstuckenhet, småaktighet, retsamhet; retsam el. småaktig kritik el. anmärkning.
geschäft **t** [*g-*, äv. *j-*] *-et, =*, vanl. neds.: affär, näringsfång, sysselsättning. (*Illustrerad svensk ordbok*, 3 uppl. 6 tr. 1985, s.v. *kitslig*, *geschäft*)

Dalins ordbok kan på sätt och vis betraktas som höjdpunkten i den tvåspråkiga svenska lexikografins utveckling. Från 1850-talet och till den korpusbaserade lingvistikens inträde i den tvåspråkiga lexikografin har teoriutvecklingen på området varit blygsam. Därför är det välkommet med det intresse som nu riktas

mot denna gren av verksamheten. Det är dock knappast troligt att de tvåspråkiga ordböckerna någonsin kommer att återerövra alla de funktioner de fyllde under det sena 1700-talet och fram till 1800-talets mitt. De kommer sannolikt inte att konkurrera med eller komplettera SAOL i funktionen som lexikografiskt hjälpmedel i fråga om stavning, avstavning och böjning av svenska ord. Däremot skulle det kontrastiva perspektiv som definitionsmässigt borde vara grundläggande för den tvåspråkiga lexikografin kunna bidra till observationer som på olika sätt kunde föra den lexikaliska beskrivningen av svenskan framåt. Ett naturligt incitament till en sådan utveckling skulle den intensifierade översättningsverksamhet som präglat den senaste tiden kunna vara. Språkliga drag som i ett språkinternt perspektiv framstår som regelbundna och oproblematiska avslöjas ofta i det kontrastiva perspektivet som oregelbundna och intressanta. Genom att tillhandahålla nya infallsvinklar på det enskilda språket skulle den tvåspråkiga lexikografin åter kunna bidra till att föra språkbeskrivningen framåt.

Det kontrastiva perspektiv som borde genomsyra den tvåspråkiga lexikografin är för all del olika väl genomfört i de äldre tvåspråkiga ordböckerna. I de flesta fall sträcker det sig inte så mycket längre än till makronivån, dvs. själva ordförrådet och dess representation på målspråket. De som bäst har utnyttjat den kontrastiva analysen inom den svenska lexikografin är faktiskt Dalins två föregångare inom den enspråkiga lexikografin: Kindblad och Almqvist. Kindblad tycks vara den som i Sverige introducerar språkhistorisk forskning av Rasks och Grimms märke, dvs. baserad på genetisk språksläktskap. I sin ordbok genomför han inte bara det historiskt-komparativa synsättet utan också ett mera allmänt kontrastivt perspektiv som anläggs i stort och smått. Någon ortodox komparatist var Kindblad dock knappast. Till skillnad från Grimm och många av hans meningsfränder såg han inte språkförändringar som ett successivt sönderfall och en förskämning av språket utan som en "oupphörigen fortgående förnygring" av språket (Kindblad 1840[-41]:XXXIX).

Kindblads kontrastiva perspektiv framträder klart i den synkrona beskrivningen, både beträffande ordens bruk och när det

gäller beskrivningen av semantiskt närbesläktade ords betydelse. I själva ordboksdelen av fragmentet, som tyvärr bara omfattar avsnittet *a-afhackning*, (totalt knappt 700 ord), relateras exempelvis främmande ord ofta till en möjlig svensk motsvarighet, såsom sker i artikeln *accentuera* (exempel 9). Där påpekar Kindblad också att intonation och betoning inte är samma sak.

- (9) **Accentuera.** V. Akt. – 2. 4. Språkläran. a) Lägga tonvigt på en eller flera stafvelser i ett ord. Till ett riktigt uttal hörer att rätt accentuera hvarje stafvelse. De Svenska orden accentueras vanligen felaktigt af utlänningar. Uti en accentuerad, det är lång stafvelse, kan sjelfva vokalen vara antingen lång eller kort. SVENSKA AKADEMIEN.
 b) Uttala vissa ord i en mening med en särdeles tonvigt. Accentuera de ord, hvarpå meningens vigt hvilat.
 c) Använda tecken, i de språk där sådana finnas, till ljudformens utmärkande. Då man skrifver Franska, bör man vara uppmärksam på att rätt accentuera orden.
 Något svenskt ord finnes ej som, med full motsvarighet, kunde användas i stället för det främmande accentuera; man brukar väl ibland, för de tvänne första betydelserna, säga Betona stafvelser, Betona ord, men detta är mindre egentligt, ty i dessa fall är det ej Tonen man utmärker, emedan alla stafvelser och ord gemenligen hafva en sådan, utan Tonvigten.
 Läs Acksänntuera. (Kindblad 1840[–41], s.v. accentuera)

Ett främmande ord kan också, såsom ordet *affekt* (exempel 10), relateras till ett svenskt ord, *sinnerörelse*, eller till ett annat främmande ord, *passion*.

- (10) **Affekt.** Subst. M. (Flt–er) – 2. En plötsigt uppstående sinnesörelse, såsom: Frukant, Ångest, Blygsel, Harm, Vrede, Glädje, Sorg m. m. Stundom betecknas äfven med Affekt: Benägenheten för häftiga sinnesörelser. *Bringa någon i affekt. Råka, komma i affekt. En våldsamt affekt. Bruket af affekternas språk är den sanna vältalarens seger.* LEOPOLD. *Vara utan affekt, full med affekt.* Ordet brukas ganska allmänt, men ersättes fullkomligt af det inhemska *Sinnerörelse*. Skillnaden emellan orden *Affekt* och *Passion* är den, att med det första blott menas En tillfälligt uppbrusande känsla, men med det sednare En fortfarande begärelse, åtrå. [...] (Kindblad 1840[–41], s.v. *affekt*)

Också svenska närsynonymers betydelse och bruk preciseras genom att de kontrasteras med varandra. Definitionen till uppslagsordet *affjädra* formulerar Kindblad som ”Borttrycka fjäderna af en lefvande fågel” och ordets användning illustreras med

språkprovet ”Den skamlösa höken affjädrar hönsen inför allas ögon”, hämtat ur Leopolds skrifter. Därefter kompletteras informationen ytterligare genom att ordet relateras till en närsynonym: ”I fråga om döda fåglar säges *Plocka*” (Kindblad 1840[–41], s.v. *affjädra*). Ibland innehåller artiklarna reflektioner kring möjliga semantiska distinktioner snarare än direkta rekommendationer. I exempel (11) presenterar Kindblad en utredning av vad orden *ackuschörska*, *barmorska* och *barnförlossarinna* har gemensamt och vad det är som skiljer dem åt.

- (11) **Ackuschörska.** Subst. F. (Flt-or) –1. En i barnförlossningskonsten inlörd och densamma utöfvande qvinna. Vara en erfaren ackuschörska. Efterfölja ackuschörskans föreskrifter. Rådfråga ackuschörskan, flera ackuschörskor.
Som vårt modersmål till detta begrepps uttryckande redan förut har, dels det urgamla, af ett forntida bruk härledda ordet Jordgumma, dels ordet Barmorska, som numera nästan uteslutande brukas, dels slutligen ordet Barnförlossarinna, utbildadt efter språkets lagar, torde ifrågasvarande utländska benämning kunna alldeles umbäras. Några anse väl denna böra användas till betecknande af en person, som inhämtat grundligare vetenskapliga kunskaper, då åter en qvinna som hufvudsakligen grundade sin barnförlossningsskicklighet på erfarenhet och vana, borde benämnas Barmorska; men om äfven en sådan begreppsskillnad skall iakttagas, är benämningen Barnförlossarinna i förra fallet användbar. (Kindblad 1840[–41], s.v. *ackuschörska*)

Också Almqvist anlägger ett kontrastivt perspektiv i sin *Ordbok öfver svenska språket* (1842–[44]). Denna omfattar en något större del av ordförrådet än Kindblads fragment, eller *abrambär*. Själva betydelsebeskrivningen tar ofta sin utgångspunkt i ordets konstruktionssätt och den kan preciseras genom en jämförelse med betydelsen hos en närsynonym. Ibland redovisas noggranna utredningar om uppslagsordets betydelse och valens relaterat till motsvarande egenskaper hos en närsynonym. I artikeln *afflå* (exempel 12) jämförs användningen av uppslagsordet *afflå* med användningen av simplex verbet *flå*.

- (12) **Afflå**, *u.* ävflå, *v. a.* 4. Ss. — Taga huden, skinnet o. s. v. af en kropp, död eller lefvande. Verbets objekt är det, som borttages, t. ex. ”afflå skinnet på en katt.” Någon gång finner man saken, hvarifrån, satt till objekt; t. ex. ”afflå en häst.” Men det vanligaste är, att nyttja det enkla *Flå*, då man vill tala om saken, hvarifrån, såsom objektet. I frasen: ”afflå hästen huden,” är huden det verkliga objektet och hästen står ablativt,

elliptiskt, för "från hästen." Afflå kan utbytas emot *Flå af*. (C.J.L. Almqvist 1842[-44], s.v. *afflå*)

Skillnaden mellan verben *afflå* och *flå* framgår tydligast av respektive verbs lexikaliska selektionsrestriktioner. *Afflå* konstrueras vanligen med objekt som *hud* eller *skinn*, medan *flå* konstrueras med ett objekt som betecknar den varelse som blir av med huden. På liknande sätt förklaras skillnaden mellan när-synonymerna *anlete* och *ansikte* (exempel 13) genom att ordens bruk kontrasteras mot varandra. Båda orden kan också konstrueras med verbet *skina*, men det är en klar skillnad om det är *anletet* eller *ansiktet* som skiner.

- (13) **Anlete**, *u.* ànnléte, *s. n.* 4. — Ansigte; framdelen af hufvudet. Ordet Anlete begagnas i högre stil än Ansigte. Figurligt förekommer det i ordasättet: *Arbeta i sin anletes svett*, strängt och oaflåtligt arbeta. **Ansigte**, *u.* ànnsi'kte, *s. n.* 4. — 1) Anlete, främre delen af hufvudet. — 2) Utseende, anseende. Ordet Ansigte brukas i ungångsspråket hellre än Anlete, hvilket tillhör den mera upplyftade stilen. Således säger man: "hvad har du fått i ansigtet?" "jag har rosen i ansigtet," men icke: i anletet. Deremot säger man: "hans anlete skiner," och man tänker dervid på en gloria; men vid uttrycket "hans ansigte skiner," tänker man på flott, svett eller följderna af ett rus. — *Se någon under ansigtet*, betrakta utan fruktan, är ett äldre talesätt. (C.J.L. Almqvist, 1842[-44], s.v. *anlete, ansigte*)

Också rent grammatiska skillnader beträffande närsynonymers fraseologi kommenteras (exempel 14). Enheterna *beggedera* och *begge två* är i det närmaste synonyma, av numeruskongruensen framgår emellertid att *beggedera*, till skillnad från *begge två*, refererar till **var och en** av de två men inte de båda sammantaget, som dualbegrepp.

- (14) **Beggedera**, *u.* b`äggedéra, *adj. sing.* — Hvardera af två. Ordet är nära liktydigt med *Begge två*, men konstrueras med verb i singularis, t. ex. "beggedera har orätt." (C.J.L. Almqvist, 1842[-44], s.v. *beggedera*)

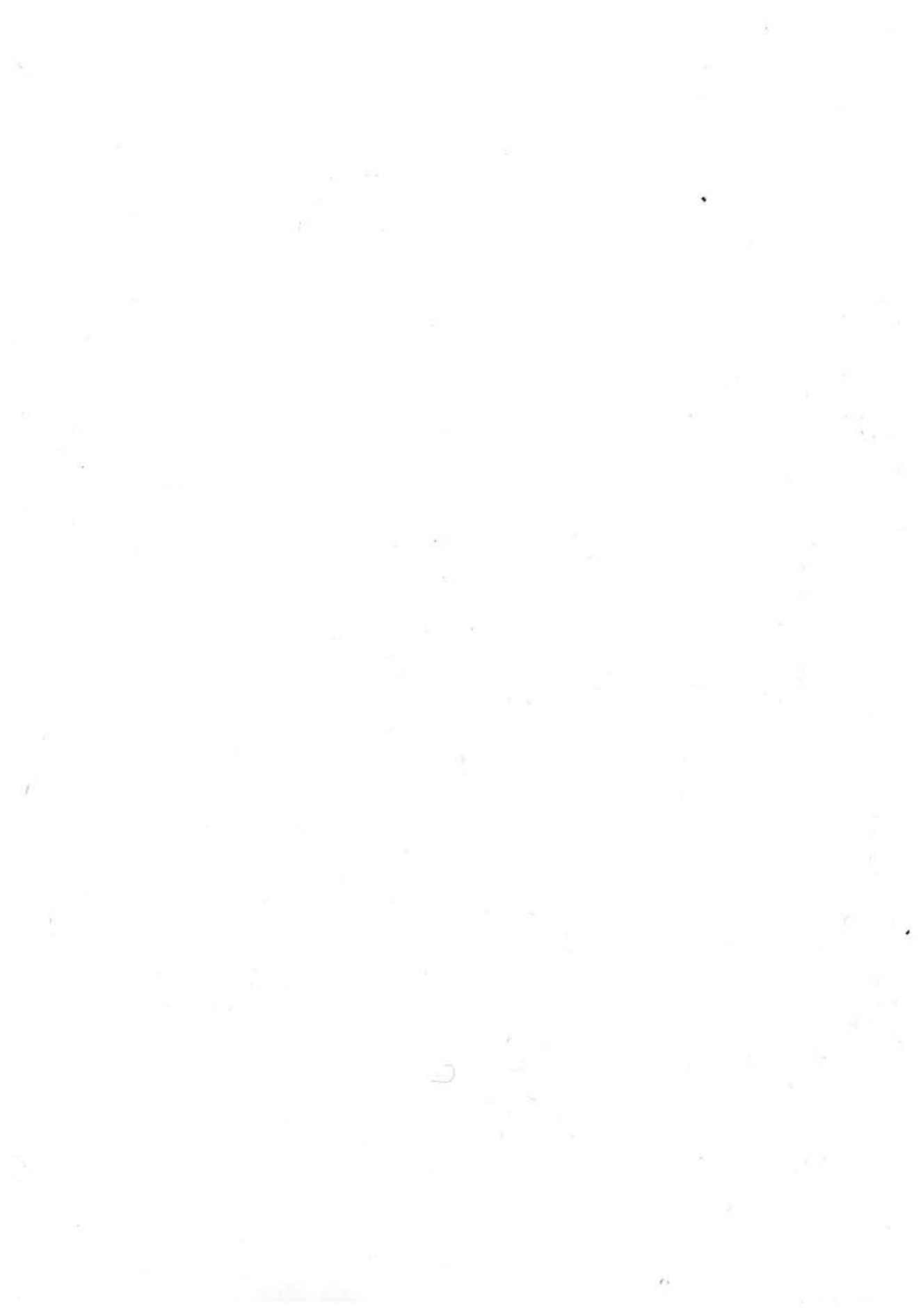
Av exempel (15) framgår hur Almqvist redovisar de betydelse-relationer som kan urskiljas inom ett semantiskt fält. Där relateras betydelsen av uppslagsordet *begjuta* till fältet 'göra våt'.

- (15) **Begjuta**, u. bejüta, v. a. 5. Böjningsformen ses under *Gjuta*. — Slå vått (vatten, vin, olja, mjölk o. s. v.) öfver något. Begjuta utmärker en starkare vätning, än som sker genom att *Fukta*; men en mindre, än som åstadkommes genom att *Öfverskölja* eller *Öfverspruta*; och ännu mindre, än genom att *Indränka*. Begjuta skiljer sig ifrån *Bespruta* deruti, att vätningen försiggår på ett mildare och saktare sätt. Konstrueras genom *med*. Verbets objekt är aldrig det våta sjelf, utan föremålet, hvarpå det våta hålles. (C.J.L. Almqvist, 1842[-44], s.v. *begjuta*)

När det gäller betydelsebeskrivningen och även betydelseindelningen använder Dalin, i arbetet med sin egen ordbok, naturligtvis de enspråkiga ordboksfragmenten – så långt de räckte. Av fragmentens kontrastiva perspektiv märks emellertid ingenting. Hans artiklar utgår strikt ifrån det enskilda uppslagsordet, som behandlas som en isolerad lexikalisk enhet. Konstruktionsuppgifter förekommer förstås, men där det finns alternativa konstruktionssätt redovisas de i bästa fall okommenterade som språkprov och de olika konstruktionernas semantiska relevans kommenteras sällan. Också betydelsebeskrivningen är begränsad till det aktuella uppslagsordet. Enstaka exempel kan dock uppbringas, som exempelvis en kommentar som kompletterar definitionen av uppslagsordet *enfaldig* (exempel 16). Sådana kommentarer är dock mycket sällsynta i Dalins ordbok.

- (16) **ENFALDIG**, ònfåldigg o. ènnfåldigg, a. 2. 1) Som har svag omdömeskraft (dock i mindre grad än *Dum*). – Syn. Se *Dum*. [---] (A.F. Dalin, 1850–1855, s.v. *enfaldig*)

Det är lätt att visa hur den första enspråkiga svenska ordboken, Dalins *Ordbok öfver svenska språket*, bygger på den lexikografiska tradition som utvecklats och konsoliderats inom den tvåspråkiga lexikografien. Verket bryter inte på något påtagligt sätt med den modell för lexikalisk beskrivning som etablerats i de tvåspråkiga ordböckerna. Dalin är emellertid inte i alla avseenden hänvisad till sina tidigare kollegers verk. Han besatt som sagt själv en hel del erfarenhet av arbete med tvåspråkiga ordböcker innan han påbörjade den enspråkiga ordboken. Det är väl kartlagt att han faktiskt utnyttjar såväl metoder som material från sin stora fransk-svenska ordbok i arbetet med den svenska ordboken. Främst är det naturligtvis artiklarnas innehållsdel som



perspektiv hos Dalin – ett drag som paradoxalt nog till en stor del skulle kunna tillskrivas Dalins beroende av den äldre tvåspråkiga lexikografin.

Vad som här är viktigt är att det inom det tidiga 1800-talets avancerade lexikografi inte nödvändigtvis behövde föreligga någon större skillnad mellan de enspråkiga och de tvåspråkiga verken i deras uppläggning. Det är därför inget anmärkningsvärt att en enspråkig ordbok kunde bygga på tvåspråkiga arbeten; eller att en enspråkig ordbok över ett språk kunde tas som utgångspunkt för en enspråkig ordbok över ett annat språk. Där emot kan man naturligtvis urskilja olika lexikografiska traditioner, men det är inte självklart att skiljelinjen mellan dessa gick mellan just enspråkiga och tvåspråkiga verk.

Något som slår åtminstone mig som tvåspråkig användare av dagens enspråkiga svenska ordböcker är att betydelseaspekten där tilldelats så stort utrymme. Ordboken är det enda forum där det enskilda ordets alla egenskaper kan behandlas utförligt. Är det givet att en svensk modersmålstalare har större svårigheter med ordens betydelse än med andra aspekter av ordets användning som till exempel ordets konstruktionsmönster och dess syntaktiska omgivningar?

De båda monumentalt anlagda men oavslutade enspråkiga fragmenten från 1840-talet representerar en lexikalisk syn som inte har manifesterats vidare inom den enspråkiga svenska lexikografin. Dalin däremot fortsatte den tradition som etablerats i de tvåspråkiga ordböckerna, och det är i allt väsentligt denna tradition som fortfarande är tongivande inom den enspråkiga lexikografin i Sverige. Med det nyväckta intresset för tvåspråkig lexikografi är det kanske dags för en alternativ, kontrastivt baserad, enspråkig lexikografi?

Litteratur

Almqvist, C.J.L. 1842[–44]. *Ordbok öfver Svenska Språket i dess närvarande skick*. Örebro: N.M. Lindhs Boktryckeri.

- Dalin, A.F. 1850–53[–55]. *Ordbok öfver svenska språket*. Stockholm: Författarens förlag.
- Fjeld, Ruth Vatvedt. 1992. "Innledning." *Nordiske studier i leksikografi. Rapport fra Konferanse om leksikografi i Norden 28.-31. mai 1991. (Skrifter utgitt av nordisk forening for leksikografi. Skrift nr. 1.)* Oslo. S. 1–6.
- Hannesdóttir, Anna Helga. 1991. "Relationer mellan Dalins olika ordböcker." *Studier i svensk språkhistoria* 2. Utg. av Sven-Göran Malmgren och Bo Ralph. (*Acta universitatis Gothoburgensis: Nordistica Gothoburgensia*. 14.) Göteborg: Göteborgs universitet. S. 78–89.
- Hannesdóttir, Anna Helga. 1998. *Lexikografihistorisk spegel. Den enspråkiga svenska lexikografins utveckling ur den tvåspråkiga. (Meijerbergs arkiv för svensk ordforskning. 23.)* Göteborg.
- Hannesdóttir, Anna Helga. 2000. "Tvåspråkig lexikografi och språkplanering i 1700-talets Sverige." *Studier i svensk språkhistoria*. 5. Utgivna av L.-E. Edlund. (*Nordsvenska*. 11.) Umeå: Umeå universitet. S. 224–235.
- Haugen, Einar. 1987. *Babels förbrødring*. Oslo: Universitetsforlaget.
- Illustrerad svensk ordbok*. [1955] 1964. Stockholm: Natur och kultur.
- Kindblad, Karl Eduard. 1840[–41]. *Ordbok öfver Swenska Språket*. Stockholm: P.A. Norstedt & Söner.
- Kromann, H.-P., Th. Riiber & P. Rosbach. 1991. "Principles of Bilingual Lexicography." *Wörterbücher/Dictionaries/ Dictionnaires. Ein internationales Handbuch zur Lexikographie*. 1989–91. Hrsg. von H. Steger & H.E. Wiegand. (*Handbücher zur Sprach- und Kommunikationswissenschaft*. Band 5.3.) Berlin & New York: Walter de Gruyter. S. 2711–2728.
- Malmgren, Sven-Göran. 1988. "Almqvist, Dalin och den svenska definitionsordbokens födelse." *Studier i svensk språkhistoria*. Red. G. Pettersson. (*Lundastudier i nordisk språkvetenskap*. A 41.) Lund: Lund University Press. S. 195–213.
- Möller, J.G.P. 1790. *Tysk och Svensk samt Svensk och Tysk Ordbok*. Tredje delen. Greifswald: A.F. Röse.

- Noreen, Adolf. 1903. *Vårt språk. Nysvensk grammatik i utförlig framställning*. Första bandet. Lund: C.W.K. Gleerups förlag.
- Sahlstedt, Abraham M. 1773. *Swensk ordbok*. Stockholm: C. Stolpe.
- Svensk-engelsk ordbok*. 1968. Stockholm: Esselte Studium.
- Weste E.W. 1807. *Svenskt och fransyskt lexicon. (Parallèle des langues Française & Suédoise. Tome III-IV.)* Stockholm: J.C. Holmberg.

Maria Arnstad

Idiombruk i talspråk

This article is a study of the use of idioms in spoken Swedish. The study is based on a corpus, consisting of fifteen transcribed interviews with three native Swedish university students. The aim of the study is to explore the kinds of idioms used in the material, examine their forms, and compare them to their corresponding forms in a number of standard Swedish dictionaries. The comparison serves to observe the possible differences between the idiom constructions represented in the dictionaries and the forms actually used in speech. In addition to this, the idioms in the material are compared with the findings of another study, with the same purpose, but concerning the use of idioms in written text (Clausén, 1996). In this article, I put forward two interrelated hypotheses: 1. The idioms used in written text are more varied than the idioms used in speech, in respect to the registered dictionary forms. 2. The idioms used in written text are of a different kind than the idioms used in speech. The first hypothesis is confirmed in my material, while the second is confirmed only to a certain extent.

För att ha hopp om att förstå varandra, måste vi dock tala i liknelser, i bilder. Det gör vi alltid till vardags i de enklaste sammanhang, och det måste vi göra nu också, då vi gemensamt ska försöka lära oss något om verkligheten.

(Ur Monika Fagerholm 1994: *Underbara kvinnor vid vatten*. Albert Bonnier Förlag.)

1. Inledning

Min undersökning om idiombruk i talspråk, som presenteras här, kan betraktas som en utvidgning av Ulla Clauséns studie om idiombruk på ledarsidor (Clausén, 1996), baserad på Dagens Nyheters ledarsidor under en vecka 1995. Clauséns studie ägnas åt att iaktta vilken typ av idiom som förekommer i löpande skriven text, undersöka vilken form idiomen har och jämföra den med den form som registrerats i "gängse ordböcker". Jämförelsen tjänar bland annat till att ta reda på om ordboksartiklarna speglar det egentliga bruket, och till att se hur idiom kan varieras i en skriven text. Detta syfte har jag "ärvt" i min egen studie, men nu överfört till talspråk. Ovanpå detta har jag även haft som

syfte att iaktta om, och i så fall hur, idiombruk och idiomvariation i talat språk skiljer sig från idiombruk och idiomvariation i skrift.

Just nu ingår jag, liksom Ulla Clausén, Erika Lyly och Ulrika Kvist Darnell, i redaktionen för en ordbok med arbetsnamnet Svensk konstruktionsordbok. Det är en fraseologisk ordbok, som ska behandla fraser och konstruktioner i svenskan. Ordboken tar bland annat upp *idiom* och *klichéer* som egna kategorier. Dessa frastyper definieras nedan (se avsnitt 4 och 5). De två typerna är ofta mycket svåra att skilja från varandra, eftersom ett idiom i många fall *används och fungerar* som en kliché (se vidare avsnitt 5). Men i de fall då ett idiom används som kliché har vi i konstruktionsordboken också alltid klassificerat det som en kliché. Det kan dock vara på sin plats att redan nu poängtera att jag i den här artikeln behandlar idiom. Jag har emellertid kommenterat de fall då idiomerna i exemplen fungerar som klichéer.

Exempel ifrån talspråk har vidare generellt varit kraftigt underrepresenterade i ordböcker, och är det än så länge även i konstruktionsordboken. Bland annat därför har jag särskilt intresserat mig för just detta område.

Ulla Clauséns resultat

Ulla Clausén har kommit fram till att de idiom som används i hennes material vanligen förekommer i en annan form än den etablerade ordboksformen. Bland annat har vissa ord (ofta verb) ersatts med ord från samma semantiska fält. Exempel:

hugga sig själv i svansen
(ordboksform: *bita sig själv i svansen* 'misslyckas i sina angrepp
så att man själv drabbas')
DN 12/3 1995

Ibland har idiomerna utökats med ett eller flera adjektiv. Exempel:

öppna en ny dörr till ngt
(ordboksform: *öppna en dörr till ngt* 'ge möjlighet till ngt')
DN 13/3 1995

eller ett substantiviskt sammansättningsled. Exempel:

hålla ngn på fredsmattan
(ordboksform: *hålla ngn på mattan* 'kuva ngn')
DN 14/3 1995

Idiomen anpassas på detta sätt till den specifika situationen och inlemmas i den övriga texten. Vidare har Clausén funnit att idiom av den typ som har en konkret och produktiv parallellfras är mycket vanliga på ledarsidorna. Dessa idiom används och varieras på ett intrikat sätt, där skribenten ofta levandegör både den konkreta och den bildliga betydelsen av frasen, t.ex. genom att bygga upp en hel situation med pragmatiskt relaterade begrepp. På så vis kan skribenten "spela på en mångtydighet och glidning mellan det konkreta och det bildliga".

2. Antaganden

I min egen idiomstudie har jag utgått ifrån två sinsemellan relaterade hypoteser:

1. Variationen av idiom är större i skrift än i tal.
2. Den typ av idiom som förekommer i skrift skiljer sig från den typ som förekommer i tal.

Jag tror att variationen av idiom är större i skrift än i tal bland annat därför att skribenten vanligen har mer tid att formulera (och omformulera) sig än talaren. I de flesta talsituationer är förutsättningarna för innovativa formuleringar begränsade, eftersom de kräver eftertanke. Och om ett yttrande dröjer för länge rubbas samtalsrytm.

De "färdiga" enheter som finns att tillgå i språket, bland annat de lexikaliserade, icke-varierade idiomerna, till exempel "ta sig vatten över huvudet" och "ta ngt med en nypa salt", ligger däremot nära till hands för talaren. Dessa utgör mallar, med tydlig pragmatisk funktion och ofta med en vidhängande attityd. Med de lexikaliserade idiomens hjälp kan talaren enkelt förmedla en viss känsla, åsikt, eller inställning till något i sin omvärld. Det blir

ekonomiskt, både vad gäller tid och tankekraft. Idiomen kan också skifta betydelse beroende på sammanhang, där betydelsen "förhandlas fram" i den aktuella situationen.

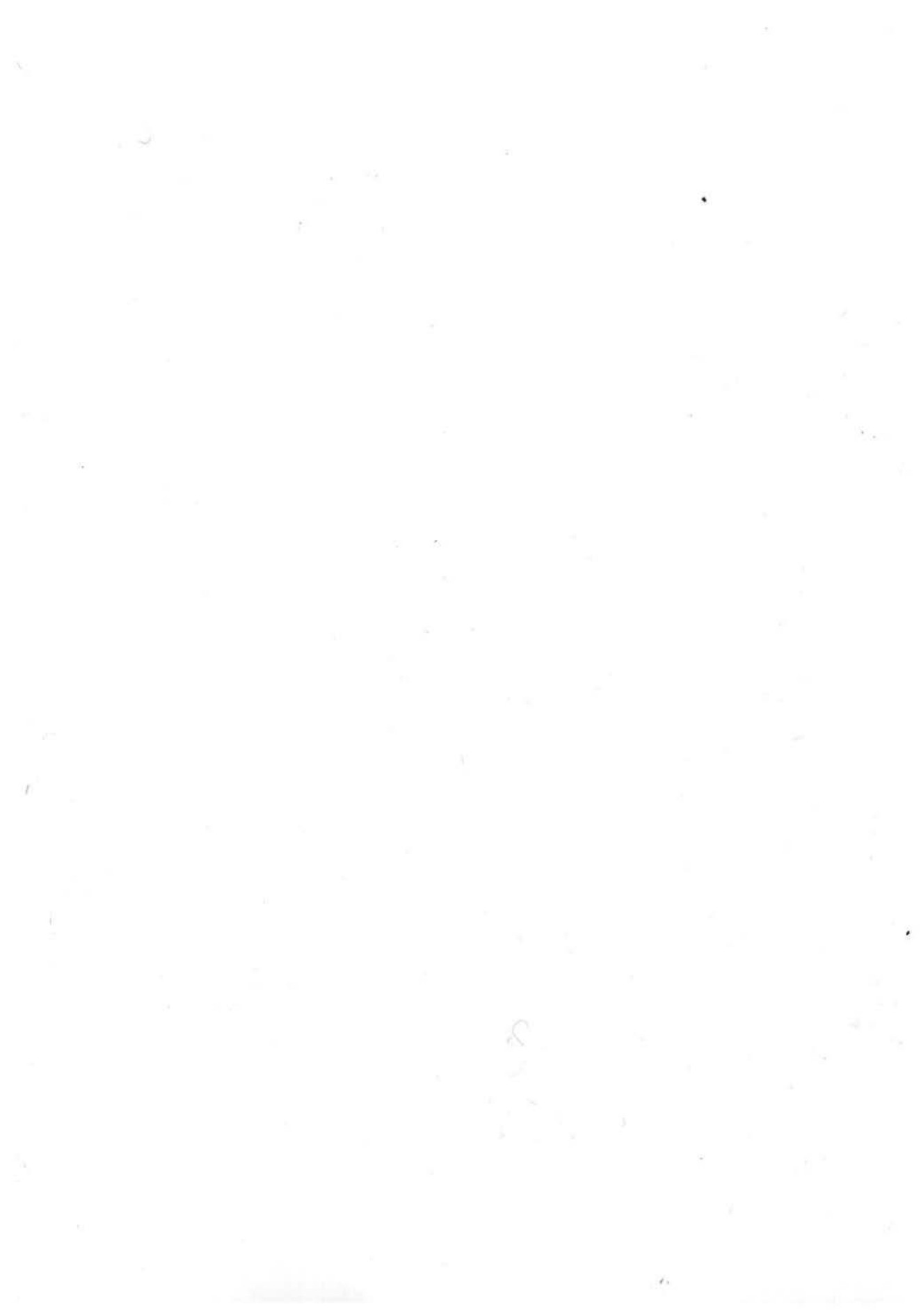
Skribenten, å sin sida, har mer tid på sig och dessutom ofta en vilja att vara originell och kreativ. Detta gäller kanske speciellt på ledarsidor och i liknande sammanhang. I ledarsidesgenren är det viktigt att formulera sitt resonemang på ett slagkraftigt och intelligent sätt, för att nå ut med sitt budskap till läsaren. Därför varierar ofta lexikaliserade idiom, och klichéer undviks, eftersom dessa ofta betraktas som talspråksmässiga.

Den andra hypotesen, att det förekommer andra typer av idiom i text än i tal, följer av den första. En förutsättning för variation av ett idiom är, naturligtvis, att idiomat *går* att variera. På ledarsidorna är, enligt Clauséns resultat, idiom av den typ som har en levande och konkret parallellfras vanlig, troligtvis för att den typen är lätt att variera. I tal, däremot, har jag antagit att det förekommer fler standardiserade idiom, och de behöver således inte *nödvändigtvis* vara av en typ som lätt kan varieras.

3. Definition och användning av idiom

I dag finns många åsikter bland språkvetare om vad ett idiom är. Åsikterna överlappar ibland, och är ibland motstridiga. De flesta är dock överens om att idiom är svårfångade och svårtolkade företeelser i språket. Patrizia Tabossi och Francesco Zardon (1993: 135) hävdar till och med att det troligtvis är en lönlös uppgift att söka definiera dem. Och Rosamund Moon (1998a) använder i sin senaste bok den övergripande termen *FEI*, med betydelsen 'fasta uttryck och idiom', för en mängd olika typer av fasta fraser i språket, såsom fasta kollokationer och talesätt. Därmed undviker hon delvis att definiera idiom som en egen grupp.

Det kan synas vara en enkel och bra lösning att inte särskilja idiomerna från andra typer av fasta fraser. De är ju ändå så svåra att få fatt i; problem medföljer nästan alla försök till beskrivning. Som ordboksredaktör, i synnerhet till en fraseologisk ordbok, blir jag inte desto mindre tvungen att söka hantera och avgränsa



dessa språkliga enheter. För de är ju trots allt mycket speciella, både till sin form och till sin funktion. Chitra Fernando (1996:3) redovisar en lista med de tre oftast omnämnda särdragen som sägs känneteckna ett idiom:

- (1) Compositeness;
- (2) Institutionalization;
- (3) Semantic opacity.

Det första särdraget beskriver idiom som ett uttryck bestående av fler än ett ord, ett sammansatt uttryck. Enligt särdrag nummer två är idiom konventionaliserade, det fasta resultatet av ett uttryck som från början var *ad hoc* och nyskapande på något sätt. Det sista särdraget, som troligtvis också är den mest spridda definitionen, gör gällande att idiom är betydelsemässigt "ogenomskinliga". De är uttryck vars innebörd inte framgår av de ingående ordens betydelser, traditionellt formulerat: "inte summan av sina delar". Nedan redogör jag för den definition av idiom som ligger till grund för min egen undersökning.

Det har även diskuterats i vilken utsträckning idiom egentligen används i skrift respektive tal. Språkforskare har på senare tid noterat att idiom används rikligt i skrift (jfr t.ex. Pawley, 1986, Tannen, 1989 och Clausén, 1996). Vissa, t.ex. John Sinclair i sitt bidrag till denna volym, hävdar däremot, att idiom används mycket sparsamt i talat språk. Sanningen i det påståendet är naturligtvis beroende av hur man definierar termen idiom. Men det har också varit, och är fortfarande, problematiskt att överhuvudtaget undersöka talat språk, eftersom det till skillnad från skrift inte är permanent till sin natur.

4. Definition av idiom i Svensk konstruktionsordbok

Min idiomdefinition i den här artikeln är densamma som finns i modellen för Svensk konstruktionsordbok (se fig. 1 i Ulla Clauséns artikel "Funktion och attityd i Svensk konstruktionsordbok" i denna skrift). Här har idiom fått en ganska vid definition,

där även liknelser och talesätt räknas in, och bildar ett *idiomkluster*. Grunddefinitionen för idiom i konstruktionsordboken är:

- en fras bestående av minst två ord,
- som metaforiserats i sin helhet.

Att idiomerna metaforiserats i sin helhet innebär att de ingående ordens konkreta betydelse "glidit över till en annan nivå" (Clausen, 1996:36), vilket kan jämföras med särdrag nummer tre i avsnitt 2 ovan. Dessa idiom kan vara en verbfras ("få kalla fötter"), en nominalfras ("tidens tand"), ett adverbialt led ("i grevens tid"), ett ordpar ("skrot och korn") eller en hel sats ("en sten har fallit från mitt bröst"). Liknelser och talesätt har båda fått en lite snävare, underordnad definition inom idiomklustret. Liknelser definieras som:

- fraser där två led jämförs (t. ex. "vit som snö" och "ngt går som en dans").

Talesätt definieras som:

- fraser som uttrycker levnadsvisdom (t. ex. "man ska inte kasta ut barnet med badvattnet" och "den som gapar efter mycket mister ofta hela stycket").

5. Definition av klichéer i Svensk konstruktionsordbok

I min studie av idiom i talspråk har jag också tagit med fraser som i ordboken går under benämningen *klichéer*, eftersom dessa i många fall även kan betraktas som idiom. Klichéer ingår, tillsammans med samtalsfraser, i det så kallade *pragmatiska klustret* i modellen för ordboken. De pragmatiska fraserna, klichéerna, är ofta knutna till specifika sociala situationer, och förekommer typiskt i talspråk. Klichéerna för vanligen också någon typ av attityd med sig. Attityden kan exempelvis vara positiv, negativ, ironisk eller mästrande. Klichéer definieras i ordboken som:

- en fras bestående av minst två ord,
- som standardiserats i en stereotyp form.

Klichéerna kan spjälkas upp i ytterligare kategorier beroende på funktion, vilket Ulla Clausén har gjort (Clausén, 1999):

- Klichéer med en *kommenterande funktion*, som Clausén kallar klyschor, tjänar till att markera någons inställning till eller åsikt om något (t.ex. "går det så går det", "man ska inte generalisera", "det är det som är grejen").
- Klichéer med en *strukturerande funktion*, som Clausén kallar textstrukturerande fraser, används för att binda ihop och strukturera ett resonemang (t.ex. "först och främst" och "helt enkelt"). Denna typ är också mycket frekvent i skriftspråk.

I arbetet med ordboken har jag ofta brottats med att kategorisera fraser som antingen idiom eller klichéer. Idiom ska, problematiskt nog, klassificeras som klichéer om de uppfyller de ovan nämnda klichéegenskaperna. Ett visst idiom kan *användas* som kliché. Två exempel är fraserna "på grund av ngn/ngt" och "det är kört". Den förra används som en textstrukturerande kliché och den senare som en kommenterande. Båda är dock också idiom, trots att det bildliga i fraserna har bleknat avsevärt, speciellt i det första fallet.

Dynamiken och glidningen mellan idiom och klichéer är mycket stor, varför det blir svårt, eller omöjligt, att fokusera enbart på den ena företeelsen i en studie av idiombruk. Därför har jag i min undersökning även plockat ut de idiom som används som klichéer.

6. Material

På senare år har det, mycket tack vare Internet, blivit enkelt att få tag på skriftspråkskorpusar av skilda typer och från olika textgenrer, för att genomföra språkvetenskapliga undersökningar. På talspråksområdet är utvecklingen inte lika snabb, och det är fortfarande svårt att få tag på bra korpusar med transkriberat talspråk,

kanske speciellt på andra språk än engelska. Värdet av transkriberat talspråk kan också diskuteras (jfr Moon, 1998b).

Till min undersökning har jag lyckligtvis fått tillgång till en del av en talspråkskorpus, som vid denna tidpunkt inte är helt färdigställd. Korpusen byggs upp inom ramen för projektet "Andra-språkets strukturutveckling" (ASU), som pågår på Institutionen för lingvistik vid Stockholms universitet. Projektet leds av professor Björn Hammarberg. Det material jag fått utgör en del av ett jämförelsematerial till en serie intervjuer med vuxna invandrare, som nyligen kommit till Sverige och håller på att lära sig svenska.

Materialet består av:

- cirka 7800 rader transkriberat tal (max 56 tecken per rad);
- femton intervjuer med tre olika informanter, cirka tjugo till tjugofem år, alla helt svenskspråkiga, universitetsstuderande samt födda och uppvuxna i Stockholmstrakten.

Intervjuerna har skett vid fem olika tillfällen med varje informant. De har utförts av två språkforskare, som båda varit närvarande vid alla tillfällen, och vars tal jag också har analyserat. Intervjuerna är ganska hårt styrda och en rad förutbestämda ämnen avhandlas. Exempelvis får informanterna beskriva vissa föremål, berätta handligen i ett antal seriestrippar, framföra sina tankar kring olika fotografier samt tala om sin bakgrund och sina framtidsplaner.

Nedan följer en presentation av de idiom jag hittat i materialet. Under varje idiom redovisas den motsvarande form som jag funnit i *Nationalencyklopedins ordbok* (1996), *Svensk ordbok* (1987), *Svensk handordbok* (1966) och *Ord för ord* (1960). Idiomens ungefärliga betydelser anges också (ungefärliga, eftersom ett idioms betydelse som sagt kan ändras efter sammanhang). I de fall då idiomerna inte registrerats i någon av de konsulterade ordböckerna har endast betydelsen angivits.

Idiomen återges så som de transkriberats i korpusen, tillsammans med lite av sin kontext. Tecken för paus, tvekan, omtagning etc. har dock utelämnats. Jag har satt idiomerna i kursiv stil.

1. i slutet *drar* artikelförfattarn *en parallell* här till *apor*
dra en parallell till ngt 'jämföra med ngt'
2. *ur moralisk synvinkel* är det väl ganska så förkastlit
ur en viss synvinkel 'enligt ett visst förhållningssätt'
3. ge sej på en annan nation bara *på grund av dom anledningarna*
på grund av ngt 'med anledning av ngt'
4. *han e* ju ganska *svartmålad* i västerländsk media
svartmåla ngn 'framställa ngn på ett negativt sätt'
5. ja tänkte försöka *fortsätta på den banan*
slå in på en viss bana 'välja en viss karriärväg'
6. jo de hänger ju ihop på nåt sätt. *de går nån röd tråd*
den röda tråden 'det genomgående temat'
7. män uppfostras till att *lägga band på känslor*
lägga band på ngt 'behärska ngt'
8. man *sopar undan alla spår* på en gång
sopa igen spåren [efter ngt] 'dölja ngt'
9. ja ja tror att jag börjar. så ser jag om jag förstår *efter hand*
efter hand 'allteftersom'
10. men då tror ja att *de ligger* så här *djupare*
'det är mer grundläggande el. djuplodande'
11. *de finns en djupare dimension*
'det är mer grundläggande el. djuplodande'
12. fast de e väldit otrevligt å börja *tänka i dom termerna* tycker ja
tänka i sådana termer 'tänka på sådant sätt'
13. så skulle man försöka ställa bra frågor också *mellan varven*
mellan varven 'under paus i aktivitet'
14. de handlar om en kommunalpolitiker i karlskrona som vill *ta hårdare tag*
mot nynazismen
ta hårdare tag 'gripa in mer seriöst'
15. annars *e man ju helt körd* idag liksom
'vara helt hopplös'
16. *de e* ju *kört* annars nästan
loppet är kört 'det är hopplöst'
17. då hade vi *trampat i klaveret* förmodlien
trampa i klaveret 'göra bort sig'

18. så han *faller på näsan* när han ställer sej upp
fallalstå på näsan 'trilla omkull'
19. *spelar ljuset en roll?*
ngt spelar [stor] roll 'ngt är av [stor] betydelse'
20. belysningen *spelar otroligt stor roll*
ngt spelar [stor] roll 'ngt är av [stor] betydelse'
21. då e de ingen mening me å *slingra sej runt me långa ord*
svänga sig med ngt 'uppträda skrytsamt med ngt'
22. de kan ja inte *på rak arm*
på rak arm 'oförberett'
23. men än så länge *har du vägvalet framför dej* då?
'inte ha bestämt sig än'
24. så vi får väl *dra ett streck här*
'sluta'
25. ja tror nästan vi får *sätta gränsen där*
'sluta'
26. då *va du inne på köksspåret* här nu
vara inne på ett visst spår 'ha en viss inriktning'
27. man kanske kan *slå två fluger i en smäll*
slå två flugor i en smäll 'nå två syften med samma åtgärd'
28. tycker du de *va ett smart drag* att ha me sej en liten fluga i en tänd-
sticksask
ett smart drag 'en taktisk handling'
29. de kan ju faktiskt *bli ställt på sin spets* ibland
ställa ngt på sin spets 'driva ngt till sin yttersta konsekvens'
30. nu kommer du å *klara de här som en dans*
ngt går som en dans 'ngt går mycket lätt'
31. de verkar som dom som gör de här tar upp ämnet i gymnasiet *blir väldigt bitna*
bli biten [av ngt] 'bli starkt intresserad [av ngt]'

7. Resultat

- Av de 31 olika idiom som används i mitt material har 18 samma form som registrerats i någon av de ordböcker jag kon-

sulterat. Jag har då emellertid bortsett från satsadverbial, som till exempel *väldigt, ju, ganska* och *oerhört*.

- Idiomen i materialet tenderar som helhet att ha en kommenterande eller textstrukturerande funktion, det vill säga: de fungerar som klichéer enligt definitionen i Svensk konstruktionsordboks modell.

- Av de sex idiom i materialet som *inte* har någon motsvarande form i någon av de konsulterade ordböckerna fungerar tre som kommenterande klichéer ("de finns en djupare dimension", "de ligger djupare" och "det är kört"). Två av idiomerna utan ordboksmotsvarigheter fungerar som kommenterande klichéer ("dra ett streck här" och "sätta gränsen där"). Det sjätte idiomet utan motsvarighet ("ha vägvalet framför sig") fungerar som ett typiskt idiom.

- Endast två av fraserna i materialet visar på *tydlig* variation: "slingra sig runt med långa ord" och "vara inne på köksspåret". Dessa båda fraser har en levande och konkret parallellfras, och är därmed också av den typ som är lätt att variera. Det första exemplet har jag dock tolkat som en sammanblandning av fraserna "svänga sig med ngt" och "slingra sig". Denna typ av sammanblandning, så kallad *katakres*, sägs vara vanlig i talspråk. Jag tror att det rör sig om en typ av omedveten, eller kanske hellre ofrivillig, variation av idiomet, som kommer sig av att man inte riktigt minns hur det "ska" låta. Denna typ av variation skiljer sig från den ofta mycket medvetna och stilmässiga variation som förekommer till exempel på ledarsidor.

- Bara en liknelse används ("klara de här som en dans").

8. Slutsatser

Idiomen i materialet rör sig åt klichéhållet i konstruktionsordbokens modell. Många av de idiom som förekommer i materialet

är så pass standardiserade att de knappt ens uppfattas som idiom. De är så gott som omarkerade i sitt sammanhang. Ofta har idiomerna som funktion att strukturera ett resonemang, eller att markera talarens inställning till det sagda på något sätt. Idiomerna fungerar med andra ord vanligen som klichéer, och skulle således också klassificeras som sådana i konstruktionsordboken.

När idiomerna används är de också för det mesta o varierade och väldigt mycket "pang på". De dyker upp i hela stycken. Detta är något som troligen är typiskt för just talspråk, som ju till skillnad från skriven text ofta är fragmentariskt och osammanhängande. Lyssnaren får själv dra inferenser, tolka det som sagts, och försöka knyta det till den aktuella situationen.

På det hela taget kan man säga att idiomerna tenderar att varieras mindre i mitt talspråksmaterial än i Ulla Clauséns skriftspråksmaterial, trots att de ofta är av den typ som lätt kan varieras. De idiom som återfinns i mitt material liknar mer de traditionella ordboksformerna. När det gäller idiombruk är alltså talspråket mer "formellt" och konservativt än skriftspråket, i alla fall om man ser till det material som undersökts här, och om man betraktar gängse ordböcker som de mest formella språkfora. Kanske är det så att de flesta nya idiom uppfins i skrift och inte i tal?

Litteratur

A. ORDBÖCKER

- Nationalencyklopedins ordbok*. 1996. Höganäs: Bra böcker.
Ord för ord. 1960. Stockholm: Norstedts.
Svensk handordbok. 1966. Uppsala: Almqvist och Wiksell.
Svensk ordbok. 1987. Stockholm: Norstedts.

B. ÖVRIG LITTERATUR

- Clausén, U. 1996. Idiom i bruk. *Språket lever! Festskrift till Margareta Westman den 27 mars 1996*. Stockholm: Norstedts. S. 36–42.

- Clausén, U. 1999. "Om pragmatiska fraser i Svensk konstruktionsordbok." *Nordiska studier i lexikografi*. 4. (Skrifter utgivna av Nordiska föreningen för lexikografi. 5.) Helsingfors.
- Fernando, C. 1996. *Idioms and Idiomaticity*. Oxford: Oxford University Press.
- Moon, R. 1998a. *Fixed Expressions and Idioms in English*. Oxford: Clarendon Press.
- Moon, R. 1998b. "On using spoken data in corpus lexicography." *Actes Euralex'98 Proceedings*. Liège. S. 347–355.
- Pawley, A. 1986. Lexicalization. D. Tannen och J.E. Atlatis (red.): *Languages and Linguistics: The Interdependence of Theory, Data and Application. Georgetown University Round Table on Languages and Linguistics 1985*. Washington D.C.: Georgetown University Press.
- Tabossi, P. och F. Zardon, F. 1993. "The Activation of Idiomatic Meaning in Spoken Language Comprehension." C. Cacciari och P. Tabossi (red.): *Idioms: Processing, Structure, and Interpretation*. S. 145.
- Tannen, D. 1989. *Talking voices: Repetition, dialogue and imagery in conversational discourse*. Cambridge/New York: Cambridge University Press.

Antti Arppe

Lärdomar från utveckling av inflekterande synonymordböcker

During 1996–1998, product development projects were carried out at Lingsoft in order to create so-called inflecting thesauri, in which the electronic form of the contents of synonym dictionaries for Swedish, Danish, and Norwegian Bokmål were integrated with computerized morphological models, created according to the two-level model, for the respective languages. The resultant computer programs provided practical insights into the interaction between structures of semantic relations, in this case representing synonymy, and with inflectional morphology. As a result, it became evident that the principle of lexical generality cannot be trusted blindly in generating across the board the inflected forms of the base-form components of a synonym dictionary. Furthermore, it would seem that synonymy between words cannot be categorically expected to extend throughout the entire inflectional paradigm of these words. This would suggest that inflected forms of words should also be considered when constructing structures of semantic relations.

1. Introduktion

Denna artikel bygger på erfarenheter från produktutvecklingsarbete av flera s.k. inflekterande synonymordböcker, som utfördes på Lingsoft 1996–1998. Dessa program utvecklades för de främsta nordiska skriftspråken i ordningen svenska (färdig 1996), danska samt norskt bokmål (1997). Först kunde dessa program bara brukas i Windows-operativsystem, men Macintosh-versioner av alla blev färdiga efteråt (1998). Redan tidigare hade man producerat ett likadant verktyg för finska (1995–1996), och nästan samtidigt med de skandinaviska språken fortsatte vi samma arbete för tyska (1997–1999), men dessa berörs inte i denna artikel.

Här skall jag först beskriva vad en inflekterande synonymordbok är och hur vi utvecklade dessa program. Därefter skall jag berätta om de lingvistiska insikter som vår produktutvecklingsgrupp fick under arbetets gång. Denna artikel bygger på en

tidigare arbetsrapport som jag författat tillsammans med Mari Voipio och Malene Würtz (Arppe, Voipio, Würtz 1999). Därtill är jag tacksam för synpunkter och kommentarer från alla andra som deltog i dessa produktutvecklingsprojekt, särskilt Jussi Birn som deltog i det svenska delprojektet.

2. Inflekterande synonymordböcker och deras utvecklingsarbete

Elektroniska ordböcker av varierande kvalitet och ambition har redan funnits hela 1990-talet för de främsta nordiska skriftspråken. Hittills har nästan alla innehållit ord bara i grundform, och har alltså mer eller mindre varit elektroniska kopior av verk i bokform, med den fördel som man får vid sökning i elektroniska datastrukturer. Om man börjar i en sådan elektronisk synonymordbok med någon böjd form av ett ord, måste man först analysera ordet självt för att mata in dess grundform i programmet. Om man därefter vill få någon synonym i samma böjningsform som det ursprungliga ordet har, måste man göra också detta själv via tangentbordet. En inflekterande synonymordbok har till uppgift att automatisera denna process så att ordboksprogrammet kan behandla inte bara ord i grundform utan också alla deras böjningsformer. Denna egenskap ökar synonymordboksprogrammets användbarhet när man skriver eller redigerar löpande text – nyttigt i alla språk som har något böjningssystem alls, vilket gäller alla nordiska språk och särskilt finska.

Schematiskt sett fungerar en inflekterande synonymordbok på följande sätt:

- I morfologiskt analys av ett böjt ord eller en fras
 (← användaren)
 ⇒ ordets eller frasen grundform eller -former
 ⇒ ordets eller frasens böjningsinformation
 ↓
- II sökning av synonymer från semantiska datastrukturen med
 det ursprungliga ordets grundform (← steg I)
 ⇒ lista av synonymer i grundform
 ↓

- III generering av respektive böjningsformer av synonymorden
 (← steg II) eller -fraser enligt böjningsinformation (← steg I)
 ⇒ lista av synonymer i respektive böjningsform(er)

Som ett konkret exempel på hur detta fungerar i praktiken kan man titta på följande synonymer av det svenska ordet *klassificerades*. Det kan analyseras antingen som ett verb i passiv preteritumform, som t.ex. i *boken klassificerades*, eller som participet i genitivform, som t.ex. i *de klassificerades ordning* – båda av verbet *klassificera*. I många fall är synonymerna organiserade enligt källan i två eller flera grupper, vilket framgår nedan av verbsynonymerna:

- klassificerades
- ⇒ ordnades (verb)
 - delades upp i klasser, indelades, klassades, grupperades,
 - rangordnades, rankades, inrangerades, graderades,
 - kategoriserades, systematiserades
- ⇒ betecknades (verb)
 - bedömdes
- ⇒ betecknades (particip)
 - bedömdas

Som råvara för en sådan inflekterande synonymordbok behöver man två grundkomponenter. Den första är en datalingsvistik morfologisk modell av språket – som i vårt fall var tvånivåmodellen enligt Koskeniemi (Koskeniemi 1983). Den danska tvånivåmodellen hade licensierats tidigare från Thomas Bilgram (Bilgram 1994) och vidareutvecklats på Lingsoft, men modellerna för svenska och norskt bokmål hade utvecklats enbart på Lingsoft (Karlsson 1992, Moshagen 1999). Den andra komponenten är en elektronisk version av en synonymordbok, som vi har licensierat från välkända nordiska förlag utom för finska, som hade utvecklats internt på företaget tidigare. Det svenska synonymordboksinnhållet, med Göran Walter som författare, licensierades från Walters lexikon, vilket material hade publicerats i bokform som *Bonniers synonymordbok* av Bonnier Alba 1995. Det danska synonyminnehållet, med Allan Karker som författare, licensierades från Politikens Forlag, och hade publicerats som *Politikens Synonymordbog* år 1994. Det norska synonyminne-

hållet, med Herbert Svenkerud som författare, licensierades från J.W. Cappelens Förlag, och hade publicerats som *Cappelens Media Synonymordbok* redan år 1983. Alla dessa synonymordböcker skilde sig från varandra i uppläggning, men passade med mindre eller större omarbetning till slutprodukten. Det svenska materialet påminde kanske mest om en traditionell synonymordbok med några tiotusen uppslagsord och en eller flera grupper av synonymer för varje uppslagsord, medan det danska materialet nästan helt saknade gruppering av synonymer, och det norska materialet var grupperat under flera tusen grunduppslagsord, som innehöll långa listor av närmare och avlägsnare synonymer – lyckligtvis semantiskt grupperade.

Utvecklingen av varje inflekterande synonymordbok tog flera månader, av vilka den största delen utgjordes av omarbetning av det ursprungliga materialet, från t.ex. enkla ordbehandlingsfiler, till en systematiskt strukturerad elektronisk form med alla förkortningar utskrivna, och därefter i testandet och finjusteringen av det resulterande programmet. I varje projekt deltog en programmerare, en eller flera lingvister samt en projektchef. Den svenska versionen utvecklades helt under året 1996 och kom ut på marknaden som en färdigintegrerad modul i den svenska versionen av Microsoft Office 97. Både den danska och norska bokmålsversionen utvecklades under 1998 och blev en del av uppgraderingsversionen 97A av respektive danska och norska Microsoft Office 97.

3. Erfarenheter med lingvistisk betydelse

3.1. Lexikal allmängiltighet

Teorin i generativ morfologi påstår att man kan anta att varje ord som hör till en viss ordklass kan böjas i alla former i ordklassens böjningsparadigm. Bl.a. Scalise (1984) säger följande:

It is in general possible to attach to any word the entire set of inflections associated with the word [base form] in question.

Bybee (1985) kallar detta för *lexikal allmängiltighet* (← *lexical generality*). Detta gäller särskilt när man tänker på nya ord, som man kan anta bli böjda på samma sätt som de andra, existerande medlemmarna i samma ordklass som de nya orden, även om dessa inte existerar (dvs. inte har använts) än. T.ex. kan man anta att alla svenska utrumsubstantiv som slutar på *-a* böjs i alla böjningsformer enligt det traditionella mönstret, vad än dessa ord betyder: *en rucka, ruckan, ruckor, ruckorna, ruckas, ruckans, ruckors, ruckornas*. Detta är en egenskap hos de öppna ordklasserna. Däremot gäller lexikal allmängiltighet inte vid derivation. Om man kan derivera ett substantiv som *tävling* från verbet *tävla*, garanterar detta inte att man kan göra likadant med verbet *finna*. I stället för **finning* bör man använda en annan derivationsändelse, t.ex. *finnande*. Från denna synpunkt kan man se datoriserade morfologiska modeller av något språk som en egentlig inkarnation av lexikalisk allmängiltighet, då dessa byggs upp av ett lexikon med ordstammar som klassificeras enligt ordklass, och en regelsamling enligt vilket ord tillhörande en viss ordklass böjs.

På det sättet kan man beskriva en inflekterande synonymordbok som en tillämpning av lexikalisk allmängiltighet över hela innehållet av en synonymordbok. Vad som då händer i stor skala har inte kunnat observeras tidigare, då det ännu inte funnits både datoriserade morfologiska modeller och synonymordböcker i elektronisk form. Vi förväntade oss att en sådan generering av former skulle behöva, lingvistiskt sett, bara minimal korrigering, men realiteten var annorlunda. Det uppstod ett behov av både *morfosemantiska* och *pragmatiska* begränsningar i det ursprungliga lexikonet.

Morfosemantiska begränsningar kallas här de böjningsformer som en infödd språkbrukare helt enkelt tolkar som felaktiga. T.ex. kan en del verb egentligen inte användas i passivform, som svenska **vars* och **blevs* eller norska **gikks*. Därtill kan en del adjektiv inte få både komparativ- eller superlativform, t.ex. danska *kæmpestor*, *skidegod* och *ekstra*, men inte **kæmpestørre*, **skidebedre* och **ekstrare* eller **kæmpestørst*, **skidebedst* och **ekstrast*. Detta är först och främst en semantisk, ordspecifik

bedömning, och därför var vi tvungna att genomgå de ursprungliga lexikonens alla verb och adjektiv för att klassificera varje ordstam i dessa ordklasser i detta avseende (i substantiv kunde man välja ut ett begränsat antal pluralia tantum-fall). Orsaken till detta var antagligen främst att i morfologisk analys, till vilket ändamål våra morfologiska modeller ursprungligen var utvecklade, behöver beskrivningen inte vara så normativt noggrann som vid morfologisk generering, vilken egenskap vi behövde i våra inflekterande synonymordböcker. Så de morfosemantiska begränsningarna var egentligen luckor i de ursprungliga morfologiska modellerna. Men i varje fall gäller sådana noggrannare klassifikationer och böjningsbegränsningar slutligen bara de specifika ord som har klassificerats, inte alla deras möjliga derivationer eller sammansättningar med andra ord.

Behovet av morfosemantiska begränsningar är också en begreppsdefinitionsfråga. Om man räknar som ordklasser bara de knappt tio som man traditionellt lär sig i skolan, substantiv, adjektiv, verb osv., blir man tvungen att efteråt granska genereringen av olika böjningsformer av ord tillhörande varje böjd ordklass, eftersom alla ord i dessa ordklasser inte uppvisar alla böjningsformer. Däremot, om man definierar en ordklass som en grupp ord som har samma böjningsparadigm, med likadana begränsningar eller luckor i paradigmet, och bygger lexikonet enligt denna strängare princip, finns den produktiva begränsningen färdig i lexikonet, och behovet av ytterligare granskning försvinner. När man tar hänsyn till alla kända undantag i mönsterparadigmet, har vi när det gäller svenska kommit fram till mer än 200 olika variationer – man borde säga miniordklasser.

Pragmatiska begränsningar gjordes för några böjningsformer som bedömdes av en infödd språkbrukare som konstiga, även om de inte var helt ogrammatiska. Som sådana klassificerades passiva preteritumformer av vissa verb, som danska *?bandtes* och *?lodes* (av *binde* och *lade*). Likaså skulle imperativformen av några verb låta konstiga, t.ex. danska *eksistere* – *?eksister!* eller norska *sykle* – *?sykl!* Som pragmatiska begränsningar räknades också de fall då en viss analys, och det därtill korresponderande genererade ordet, skulle vara väldigt sällsynt och därför inte för-

väntat. Sådana var genitivform av adjektiv i singularis, t.ex. svenska *?stors*, *?storts* och de motsvarande participformerna, t.ex. danska *?spists*, *?elskets*. Likadana bedömningar gjordes för komparativ- och superlativformer av vissa adjektiv. Medan några adjektiv som betecknar färg lätt kan kompareras, t.ex. svenska *grön – grönare – grönast*, skulle andra låta konstiga, t.ex. danska *orange – ?orangere – ?orangest*.

Pragmatiska begränsningar behövdes också på grund av homonymi, som förekommer relativt ofta i de skandinaviska språken. T.ex. danska *flyver* kan analyseras både som ett verb i presensform och som ett substantiv i grundform, och svenska *anklagades* antingen som passiv preteritum av verbet *anklaga* eller som den bestämda definitiva genitivformen av participet *anklagad*. Om man hade ordens kontextinformation, så kunde man lätt skilja mellan dessa alternativa analyser med existerande programverktyg. Problem uppstår när de existerande gränssnitten med ordbehandlingsprogram inte erbjuder någon kontextinformation. Detta är inte särskilt problematiskt då de båda analyserna är relativt frekventa och brukaren själv kan välja det alternativ som gäller. Men när någon analys är relativt infrekvent valde vi att begränsa dessa former för att minska brukarens förvirring, särskilt eftersom många inte betraktar dessa ord genom en lingvists glasögon. Man kan anta t.ex. att svenska *gås* hänvisar till fågeln och inte är presens passiv form av verbet *gå*.

3.2. Synonymi och böjning

Den föregående delen av denna uppsats har behandlat skillnader i böjningen av enstaka ord inom olika ordklasser, som först och främst är en fråga begränsad till den morfologiska analys- och genereringskomponenten av en inflekterande synonymordbok. När man kombinerar en semantisk relationsbeskrivning, i detta fall innehållet i en synonymordbok, med en sådan automatisk böjningsapparat, mötte vi ett svårare problem: Hur "bra" kommer det synonyma förhållandet mellan två eller flera ord att kvarstå i de respektive ordens olika böjningsformer?

Den nuvarande lexikografiska och semantiska teorin behandlar inte denna fråga – åtminstone inte direkt. Allmänt verkar böjningens interaktion med och påverkan på ordens semantiska egenskaper ha tillmätts en underordnad betydelse. T.ex Zgusta tonar ned böjningen som bara ett av många olika drag som ett ords betydelse består av, och Cruse förklarar böjningen som ett sekundärt drag semantiskt sett, som inte ändrar grundformens centrala betydelse:

The lexical meaning comprises of a great number of different components and there are many pertinent phenomena and relations which must be studied completely, but preferably as distinct, interdependent factors, if the whole meaning of a word (lexical unit) is to be analyzed, understood, and described. It will be useful to discern the following main components of lexical meaning: 1) the **designation**, 2) the **connotation**, and (possibly) 3) the **range of application**. (Zgusta 1971: 27)

We have so far assumed that it is a word form associated with a single sense, and that a difference of word form entails a difference of lexical unit. [...] Strictly speaking, we would be obliged, on this view, to regard, for instance *obey*, *obeys* and *obeyed* as representing different lexical units. It would, however, be more advantageous for our purposes to be able to say that they were **alternative manifestations of the same lexical unit** *obey*. (Cruse 1986: 76–77)

Senare, efter diskussion av grundkomponenterna i ordets betydelse, nämner Zgusta dock böjningens betydelse, och påminner om den variation i betydelse inom ett ords böjningsparadigm som påträffats i flera språk:

It cannot be stressed enough that in different languages, any members of any paradigm, whether 'regular' or 'irregular', can have a lexical meaning more or less different from the rest of the paradigm. [...] In any case, the lexicographer should always count on the possibility that a form – any form – of a paradigm can show a peculiarity in respect to its lexical meaning when compared with the other forms of the paradigm. (Zgusta 1971: 123, 126)

Trots denna observation, som Zgusta stöder med flera exempel, har han inte valt att betrakta böjningen som en av ordbetydelsens

huvudkomponenter. Senare har Sinclair också antytt detsamma utgående från sina erfarenheter i engelsk lexikografi:

... it is not yet understood how meanings are distributed among forms of a lemma, and a new branch of study is looming – the interrelationships of a lemma and its forms. (Sinclair 1991: 41)

När man fortsätter i litteraturen till beskrivning av ordens semantiska förhållanden verkar böjningens semantiska betydelse försvinna helt och hållet. I fall av synonymi koncentrerar man sig på ordens likhetsgrad, men antyder ingenting om böjning. Inte heller Zgusta tar upp böjningen när han beskriver synonymi:

two words are synonymous, if they have the same sense. (Lyons 1968)

Synonyms, then, are lexical items whose senses are identical in respect of 'central' semantic traits, but differ, if at all, only in what we may call 'minor' or 'peripheral' traits. (Cruse 1986: 267)

It is, however, advantageous to reserve the term synonymy, synonyms [...] only for the cases of absolute identity in meaning. [...] The absolute identity of meaning requires the identity of all the three basic components of meaning [being designatum, connotation and range of application]. [...] If there is a difference at least in one of the three basic components of meaning [...], the respective words are near-synonyms only. (Zgusta 1971: 89, 90)

Teorin har också följts i praktiken. I WordNet har man beslutat att lämna den morfologiska böjningen utanför den centrala semantiska beskrivningen:

Initially, interest was limited to semantic relations; no plans were made to include morphological relations in WordNet. [...] English nouns, verbs, and adjectives are organized in synonym sets, each representing one underlying lexical concept. [...] The three papers [describing the principles of WordNet] have little to say about lexical relations resulting from inflectional morphology, since those relations are incorporated in the **interface** to WordNet, **not in the central database**. (Miller 1993: 9)

I vårt testande av de första versionerna av inflekterande synonymordböcker märkte vi att det inte i praktiken går att bortse från böjningens påverkan på ordens synonymiförhållanden. Särskilt

var verbsynonymer känsliga för böjningen. Låt oss titta på de svenska synonymparen *hitta-finna* och *säga-betyda*. Fast dessa ord är synonyma i meningarna (1–4), är det inte längre möjligt att använda dem som synonymer för varandra i meningarna (5–8):

- (1) Sen *hittade* jag boken.
- (2) Sen *fann* jag boken.

- (3) Vad vill det *säga*?
- (4) Vad vill det *betyda*?

- (5) Boken *hittades*.
- (6) Boken *?fanns*.

- (7) *Säg* någonting!
- (8) **Betyd* någonting!

I dessa exempel försvinner det synonymiska förhållandet antingen på grund av att ordet ändrar sin semantiska betydelse i en eller flera böjningsformer (6), eller att den korresponderande böjningsformen är ogrammatisk för synonymordet (8). För svenskans del fann vi att denna inkongruens återfanns relativt ofta i verbsynonymgrupper och särskilt uppstod vid passiv- och imperativform, som vi därefter var tvungna att granska manuellt i detta avseende. Likadana fenomen fanns också i de danska och norska inflekterande synonymordböckerna.

Sammantaget var vår slutsats att man inte kategoriskt kan förvänta att synonymiförhållanden mellan två eller flera ord (i grundform) automatiskt också kvarstår i alla böjningsformer av dessa ord. En orsak till att detta fenomen inte behandlats särskilt mycket i litteraturen är kanske att hela idén att generera synonymbeskrivningar automatiskt inte riktigt har uppstått tidigare. Om man utgår från enbart de tryckta synonymordböckerna, är frågan inte så problematisk, för de synonymrelationer som beskrivs i ordboken gäller ju bara de former som finns tryckta i ordboken. I praktiken garanterar man ingenting beträffande böjda former av dessa ord, men psykologiskt sett pekar valet av nästan bara grundformer i sådana verk starkt i den riktningen att dessa semantiska förhållanden också skulle kvarstå i grund-

formernas böjda former. Men om man trovärdigt vill ange synonymer till verbens passivformer borde man egentligen skapa en separat synonymordbok för detta ändamål, utgående från varje böjningsform.

4. Sammanfattning

Vårt arbete för att åstadkomma inflekterande synonymordböcker för danska, norskt bokmål och svenska genom att integrera datoriserade morfologiska modeller för respektive språk med elektroniska versioner av synonymordböcker har gett betydande insikt i flera lingvistiska fenomen, särskilt i skärningsplanet mellan lexikalisk morfologi och lexikalisk semantik, mer specifikt synonymi. De viktigaste lärdomarna är följande:

1. Lexikalisk allmängiltighet kan man inte lita på i den utsträckning som man skulle anta, särskilt om man bara använder en grov ordklassindelning. Man behöver både morfosemantiska, morfologibaserade begränsningar i modellens ordformsgenerering och pragmatiska begränsningar på grund av frekvens- eller förväntningsinformation.
2. Man kan inte kategoriskt förvänta att synonymiförhållanden mellan två eller flera ord (i grundform) automatiskt också kvarstår i alla böjningsformer av dessa ord.

Till slut verkar våra erfarenheter antyda att det i det här området finns många frågor att forska vidare i.

Kontaktinformation

Antti Arppe, Lingsoft Ab/Helsingfors Universitet
antti.arppe@iki.fi

Litteratur

A. ORDBÖCKER

- Politikens Synonymordbog*. 1994. Karker, Allan. Politikens Forlag.
- Cappelens Media Synonymordbok*. 1983. Svenkerud, Herbert. J.W. Cappelens Forlag.
- Bonniers Synonymordbok*. 1995. Walter, Göran. Bonnier Alba.

B. ANNAN LITTERATUR

- Arppe, Antti; Voipio, Mari; Würtz, Malene. 1998. "Creating inflecting electronic thesauri." Working paper presented at the 17th Scandinavian Conference in Linguistics, August 20–22, 1998.
- Bilgram, Thomas. 1994. "Computerstyret analyse af dansk. En praktisk analyse af en væsentlig kilde til homonymi i dansk med forslag til kontekstuell disambiguering ved hjælp af constraint-regler." Upubliceret speciale, Institut for Lingvistik, Aarhus Universitet.
- Bybee, Joan L. 1985. *Morphology*. John Benjamins Publishing Company.
- Cruse, D.A. 1986. *Lexical Semantics*. Cambridge: CUP.
- Karlsson, Fred. 1992. SWETWOL: A Comprehensive Morphological Analyzer for Swedish. *Nordic Journal of Linguistics* 15:1–45.
- Koskenniemi, Kimmo. 1983. *Two-level Morphology: A General Computational Model for Word-Form Recognition and Production*. Helsinki: Publications of the Department of General Linguistics 11, Department of General Linguistics, University of Helsinki.
- Lyons, John. 1968. *Introduction to theoretical linguistic*. Cambridge: CUP.

- Miller, George A. et al. 1993. Introduction to WordNet: An On-Line Lexical Database. URL.
[http://www.cogsci.princeton.edu/~wn/...](http://www.cogsci.princeton.edu/~wn/) (visited June 1999)
- Moshagen, Sjur. 1998. "Morphological Analyzers for Norwegian (Bokmål and Nynorsk)." Working paper presented at the 17th Scandinavian Conference in Linguistics, August 20–22, 1998.
- Scalise, Sergio. 1984. *Generative Morphology*. The Netherlands: Foris Publications.
- Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Zgusta, Ladislav. 1971. *Manual of Lexicography*. (*Janua Linguarum, Series Maior*. 39.) The Hague: Mouton.

Tove Bjørneset

Introduksjon til ordboksprosjektet NORDLEXIN-N

Since 1996 a project team at the HIT centre, University of Bergen, has been working on transferring the source material of the Swedish dictionary series LEXIN into Norwegian. The LEXIN dictionaries are customized for non-native speakers who have limited proficiency in the source language and who may also be unfamiliar with the use of dictionaries as such. The Swedish source material amounts to approx. 30 000 entries distributed on the quantitative selections MINI (approx. 5 000 entries), MIDI (approx. 8 000 entries), STORA (approx. 17 000 entries) and MAXI (approx. 30 000 entries). The objective of the project is to exploit as much of the Swedish source material as possible in order to develop a similar Norwegian database. This may in turn be linked to the already existing translations (approx. 20 languages) so that a similar dictionary series can be automatically generated in Norway. This article gives a short outline of the work carried out so far.

Bakgrunn

Den svenske Skolöverstyrelsen (SÖ) tok i 1979 initiativ til å starte et forsknings- og utviklingsprosjekt med den hensikt å produsere ordbøker som egnet seg for svenskundervisning for innvandrere. Prosjektet fikk navnet *Språklexikon för invandrare* (*LEX-IN*) og ble gjennomført i samarbeid med Statens institut för läromedelsinformation (SIL) og Statens invandrarverk (SIV). Det grunnleggende utviklingsarbeidet ble avsluttet i 1984. Deretter overtok SIV hovedansvaret for å produsere ordbøker for innvandrere. Siden 1991 har Skolverket hatt dette ansvaret. Bakgrunnen for prosjektet var den svenske Riksdagens beslutning om at det svenske samfunnet skulle arbeide for aktiv tospråklighet og bedre svenskopplæringen for innvandrere.

LEXIN-ordbøkene er oversiktlige og enkle å bruke, noe som gjør at også innvandrere med svært begrensede leseferdigheter eller erfaring i bruk av ordbøker kan ha nytte av dem. Ordbøkene utgis i ulike størrelser; MINI (ca. 5 000 ord), MIDI (ca.

8 000 ord), STORA (ca. 17 000 ord) og MAXI (ca. 30 000 ord). Oversettelsene omfatter per i dag rundt 20 språk, deriblant tyrkisk, albansk, arabisk, finsk, gresk, makedonsk, persisk, tigrinsk, polsk, somalisk, armensk og turabdisk. Antall lemma i disse varierer, og flere ordbøker er under utgivelse.

Tidlig på 1990-tallet fikk samtlige nordiske land tilbud om å disponere den svenske databasen med ordboksgrunnet for LEXIN vederlagsfritt, for eventuell produksjon av sine nasjonale LEXIN-ordbøker. Det kongelige kirke-, utdannings- og forskningsdepartementet (KUF) var oppdragsgiver da det norske NORDLEXIN-prosjektet ble satt i gang høsten 1986 ved HIT-senteret (Forskningsprogram for humanistisk informasjonsteknologi), avdeling Norsk termbank, Universitetet i Bergen. Arbeidet hadde de første månedene status som et pilotprosjekt, der det svenske ordboksunderlaget for MINI-utvalget ble overført til norsk. I juni samme år startet arbeidet med overføringen av ordboksunderlaget for STORA-utvalget, et arbeid som pågikk til slutten av 1998. I 1999 overdrog KUF oppdragsgiverfunksjonen til Nasjonalt Læremiddelsenter (NLS), men samarbeidet med HIT-senteret fortsetter. Det norske ordboksunderlaget skal nå utvides til samme omfang som det svenske.

De svenske ordboksartiklene inneholder informasjon om ordklasse, fullt utskrevne bøyingsformer, forklaringer i form av omskrivninger og/eller synonym, setningskonstruksjon og uttalemarkeringer. I tillegg kommer ofte også stilistiske og grammatiske kommentarer samt språkeksempler og idiomer. Rundt 1 700 av de svenske lemmaene er illustrert. Illustrasjonene er inndelt i 32 emneområder (blant annet 'Familj och släkt', 'Skola och utbildning' og 'Mått och vikt. Geometri') og er samlet i et eget vedlegg i ordbøkene. De er nå også utgitt som et eget hefte, *Bildtema*. Blant annet som følge av særsvensk utforming og alder kan ikke de svenske illustrasjonene brukes i Norge, men en lignende løsning vil trolig bli valgt. Foruten den primære målgruppen, som består av innvandrere og flyktninger, kan et slikt bildehefte være et godt pedagogisk hjelpemiddel også for andre elevgrupper, som døve eller elever med lese- og skrivevansker.

Overføring av et ordboksgrunnlag fra svensk til norsk

Det som kanskje i størst grad særpreger det norske NORD-LEXIN-prosjektet, er at det er selve kildespråksunderlaget for en ordboksserie som overføres fra svensk til norsk. Denne overføringen skjer fortløpende under gjennomgangen og vurderingen av de alfabetiserte svenske ordboksartiklene. Arbeidet kan grovt deles inn i to hovedfaser:

1. ekvivalensarbeid;
2. tilrettelegging for norske forhold.

Under gjennomgangen av de svenske ordboksartiklene kartlegger og etablerer vi ekvivalenser mellom svensk og norsk i alle aktuelle felt. I denne fasen konsentrerer vi oss om å utnytte det svenske grunnlagsmaterialet best mulig, noe som innebærer oversettelse av alle lemma, fraser, idiomer og sammensetninger. Alle ekvivalenser kontrolleres og godkjennes av en svensk filolog med lang erfaring i norskundervisning for fremmedspråklige. Det satses på at en vesentlig del av oversettelsene til målpråkene skal kunne gjøres maskinelt. For at dette skal være mulig, må alle norske felt knyttes til de svenske feltene som de er ekvivalente med.

Under tilretteleggingen for norske forhold er det naturlig å frigjøre seg noe mer fra det svenske ordboksunderlaget, for å sikre at lemmautvalget i større grad avspeiler det norske samfunnet. I denne fasen blir basen systematisk supplert med nye norske ordboksartikler. Nye fraser, sammensetninger og idiomer legges inn etter skjønn i både ekvivalerte og nye norske ordboksposter. Til slutt legges det inn morfologisk informasjon og uttalemarkeringer.

Ordboksartiklene inneholder flere felt som beskriver bruksområdet for lemmaet (blant annet definisjon, frase, idiom og sammensetninger). Det varierer hvilke felt som har blitt oversatt i de ulike LEXIN-ordbøkene, men vi har likevel valgt å etablere samtlige ekvivalenser vi har funnet. Det vil dessuten være av betydning for en eventuell fellesnordisk LEXIN-base at oversettel-

sene av det svenske underlaget er så fullstendige som mulig. Vi har tatt utgangspunkt i det svenske postformatet, men bearbeidet det noe og utviklet et eget postformat i en ORACLE-base. Ordbokspostene har, noe grovt skissert, denne strukturen:

S-lem	svensk lemma
E-lem	norsk lemma
S-kat	svensk ordklasse
E-kat	norsk ordklasse
S-def	svensk definisjon/synonym
E-def	norsk definisjon/synonym
S-til	svensk tilleggsdefinisjon (ved samfunnstermer)
E-til	norsk tilleggsdefinisjon (ved behov)
S-eks	svensk frase
E-eks	norsk frase
S-idi	svensk idiom med forklaring
E-idi	norsk idiom med forklaring
S-div	henvisning til svensk illustrasjon
E-div	henvisning til illustrasjon ikke etablert
S-sms	svensk sammensetning
E-sms	norsk sammensetning
S-mor	fullt utskrevne bøyingsformer
E-mor	fullt utskrevne bøyingsformer (med alle lovlige varianter)
S-utt	uttalemarkering med ordinære bokstaver, markering av tonem
E-utt	uttalemarkering med ordinære bokstaver, ikke markering av tonem

S-feltene inneholder svensk ordboksmateriale, mens E-feltene er opprettet av oss og skal fylles med ekvivalent, norsk materiale. Alle ordboksposter og felt har sine unike koder med henblikk på pilotprosjektet med den datamaskinelle oversettelsen. Om det ikke eksisterer noen ekvivalent for det svenske lemmaet, inaktiveres ordboksposten, slik at den ikke lenger er en del av den norske databasen. Mangler det en norsk ekvivalent for et av de andre feltene, mens selve lemmaet er ekvivalent, inaktiveres feltet ved at det settes inn en bindestrek. Eventuelt legger vi inn et særnorsk felt med innhold som ikke skal knyttes til noe svensk felt, og som da må oversettes separat til det aktuelle målspåket. Den ekvivalerte ordboksposten under kan illustrere dette. Posten er fullstendig ekvivalent. Det svenske og det norske begrepet er det samme, men definisjonen er utdypet:

S-lem	glas
E-lem	glass
S-kat	subst.
E-kat	subst
S-def	ett hårt och genomskinligt ämne
E-def	-
N-def	et hardt, knusbart og gjennomsiktig materiale
S-div	bild 7:14
E-div	
S-sms	glas~tillverkning -en
E-sms	glass~produksjon -en
S-mor	glaset glas glasen
E-mor	glasset glass glassene (el glassa)
N-alt	glas
S-utt	glɑ:s
E-utt	glas

Ekvivalens

NORDLEXIN-N-prosjektet er ikke et oversettelsesprosjekt i vanlig forstand, men kan bedre karakteriseres som et *ekvivale-
ringsprosjekt*. Å kartlegge og etablere ekvivalenser for lemma-
feltene har naturlig nok vært det mest sentrale i dette arbeidet. I
Nordisk leksikografisk ordbok (1997) defineres *ekvivalens* slik:

semantisk og funksjonsmessig overensstemmelse mellom ord
eller uttrykk i to eller flere språk.

Begrepene kan ofte være vagere og mer upresise innen allmenn-
språk enn fagspråk. Den typiske termgruppen består hovedsake-
lig av substantiv, mens vi i dette ordboksprosjektet i større grad
har hatt leksem fra alle ordklasser. *Fullekvivalens*, det vil si full-
stendig samsvar i tegninhold og betydningsomfang, forekom-
mer sjelden innen allmennspråk, men oftere innen fagspråk. Det
er kartlagt flere fullekvalenser enn forventet i prosjektet. Ved
partiell ekvivalens er det delvis likhet mellom kildeord og
målord. Forskjellen kan blant annet vise seg ved at begrepet i
målspråket har færre karakteristiske trekk enn begrepet i kilde-
språket. Eksempelvis kan det svenske lemmaet *betyg* i mange
tilfeller oversettes til *karakter*, men vi må tenke generelt og
begrepsorientert. I tillegg til denne betydningen, dekker *betyg*

også det norske ordet *vitnemål*. Vi kan derfor ikke ekvivalere det svenske lemmaet med *karakter*. Fordi det ikke finnes noen norsk overterm som dekker både *karakter* og *vitnemål*, kan vi heller ikke sette inn en norsk ekvivalent for *betyg*. Vi må derfor inaktivere den svenske posten og opprette to nye poster for *karakter* og *vitnemål*. Det er også kartlagt en viss mengde av *nullekvivalenser*. Ikke sjelden mangler norsk fullstendige eller partielle ekvivalenter til de svenske leksikalske størrelsene. Som regel inntreffer dette i møte med kulturspesifikke svenske begreper, som for eksempel med *De Aderton* (medlemmene i Svenska Akademien).

Nye norske lemma

Svensk og norsk språk er mer ulike enn man gjerne vil tro. Tilsynelatende identiske ord viser seg gjerne å være falske venner eller tilsløre ofte viktige betydningsnyanser og -forskjeller. Ord som er synonyme i ett av språkene, kan ha klart atskilte betydninger i det andre språket. I en del tilfeller har slike og lignende ulikheter gjort det nødvendig å lage nye norske poster.

Samfunnsinstitusjoner med tilhørende stillingskategorier og funksjoner er ofte ulikt organisert i Sverige og Norge. En rekke ord som er knyttet til det svenske samfunnet og samfunnssystemet har vi valgt å inaktivere i den norske versjonen av ordboksgrunnlaget. I stedet har vi lagt inn tilsvarende norske ord som nye oppslag. Videre har det vært nødvendig å legge inn en del ord som refererer til nyere fenomen, blant annet dataterminologi. Det er også ekserpert en del fra brosjyrer og foldere utgitt av kommunale etater, apotek, bank, postkontor, arbeidskontor og offentlige kommunikasjonsbedrifter, for å fange opp flest mulig ord som man møter i hverdagslivet. Tilsvarende er det lagt inn ord og uttrykk knyttet til skoleverk og skolefag, sport og fritidsaktiviteter. Vi har valgt å utelate opplagt tidsavgrenset informasjon, med unntak av blant annet de største norske politiske partiene og organisasjonene.

Prosjektnormal

Alle former innen læreboknormalen skal være representert i ordboksunderlaget for NORDLEXIN-N. De praktiske konsekvensene av dette er blant annet et nærmest uendelig mangfold av valgmuligheter: Hvilken rekkefølge bør de ulike formene presenteres i? Hvilke former bør vi bruke i definisjoner og eksempler? Vi kom fram til at det ikke ville være riktig å bruke bare én variant av norsk bokmål. I ordboksprosjekter som dette er det viktig å ikke tilsløre de relativt store variasjonsmulighetene i norsk språk, men tvert imot å gjøre dem synlige for brukeren.

Vi valgte likevel å etablere og operere med en allmennspråklig prosjektnormal til bruk i lemma- og definisjonsfeltene. Prosjektnormalen kan karakteriseres som et relativt moderat bokmål. I andre felt har vi bevisst vekslet mellom de ulike tillatte formene i læreboknormalen. Innbyrdes inkonsekvens har ikke blitt vurdert som problematisk, men som ellers i bokmål, er det gjerne konkrete substantiver og visse stilistiske kontekster som i særlig grad motiverer valg av hokjønnsord og andre "radikale" former.

Valgfrie former

De ulike formene er atskilt med mellomrom. Sidestilte former settes i vanlige parenteser og innledes med "el" for "eller". Den store valgfriheten i norsk gjør imidlertid at denne løsningen kan virke lite oversiktlig og derfor fungere dårligere her enn i Sverige. Vi ønsker derfor å endre dette oppsettet.

I morfologifeltet er de valgfrie formene lagt inn i samme rekkefølge som de ulike sidestilte formene av lemmaet. Lemmaet *beruset* har *berust* og *berusa* som sidestilte former. *Beruset* er valgt som hovedinngang, mens *berust* og *berusa* er lagt inn i felt for alternative former. I morfologifeltet settes de finite formene opp i samme rekkefølge. I tilfeller der lemmaet bare har én variant innenfor læreboknormalen, men det er flere valgfrie bøyingsformer, har vi som et grunnprinsipp valgt å sette de mest konservative formene først.

Vi har forsøkt å rasjonalisere med plassen i de norske morfologifeltene, der dette har vært mulig. Når variasjonen mellom ulike former ikke gjelder bøyingsparadigme, men annen variasjon, som for eksempel monoftong/diftong, mener vi at det er tilstrekkelig å illustrere bøyningen med den varianten som er oppført i lemmafeltet. Ordene *sen* og *sein* har identisk bøyning, og det er derfor ikke nødvendig å gjenta informasjonen om at også diftong er tillatt, i morfologifeltet.

Uttalemarkering

Det finnes ingen normert uttale for norsk, og heller ikke noe "høyspråk". Det har etter hvert blitt akseptert å snakke sin egen dialekt i alle sammenhenger, uansett posisjon og tilholdssted. Det nærmeste vi kommer en standardisert uttale, er kanskje scenespråket på teater og lignende steder, men dette kan høres stivt og unaturlig ut for mange. I vår statlige nyhetskanal NRK må journalistene snakke bokmål eller nynorsk i nyhetssendingene, men uttalen kan likevel variere mye, avhengig av hvor i landet personene kommer fra.

På bakgrunn av dette er det etter vår oppfatning ikke mulig å gi en nøyaktig uttalemarkering for moderne norsk. En uttalemarkering som skal være funksjonell for flest mulig i målgruppen, må nødvendigvis være så generell at den ikke blir direkte misvisende for svært mange av de norske dialektene. Likevel må den gi mer informasjon enn hva selve skriftbildet gir. Der det har vært nødvendig å velge mellom to uttalemåter med geografisk variasjon, har vi valgt den som dominerer i Østlandsområdet. Årsaken er at de største delene av målgruppen bor der. Er to uttalemåter utbredt over hele landet, er begge tatt med. Det er ikke brukt IPA-tegn i uttalemarkeringene, da vi har vurdert disse som for avanserte for målgruppen. I stedet brukes den vanligste bokstaven eller bokstavkombinasjonen for en norsk lyd. Hver lyd er vanligvis representert av én bokstav. Diftonger markeres med bue for å vise at de to lydene hører sammen.

Kvantitet er bare markert for vokalene. Markeringen består av et kolon satt direkte etter den lange vokalen, fordi lengden på konsonantene kan utledes fra lengden på vokalene:

lete etter noe	[le:te]
flyet skal snart lette	[lete]

Trykk markeres med et punkt under den aksentuerte vokalen, med unntak av i enstavelsesord. Hovedtrykk markeres i alle posisjoner. Bitrykk markeres ikke:

bokst.a:v
arb.æ\ider

Norsk har to tonem (tonelag): Det er ulik uttale på ¹*tanken* (bestemt form av substantivet *tank*) og ²*tanken* (bestemt form av substantivet *tanke*). Dette markeres likevel ikke, fordi det er store dialektvariasjoner med hensyn til hvilket tonem et ord har, og også med hensyn til hvilke ord som har tonemvariasjoner eller ikke.

Ordklasseinndeling

Vi har tatt utgangspunkt i *Norsk referansegrammatikk* og bruker disse ordklassene:

Substantiv	(en gammel mann)
Pronomen	(han spurte henne om de skulle gifte seg)
Determinativ	(en gammel mann)
Adjektiv	(en gammel mann)
Adverb	(det regner ikke nå)
Verb	(jeg heter Nina og bor i Norge)
Preposisjon	(jeg bor i Bergen)
Interjeksjon	(huff! , æsj! , au!)
Konjunksjon	(han var snill og grei)
Subjunksjon	(han spurte om jeg kom)

Sluttkommentar

Vi blir ofte spurt om årsaken til at vi går via et svensk ordboksunderlag i arbeidet med å lage norske ordbøker for minoritetsspråklige innvandrere. Ordboksproduksjon kan foregå på ulike måter, og NORDLEXIN-N-prosjektet kan sies å være tuftet på en kombinasjon av disse. Ved å overføre det svenske ordboksunderlaget for LEXIN-ordbøkene til norsk, satser vi på å få elektronisk tilgang til de rundt 20 målpråkene som LEXIN til nå er oversatt til. Videre bør det ligge et ikke ubetydelig kvalitetssikringsaspekt i det å bygge på et svensk ordbokskonsept som har fungert godt i en svært differensiert målgruppe gjennom snart 20 år. På dette grunnlaget håper vi å kunne utvikle brukervennlige og funksjonelle ordbøker i trykt og elektronisk form for innvandrere også i Norge.

Litteratur

- Beijer, Maj og Kiros Fre Woldu. 1997. *Detta är LEXIN. Lexikon för invandrare*. Stockholm: Skolverket.
- Bergenholtz, Henning m.fl. 1997. *Nordisk leksikografisk ordbok*. Oslo: Universitetsforlaget.
- Faarlund, Jan Terje m.fl. 1997. *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.
- Gellerstam, Martin. 1999. "LEXIN – lexikon för invandrare." *LexicoNordica* 6/1999.
- Svensén, Bo. 1997. *Handbok i lexikografi. Principer och metoder i ordboksarbete*. Stockholm: Norstedts.
- Svenska ord med uttal och förklaringar*. 2 uppl. 1992. Stockholm: Norstedts.

Erica Brewitz
Anna-Lena Bucher

TNCs arbete med Eurodicautom, Europeiska kommissionens termdatabas¹

The paper gives a brief orientation of the goals and activities of The Swedish Centre for Technical Terminology (TNC) and the reasons why subject field specialists feel the need of determining and describing subject field related concepts and terms. The principles and methods of terminology work are described, and the importance of having subject field specialists participating in the work is particularly emphasized.

Since Sweden's membership in the EU in 1995, TNC's activities have been dominated to a large extent by the comprehensive task from the European Commission of updating the terminologic database Eurodicautom with Swedish terminology in various fields. In four different projects, each running for twelve months, TNC has been supplementing approximately 140 000 term records with Swedish terms and references, and to some extent also with definitions and comments. The working principles in some respects diverge from the traditional methods of terminology work. In other respects, e.g. the close cooperation with subject field specialists, it has been possible to retain the principles. However, TNC has had to adapt some of the working methods to the specific conditions of these projects. The last part of the paper is a report from one of the specialists, a biologist. She describes her motives for applying for a job within this project, and her experiences after having been involved in the work for more than a year.

1. TNC och terminologiarbete

Tekniska nomenklaturcentralen, TNC, har regeringens uppdrag att åstadkomma terminologi som är ändamålsenlig och lämpad för svenska förhållanden. TNC etablerades 1941 på initiativ av tekniker, i huvudsak patentingenjörer, standardiserare och upp-

¹ Artikeln är uppdelad i tre delar:

1. TNC och terminologiarbete.

2. TNCs uppdrag att komplettera Eurodicautom med svenska termer.

3. Vittnesmål från en av TNCs anlitade experter i Eurodicautom-uppdragen.

Anna-Lena Bucher svarar för de två första delarna och Erica Brewitz för den tredje delen.

finnare. Ingenjörsvetenskapsakademien och ASEA var med bland grundarna och TNCs förste föreståndare, John Wennerberg, kom från ASEA. Resultatet av arbetet genom åren är bl.a. ordlistor utarbetade enligt terminologiska principer och med terminologiska metoder. Typiskt för det traditionella terminologiska ordlistearbetet är följande:

- Företrädare för en bransch tar initiativ till en ordlista; orsaken är i regel att man vill undanröja missförstånd och få en förbättrad och effektiv kommunikation inom branschen.
- Ämnesspecialister deltar alltid i arbetet tillsammans med (minst) en terminolog .
- Arbetet är begreppsorienterat vilket bl.a. innebär att man i första hand analyserar, preciserar och beskriver begreppen i definitioner; därefter rekommenderar man också vilken eller vilka termer som i första hand bör användas.
- Resultatet är normerande.

Den vaghet i betydelsen som finns hos vissa ord kan vi hantera i vardagsspråket. I den fackspråkliga kommunikationen är det däremot viktigt att ha precision och entydighet. Ordet *träd* har av Ruth Vatvedt Fjeld nämnts som exempel på ett ord med vag betydelse enligt vissa allmänspråkliga lexikon. I TNCs *Skogsordlista*, en ordlista som omfattar begrepp inom skogbruk och skogsvård, betraktas *träd* som en term och ges där en precis avgränsning i definitionen.

<p>träd, magnolignid (som livsform:) vedväxt som är högvuxen och icke klättrande och som har genomgående huvudstam Höjd hos träd överstiger normalt 1,3 m (från praktisk skogsförnyringssynpunkt). Internationell livsformsterm är <i>magnolignid</i>. På latin heter träd <i>arbor</i>. <i>en</i> tree, magno-lignid <i>de</i> Baum, m <i>fi</i> puu</p>	<p>term synonym definition</p> <p>anmärkning</p> <p>engelsk term synonym tysk term finsk term</p>
<p>Källa: TNC 96 Skogsordlista</p>	

Det överordnade begreppet *vedväxt* definieras så här.

<p>vedväxt, lignos [-å's], <i>en, -er</i> växt vars stam och grenar är förvedade och mångåriga <i>en</i> lignosis <i>de</i> Holzpflanze, f, Holzgewächs, n <i>fi</i> puukasvi</p> <p>Källa: TNC 96 Skogsordlista</p>	<p>term synonym grammatik definition</p> <p>engelsk term tysk term synonym finsk term</p>
--	--

Det som visas här är *termposter*. En termpost innehåller utsagor om ett begrepp, t.ex. term, definition, anmärkning, korshänvisning till relaterade begrepp liksom termmotsvarigheter på andra språk. Termposten är väl strukturerad i olika fält avsedda för dessa olika informationskategorier. Vilka utsagor som väljs beror på fackområdet i fråga och på de önskemål och behov som branschen har.

Under TNCs snart 60-åriga verksamhet har ca 30 000 sådana här termposter utarbetats och presenterats i ordlistor omfattande ca 50 olika fackområden.

2. TNCs uppdrag att komplettera Eurodicautom med svensk terminologi

Sedan Sveriges EU-inträde 1995 har TNC i omgångar fått i uppdrag att komplettera Europeiska kommissionens termdatabas Eurodicautom med svensk terminologi. Nedan följer några fakta om Eurodicautom:

Eurodicautom är världens största termdatabas och innehåller drygt en miljon termposter; den 4 maj 1999 var den exakta siffran 1 241 779 termposter. De flesta termposterna innehåller termer, men det finns också fraser och förkortningar, speciellt akronymer. Eurodicautom började byggas upp under de första åren av 70-talet och var inledningsvis huvudsakligen avsedd för kommissionens översättare. I dag används den även av övriga EU-tjänstemän samt av andra yrkesaktiva, kanske framförallt av

översättare utanför EU-institutionerna och ute i medlemsnationerna.

Innehållet i Eurodicautom utgjordes från början till största delen av flerspråkiga ordlistor, vissa med definitioner, som senare kompletterats med termmotsvarigheter på andra språk. Många av de ingående ordlistorna är lexikografiskt utarbetade, men bland dem finns också en del terminologiska ordlistor. Flera av TNCs ordlistor har t.ex. funnits inmatade i Eurodicautom långt före Sveriges medlemskap, men då förstås utan svenska motsvarigheter; ett antal TNC-ordlistor med danska ekvivalenter fördes in i databasen i samband med Danmarks EG-inträde 1974.

I princip är Eurodicautom uppbyggd enligt den princip som terminologiska ordlistor är uppbyggda på, nämligen en termpost-ett begrepp. En ideal termpost kan se ut som på nästa sida. Termposten är ideal i det avseendet att den faktiskt innehåller endast ett begrepp som dessutom har termmotsvarigheter på alla de officiella EU-språken. Att det är samma begrepp som avses kan man utläsa i definitionerna.

TNC har sedan 1995 i fyra omgångar arbetat med att tillföra svenska termer i Eurodicautom, och vi befinner oss just nu i slutfasen av det fjärde projektet som skall slutlevereras i början av juni. När detta projekt är avslutat har vi totalt analyserat och bearbetat ca 140 000 termposter för svenskans del.

Projektet har varit en utmaning för oss i många avseenden; de har utvecklat TNCs kompetens i översättningsorienterat terminologiarbete, och den stora utmaningen har varit att kunna tillämpa de terminologiska principerna inom mycket begränsade ramar.

Här är några jämförelser. Den sammanlagda tiden för allt arbete med en traditionell termpost varierar mellan 2–6 timmar. För varje termpost som vi arbetat med i Eurodicautom har vi haft 15–20 minuter till förfogande.

	Terminology Office	Terminology Office, European Commission, Brussels (=BTB)
	Collection	Civil Protection (=CIF90)
	ID Number	0000842
	Date	000628
	Reliability	4
	Subject	public safety services (=AD7) Unknown CM (=CIV) Unknown CM (=)
English	Term	<i>isotope</i>
	Definition	Nuclides having the same atomic number and thus identical chemical properties
	Reference	Multilingual Glossary on Civil Protection 1990, European Commission/Member States Working Party
Danish	Term	<i>isotop</i>
	Definition	Nukleider, som har samme atomtal og således har ens kemiske egenskaber
	Reference	Fler sproget glossar vedrørende civilbeskyttelse 1990, Arbejdsgruppe Europa-Kommissionen/Medlemsstaterne
Spanish	Term	<i>isótopo</i>
	Definition	Nucleido que tiene el mismo número atómico que otro y, consiguientemente, sus mismas propiedades químicas
	Reference	Glosario multilingüe de protección civil 1990, Grupo de Trabajo Comisión Europea/Estados Miembros
French	Term	<i>isotope</i>
	Definition	Chacun des nucléides présentant le même numéro atomique ainsi que des propriétés chimiques similaires
	Reference	Glossaire multilingue en matière de protection civile 1990, Groupe de travail Commission Européenne/Etats membres

German	Term	<i>Isotop</i>
	Definition	Atomkerne mit gleicher Ordnungszahl und mit gleichen chemischen Eigenschaften
	Reference	Mehrsprachiges Glossar für den Katastrophenschutz 1990, Arbeitsgruppe Europäische Kommission/Mitgliedstaaten
Italian	Term	<i>isotopo</i>
	Definition	Nuclide avente lo stesso numero atomico e quindi uguali proprietà chimiche di un altro
	Reference	Glossario multilingue della protezione civile 1990, Gruppo di lavoro Commissione europea/Stati membri
Dutch	Term	<i>isotoop</i>
	Definition	Nuclide met hetzelfde atoomgetal als een ander en met derhalve dezelfde scheikundige eigenschappen
	Reference	Meertalig Glossarium Civiele Bescherming(B)/Rampenbestrijding (NL) 1990, Werkgroep Europese Commissie/Lid-Staten
Portuguese	Term	<i>isótopo</i>
	Definition	Núclidos com o mesmo número atómico e, por consequência, com propriedades químicas idênticas
	Reference	Glossário multilingue de protecção civil 1990, Grupo de Trabalho Comissão Europeia/Estados-membros
Finnish	Term	<i>isotooppi [1]</i>
	Definition	osa kemiallisten aineiden sarjassa, jossa kaikilla on samat kemialliset ominaisuudet, mutta eroa atomipainossa tai sähkövarauksessa
	Reference	Salmenlinna, Anti-Jussi, palomestari, Uudenmaan Pelastusliitto ry, Helsinki, 1998 [VE 1]
Swedish	Term	<i>isotop</i>
	Definition	en av flera nukleider med samma atomnummer men olika masstal
	Reference	Kärnenergiordlista. -1990(TNC 90)
	Note	{ <i>NTE</i> } isotop är en mindre lämplig benämning på radioaktiv nuklid

I ett traditionellt terminologiprojekt arbetar man i en grupp bestående av en eller två terminologer och fem–sex ämnesspecialister och arbetet handlar i korthet om att:

- välja ut begrepp
- analysera begreppen och upprätta begreppssystem
- skriva definitioner och eventuellt anmärkningar
- rekommendera en eller flera termer
- tillföra ekvivalenter.

Principiellt skiljer sig Eurodicautom-projekten från ett traditionellt terminologiprojekt på följande punkter:

- Urvalet av begrepp är redan gjort och kan inte påverkas.
- Termposterna har redan, på gott och ont, bearbetats av flera andra upphovsmän som har tillfört termer och annan information på andra språk.
- Det ges inga hänvisningar till närbesläktade begrepp vilket kan ge en fragmenterad bild av verkligheten.
- Den huvudsakliga målgruppen är översättare och inte ämnesspecialister som i traditionella projekt.

Hur har vi då gjort? Grovt sett har vi delat in arbetet i fyra steg:

Steg 1

Sortering och analys av filerna.

Kontakt med lämpliga experter.

Detta steg sköttes av terminologer eller handläggare på TNC.

Steg 2

Begreppet identifieras.

Detta steg sköttes av en person med ämneskompetens – ibland någon internt – oftare externt. Totalt har vi anlitat ca 100 externa experter bara för det fjärde Eurodicautom-projektet.

Steg 3

Termposten kompetteras med svensk term motsvarighet och annan information.

Detta steg skötte av fackexperten ensam eller i samarbete med TNCs terminolog eller handläggare

Steg 4

Terminologisk och språklig granskning.

Detta steg skötte av terminolog eller handläggare på TNC.

De svårigheter vi stötte på berodde i huvudsak på något av följande:

- Det var svårt eller omöjligt att identifiera vilket begrepp som avsågs eftersom informationen på de övriga språken var bristfällig eller disparat.
- Vissa poster innehöll kulturspecifika begrepp som saknas i Sverige.
- Vissa poster innehöll föråldrade begrepp.
- Vissa poster innehöll flera begrepp.

Här nedan följer några exempel (s. 61ff.).

I det interna anmärkningsfältet visas förarbete som gjorts på TNC för att underlätta expertens identifiering av begreppet: den franska definitionen har grovöversatts till svenska och därtill har ett belägg från en ordlista över maträtter redovisats. Expertens kommentar och kännedom om denna konfekt syns också. Slutsatsen som terminologen och experten i samråd kommit fram till visas i de svenska fälten: termen *calisson* föreslås också som svensk term med synonymen *mandelkonfekt* och begreppet preciseras genom att det står i definitionen att det är konfekt.

French	<i>Keyword</i>	calisson
	<i>Definition</i>	confiserie à base de pâte d'amande, à dessus glace, à dessous en pâte azyne. De form losangique spécialité d'Aix-en-Provence
	<i>Reference</i>	Clement, Dictionnaire des industries alimentaires, Paris, 1978
English	<i>Keyword</i>	marzipan sweet
	<i>Reference</i>	Clement, Dictionnaire des industries alimentaires, Paris, 1978
Swedish	<i>Keyword</i>	calisson; mandelkonfekt
	<i>Definition</i>	konfekt av rostad mandelmassa med melon, utstruken på oblat och med glacerad vit ovansida
	<i>Reference</i>	Jan Hedh, Konditor Konsult (1997)
	<i>Note</i>	romboidformad
<hr/>		
TNCs internanmärkning:	<p>TNC: Grovöversättning av fransk definition: konditorivara (eller konfektyr) med mandelmassa som huvudingrediens, med glacerad ovansida och med en undersida gjord av osyrat bröd. I form av en romb, specialitet i Aix-en-Provence.</p> <p>TNC: Förarbete: Zenk, G: Köksfranska. – 1996: "mellanting mellan konfekt och kaka som görs av mandel; från Aix-en-Provence".</p> <p>Expert: Provensalsk specialitet av rostad mandelmassa med melon, utstruken på oblat. Glacerad vit ovansida. Romboidformad. Mandelkonfekt.</p>	

Det finns också exempel på fraser:

French	<i>Phrase</i>	chercher un emploi
	<i>Definition</i>	être à la recherche d'un poste de travail
	<i>Reference</i>	FRANKL
German	<i>Phrase</i>	eine Stelle suchen; einen Arbeitsplatz suchen
	<i>Definition</i>	
	<i>Reference</i>	FRANKL
Danish	<i>Phrase</i>	søge beskæftigelse; søge arbejde
	<i>Definition</i>	
	<i>Reference</i>	K.Thrysoe, Arbejdsministeriet
Swedish	<i>Phrase</i>	söka ett arbete
	<i>Reference</i>	Kent Källström, Stockholms universitet (1999)

French	<i>Phrase</i>	chercher de l'emploi
	<i>Definition</i>	être à la recherche d'une activité professionnelle
	<i>Reference</i>	FRANKL
German	<i>Phrase</i>	Arbeit suchen; eine Beschäftigung suchen
	<i>Definition</i>	
	<i>Reference</i>	FRANKL
Swedish	<i>Phrase</i>	söka arbete
	<i>Reference</i>	Kent Källström, Stockholms universitet (1999)

Jag har med den här korta genomgången velat fästa uppmärksamheten på en del av terminologiarbetet som är oerhört viktig och som vi, de begränsade resurserna till trots, försökt göra så lite avkall på som möjligt i Eurodicautom-arbetet, nämligen expertmedverkan. Erica Brewitz är en av de experter som TNC har anställt på projektbasis för det här arbetet. Hon har arbetat hos oss som både handläggare och expert och har förutom kunskaper inom sitt fackområde också tillfört oss sitt kontaktnät.

3. Vittnesmål från en av TNCs anlitate experter i Eurodicautom-uppdragen

Jag är naturvetare. Biolog. Det innebär att jag har fått studera många skiftande områden, allt från hur djur och växter har utvecklats genom åren till de komplicerade biokemiska reaktioner som sätter igång i min kropp när jag blir förälskad. Under flera år var labbrocken som en andra hud för mig eftersom jag är en så kallad "vit biolog". En vit biolog står vid en bänk på ett laboratorium och gör experiment av olika slag till skillnad från den gröna biologen som bär gummistövlar och är ute i skogen.

Från biologi till terminologi – steget från laboratoriet till TNC kan tyckas vara långt. Det är det inte. När jag första gången hörde talas om TNC sökte man folk till Eurodicautom-projektet, något som jag då uppfattade som ett översättningsuppdrag. Det tyckte jag lät intressant, jag har alltid varit intresserad av språk, och i detta projekt skulle jag kunna kombinera både språkinträsse och naturvetenskapligt kunnande. Efter att ha fått en kort introduktion i vad arbetet verkligen innebar och vad terminologi är, insåg jag att det för mig var ett drömjobb. Varför? Jo:

- Som student traggade jag mig genom 17 kursböcker, alla utom en på engelska.
- Den senaste svenska bok som behandlar det område jag har specialiserat mig på, växtfysiologi, är från 1950.
- Den enda svenska biologiska ordlista som finns (från 1976) är till innehållet mycket mager, den skulle behöva fyllas ut med 3–4 gånger så många termer och definitioner för att matcha de engelska motsvarigheterna.

Med andra ord: Den svenska terminologin inom biologin håller på att utarmas och det vill jag inte ska ske. Så när jag fick chansen att arbeta med Eurodicautom blev jag själagläd! Trots att Eurodicautom är en databas som först och främst riktar sig till översättare har jag ändå hela tiden sett den som en källa för biologer och andra naturvetare, att utnyttja.

Arbetet med Eurodicautom har fört mig i kontakt med många olika ämnesområden det senaste året: molekylärbiologi, miljö-vård, skadeinsekter, trädgårdsbruk, biokemi, bioteknik, exotiska träd ... Naturligtvis behärskar jag inte alla dessa skilda områden till fullo, men jag har ändå kunnat agera som expert till en början genom att tillföra svenska termer så långt det har varit möjligt för mig. Därefter har jag iklätt mig handläggarens roll och anlitat andra experter. En hel del experter har jag funnit i det digra expertregister som TNC har byggt upp genom åren, men jag har också skamlöst utnyttjat alla de människor jag har lärt känna som student och doktorand. Totalt har jag varit i kontakt med ca 35 ämnesspecialister, några av dem har hjälpt till med flera hundra termer, andra bara med ett par.

Anna-Lena poängterade tidigare vikten av att i allt terminologiarbete samarbeta med experter och att detta även gäller när man tillför termer i Eurodicautom. Jag kan inte annat än instämma. Som exempel kan nämnas de termposter som behandlar exotiska trädarter (ca 350) och de som behandlar skadedjur (ca 1600) som vi har arbetat med. För en lekman kan det vara oerhört frestande att göra en direktöversättning om man ser att ett skadedjur heter "clay coloured weevil" på engelska och låta den svenska motsvarigheten få bli "lerfärgad vivel" när den i stället är "öronvivel" enligt experter på skalbaggar. Tyvärr råder det redan idag viss förvirring i nomenklaturen när det gäller många arter, så här måste absolut experter medverka så att inte kaos bryter ut.

Majoriteten av de skadedjur och trädslag som har ingått i det här projektet förekommer inte i Sverige, eller ens i Europa, och har därför ofta saknat svenskt artnamn. Experterna som hjälpt oss bringa ordning på alla småkryp och träd har alla blommat upp när de insett att de här har haft chansen att som förste svensk namnge arterna (jag kallar det "Carl von Linné-syndromet"). De har tagit sin uppgift på största allvar och det kan man småskratta åt. Men samtidigt betyder det att detta EU-finansierade projekt har bidragit till att vi i Sverige i dag har fått en markant ökad svensk nomenklatur vad beträffar exotiska trädarter och skadedjur och det är inte illa.

Sammanfattningsvis så har jag verkligen njutit av att arbeta med Eurodicautom. Jag har lärt mig att tänka som en terminolog och det kommer jag alltid att ha nytta av i framtiden. Jag har periodvis fått leka detektiv, både i mitt sökande efter den rätta termen och efter den rätte experten, vilket tilltalar den forne forskaren i mig. Det har varit mycket stimulerande att göra bekant-skap med Eurodicautom och jag är faktiskt ganska stolt över att ha fått hjälpa till med att tillföra svenska naturvetenskapliga termer till denna gigantiska databas.

Litteratur

Preparing data for Eurodicautom. 1997. Luxembourg: European Commission: Terminology and Language Support Services.

Skogsordlista. 1994. Solna: TNC. (*Tekniska nomenklaturcentralens publikationer*. 96.)

Westerberg, Kjell & Gustafsson, Maria. 1998. "Kan man tillämpa principerna för terminologiarbete när man kompletterar Eurodicautom med svenska termer?" *LexicoNordica* 5:117-127.

Yvonne Cederholm
Lena Rogström

Gamla ord blir som nya OSA-databasen som källa och resurs

The Swedish historical dictionary, *Ordbok över svenska språket utgiven av Svenska akademien*, is now available on the Internet as the OSA database. The dictionary as such is a very valuable source for lexicological research, and the OSA database makes it possible to extract separate structural categories in the dictionary in a much easier and faster way than you can accomplish by looking up the actual category in the 32 written volumes.

One way of using the database is to "recycle" the examples from the authentic texts used in the dictionary, in order to create synchronical corpuses from different periods. These corpuses can then be used as primary material for further examinations.

For this paper we extracted a textcorpus of all the 18th century examples from the database – approximately 850 000 words. We then made comparisons with two older studies of 18th century texts, the first one concerning morphology and the other spelling principles, in order to confirm, and hopefully improve, the earlier results.

Thus, our main purpose of this paper is to show the advantages of creating your own corpus out of already existing data, a new opportunity that the OSA database provides for scholars in historical linguistics.

Sedan 1983 har det s.k. OSA-projektet funnits vid Språkdata, Institutionen för svenska språket, Göteborgs universitet. Projektet syftar till att omvandla den svenska historiska ordboken *Ordbok över svenska språket utgiven av Svenska akademien*, vanligen kallad SAOB, till en sökbar databas – ett arbete som efter en lång struktureringsfas nu börjar ge resultat. (För mer ingående information om databasens utformning se Cederholm 1996.)

OSA finns sedan maj 1998 tillgänglig för sökning på Internet. Förutom de mest grundläggande sökningarna på enskilda artiklar i ordboken erbjuder gränssnittet också mer avancerade sökningsmöjligheter som bygger på att man kan söka enskilda kategorier genom hela ordboken – ett arbete som är möjligt att genomföra också i den tryckta ordboken men som där tar oerhört mycket längre tid i anspråk. Exempel på sådana sökningar är

t.ex. alla ord som kategoriserats enligt en viss fackordstillhörighet, har en viss etymologisk härkomst eller är bildade med ett visst avledningsmorfem.

Utgångspunkten för detta föredrag är en korpus bestående av alla språkprov från 1700-talet i SAOB, här kallad OSA-korpusen. Denna ska ses som ett första steg i uppbyggnaden av en större OSA-korpus, där användaren kan söka i delkorpusar genom att ange parametrar som genre, författare eller tidsperioder. Ett exempel kan vara en historisk dagstidningskorpus innehållande alla språkprov från Aftonbladet, Stockholmstidningen, Svenska Dagbladet etc. Detta är så vitt vi kan se ett nytt sätt att etablera ett källmaterial för exempelvis språkliga undersökningar av språkhistorisk karaktär.

SAOB är i kraft av sin storlek och omfattning en viktig källa i lexikalisk forskning. Ordboken är oerhört detaljerad och ger upplysningar om enskilda ords alla sidor: uttal, böjning, härkomst, betydelse och ordbildning. Alla uppgifter bygger på autentiska språkprov. Just det faktum att SAOB utnyttjar så många autentiska källor gör att den lämpar sig väl som källmaterial. Merparten av det skriftliga material som finns från gångna tider har utnyttjats av SAOB, vilket innebär att täckningsgraden över tid, genre och texttyp borde vara optimal i förhållande till befintliga källor. Givetvis kan man själv samla in samma material och skapa en textkorpus på egen hand, men problemet med att skapa korpusar av gamla texter är att scanningsmöjligheterna är begränsade på grund av den tryckta textens ojämna kvalitet. Att skriva in texterna för hand innebär att insamlingsfasen blir oerhört lång och kostsam. Fördelen med att använda SAOB som källa till nya delkorpusar är att materialet redan finns tillgängligt, att det bygger på olika typer av källor och att det är kontrollerat flera gånger. Denna "återanvändning" av en redan befintlig korpus gör att man nu kan använda SAOB:s material på ett nytt sätt.

För att visa hur sådana delkorpusar kan användas görs här en jämförelse med tre större undersökningar som bygger på material från 1700-talet.

1. 1700-talskorpusens utformning

Kraven på hur en korpus skall vara sammansatt för att på bästa sätt göra språket rättvisa är stora. En historisk korpus kan emellertid inte bli bättre än vad källäget tillåter eftersom korpusen inte kan innehålla fler texter än de som finns tillgängliga från epoken i fråga. Att det är stor skillnad på antalet källor från 1500-, 1600- och 1700-talet i SAOB visar sig när man söker i den bibliografiska databasen:

- 1500-talet: 114 titlar, 24 namngivna författare (79 % anonyma källor)
- 1600-talet: 640 titlar, 254 namngivna författare (60 % anonyma källor)
- 1700-talet: 1 689 titlar, 714 namngivna författare (58 % anonyma källor)

En närmare undersökning av språkproven från respektive sekel visar att det är vissa årtal som är mer frekvent förekommande än andra, vilket betyder att den kronologiska täckningsgraden är något skev för de äldsta seklen. Man kan också anta att flera språkprov kommer från samma källa. I bild 1 visas språkprovets fördelning över nästan fem sekler uppdelade på 25-årsperioder:

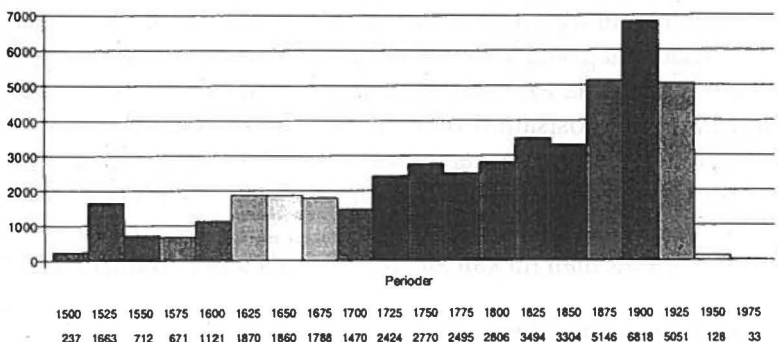


BILD 1. Språkproven i SAOB fördelade på 25-årsperioder.

Vissa perioder innehåller fler språkprov än andra, vilket troligen beror på tillgången till källor från den aktuella tidsperioden. Av denna orsak är perioden 1525–1550 starkt representerad genom bibelöversättningarna från denna tid samt Gustav Vasas registratur. Samma förklaring kan tillämpas på perioden 1900–1925, med den nya bibelöversättningen från 1917.

Vi har valt att arbeta med 1700-talsspråkproven för att få en korpus som innehåller många olika genrer och har en förhållandevis jämn spridning över tiden. Korpusen består av 1 416 020 ord, fördelade på 72 910 språkprov. I ordantalet är inbegripet källhänvisningar och redaktionell, supplerad, text. Tar man bort källhänvisningarna återstår ca 925 000 ord. Genom att räkna bort antalet parenteser, som uppskattningsvis innehåller ca tre ord supplerad text, hamnar man på en summa av 850 000 ord totalt. Dessa utgörs av 207 000 olika ord (*types*).

Materialet är hämtat från 1 689 olika titlar, och det finns 714 namngivna författare. Det skall dock påpekas att alla källhänvisningar inte beledsagas av språkprov. En stor del av beläggen återges bara med källhänvisning.

2. En jämförelse med äldre undersökningar

2.1. Morfologi

För att närmare exemplifiera korpusens möjligheter jämför vi med några äldre undersökningar där 1700-talssvenska ingår antingen som huvudsakligt material eller som delmaterial. Undersökningarna behandlar morfologi, ortografi och lexikon.

Den första undersökningen är Östman (1992) vars huvudsakliga syfte är att undersöka hur korta former av vissa ord introduceras i svenskt skriftspråk, och under vilka språkliga betingelser kortformerna används. Östman menar att kortformerna säkerligen använts i talspråk sedan medeltiden, och att de får insteg i svenskt skriftspråk under nysvensk tid (Östman 1992:12). Somliga former etableras snabbt såsom *mor*, *far* och *bror* medan andra former tar längre tid på sig och används olika beroende på

genre och språkliga omständigheter. Somliga former är ännu inte helt etablerade som kortformer t.ex. *nåt*, *sånt* och *stan* (Östman 1992:11).

Undersökningen spänner över äldre nysvensk tid fram till modern tid. För undersökningen av de äldsta texterna utnyttjar Östman en textbank över nysvenska, sammanställd vid Institutionen för nordiska språk vid Uppsala universitet. Denna uppgår till 300 000 ord totalt. Från 1700-talet finns enbart två texter i textbanken: *Jon Stålhammars brev till hustrun Sofia Drake* (1700–1708) på 44 000 löpord och *Johan Runius prosastycken på svenska* (1710) på 30 000 löpord. Förutom dessa två texter undersöker Östman också Dalins *Then Swänska Argus* (1732–42), 210 000 löpord, samt jämför med den andra upplagan av Argus från 1754. Totalt har Östman således ett material som för 1700-talssvenskans vidkommande uppgår till 284 000 ord, vilket kan jämföras med vårt material på 850 000 ord. Östmans 1700-talstexter är koncentrerade till första hälften av 1700-talet och den senare delen av seklet behandlas inte alls. (Den författare som kommer efter Dalin är Strindberg, vilket alltså ger ett kronologiskt gap på över 100 år.) Vårt material är hämtat från hela perioden.

Även om OSA-korpusen till viss del innehåller material som Östman av olika skäl inte vill ha med som undersökningsmaterial (t.ex. formelartat språk som lag- och bibelspråk, samt genrer som står talspråket alltför nära såsom dramer), så skulle man genom en utrangering av denna typ av källor ändå få en korpus som i storlek och genremässig spridning är större än Östmans och därmed troligen också skulle säga mer om utvecklingen av kortformerna generellt sett.

I tabell 1 ges en enkel uppställning över antalet belägg ur vardera materialet. Man kan tydligt se att OSA-korpusen ger ett bättre utgångsläge för undersökningen genom att framför allt kortformerna är bättre representerade. Också en del långformer, t.ex. *moder* som saknas helt hos Stålhammar och Runius, är bättre belagda i OSA, tillsammans med kortformen *mor*. Även kortformen av best. form. pl. *dar* är talrikare i OSA, och kan

bilda underlag för en grundligare undersökning om utvecklingen av formen.

TABELL 1. *Antal belägg av substantiven dag, stad och moder.*

	Stålhammar	Runius	Dalin	OSA
dagen	–	–	60	167
daggar	15	9	59	151
dan	–	–	–	2
dar	–	1	18	30
staden	–	–	55	167
stan/sta'n	–	–	1	1/3
moder	–	–	8	64
mor	–	2	39	49

På samma sätt är det med de pronomen som återges i tabell 2.

TABELL 2. *Antal belägg av pronomenet er.*

	Stålhammar	Runius	Dalin	OSA
er	9	25	456	267
eder	94	4	30	190

Också här ger OSA-korpusen en god komplettering. Stålhammars flitiga bruk av *eder* förklarar Östman med att ordet ingår som en ritualiserad välgångsönskning i breven (1992:42), vilket är fullt rimligt. Dalins *Argus* innehåller många partier med direkt tilltal till läsaren vilket förklarar hans omfattande användning av *er*. Av denna anledning kan det vara behövligt med ett material som innehåller fler texttypsmönster, vilket OSA gör.

Östman undersöker också flera adverb. I denna undersökning har endast *sedan* och *neder* med respektive kortformer undersökts. Resultaten redovisas i tabell 3. *Sedan/sen* förefaller följa de regler som finns också idag, genom att *sedan* troligen är den mer skriftspråkliga varianten. Utan att ha undersökt det närmare kan man anta att *sen* bara återfinns i de mer talspråkliga texterna.

En jämförelse med OSA visar att *ned* har fler förekomster än *neder*, och att de båda kortformerna *ned* och *ner* tillsammans vida överstiger långformen. På samma sätt är det hos Dalin,

medan de båda andra texterna inte har så många belägg alls. Förmodligen skulle OSA kunna ge en mer differentierad bild av utvecklingen av dessa former.

TABELL 3. *Antal belägg av adverbena sedan och neder.*

	Stålhammar	Runius	Dalin	OSA
sedan	97	45	201	600
sen	1	3	90	64
se'n	—	—	—	18
neder	3	5	8	156
ned	2	6	64	173
ner	—	—	1	56

Vad beträffar verben har vi valt att jämföra endast två verb, *draga* och *säga*, i olika former.

TABELL 4. *Antal belägg av verben draga och säga.*

	Stålhammar	Runius	Dalin	OSA
draga (inf + plur)	—	5	22	125
dra (inf + plur)	2	1	16	16
drager	—	—	4	37
drar	—	—	20	26
sade	—	46	391	118
sa	—	6	6	25

Vad som är tydligt är att den tidsmässiga spridningen för dessa belägg är ringa hos Östman, framför allt vad gäller *draga*. Det är egentligen bara Dalin som har godtagbart många belägg på detta verb. Också här skulle en kompletterande undersökning vara av stort värde.

Östman är intresserad av att kartlägga kortformernas utveckling i förhållande till andra språkliga faktorer som kunnat påverka deras uppkomst och etablering. En sådan faktor är formernas semantiska sida, där bl.a. polysema ord ibland kan ha en form kopplad till en viss betydelse. Ett sådant exempel är *fader/far* där den långa formen, framför allt hos Dalin, reservera-

tas för betydelsen 'andlig fader' och den korta formen för betydelsen 'köttslig far'. För att se om denna fördelning går att skönja också i OSA-korpusen undersökte vi dessa ord närmare.

TABELL 5. *Fördelningen av fader och far.*

RUNIUS		STÅLHAMMAR		DALIN		OSA	
FADER	FAR	FADER	FAR	FADER	FAR	FADER	FAR
K: 6	K: 3	K: 1	K: 9	K: 3	K: 24	K: 55	K: 23
A: -	A: -	A: 4	A: -	A: 29	A: 12	A: 10	A: 1

(K= köttslig far, A= andlig far)

Man kan av siffrorna se att kortformen oftast har betydelsen 'köttslig far', också i OSA-korpusen. Sätillvida stämmer Östmans iakttagelse. Den långa formen, däremot, används också den till övervägande delen om 'köttslig far' i OSA-korpusen, vilket inte stämmer med Östmans antaganden.

Sammanfattningsvis kan man säga att OSA-korpusen ofta bekräftar vad Östman kommit fram till, men att man i flera fall hade nått säkrare resultat om man använt en större korpus, speciellt i de fall då hennes resultat bygger på endast en eller två av 1700-talskällorna. Man hade också fått en bättre uppfattning om utvecklingen under senare hälften av seklet.

När OSA-korpusen blir genreindelad får man dessutom ett bättre utgångsläge när det gäller att se vilka val av kort- respektive långform som är genrebundna.

2.2. Ortografi

Den andra undersökningen vi jämför med är Santesson (1986) som intresserar sig för det inflytande som boktryckare Lars Salvius hade på språknormeringen, speciellt den ortografiska, under sin verksamhet mellan åren 1743 och 1777. Santessons utgångspunkt är en jämförelse av böcker som tryckts i två upplagor, varav den andra upplagan tryckts hos Salvius men den första på ett annat tryckeri. Dessa speciella förutsättningar reducerar hennes undersökningsmaterial till 17 texter, som var och en

utgörs av ca 4 500 ord, vilket ger en totalmängd ord på ca 76 500 ord. Eftersom materialurvalet styrs av utomspråkliga premisser, kan inte OSA-korpusen förbättra källäget. Däremot kan den större korpusen användas för att undersöka om Santessons resultat är giltiga också för andra texter än sådana som är tryckta hos Salvius, och om man kan skönja någon språklig förändring i det senare 1700-talets stavningsnormer som kanske går att förknippa med Salvius insatser. Salvius verkade under mer än 30 år och tryckte många inflytelserika skrifter, t.ex. Vetenskapsakademins handlingar, så det är inte helt orimligt att tänka sig att vissa av Salvius ortografiska principer slog igenom på större bredd.

Santesson undersöker en mängd olika ortografiska förändringar, men vi koncentrerar oss huvudsakligen på stavningen av två olika ljud representerade i ett och samma ord: *människa*. Genom att undersöka just detta ord kan man dels se hur stavningen av ä-ljudet varierar, men också stavningen av sje-ljudet. Stavningen av /ä/ är snarast lexikalt betingad enligt Santesson (1986:144) I hennes material kan man tydligt se att det sker en förändring från stavning med <e> i första upplagan till stavning med <ä> i andra upplagan. Hennes material uppgår till 117 belägg, inklusive avledningar, och i 101 av dessa belägg sker en förändring från stavning med <e> till <ä>. Vad som intresserar oss är i första hand om man kan se att det också fortsättningsvis, efter 1777, är en högre förekomst av stavning med <ä>.

Sökningen i OSA-korpusen gav totalt 423 belägg, varav 21 var avledningar. Två av beläggen har endast källangivelsen DENS. (densamme), så de går inte att tidsfästa eftersom språkproven i korpusen inte kommer i samma ordning som i SAOB. Totalt behandlas således 404 belägg.

I materialet syns tydligt att stavningen av sje-ljudet ändras från början till slutet av seklet (se tabell 5). Under första hälften av seklet dominerar stavningen med <i>-inskott, *menniskia*, för att mot slutet av seklet övergå till den stavning av sje-ljudet vi känner idag i detta ord: *menniska*. Man kan också se att stavningen av sje-ljud med <i>-inskott oftare hör samman med stavning av ä-ljudet med <e> (*menniskia*). Utvecklingen av ortografin för sje-ljudet i OSA-beläggen bekräftar Santessons resultat och hör

samman med en allmän utveckling av ortografin under 1700-talet.

Vad gäller stavningen av ä-ljudet dominerar <e>-stavning i OSA-korpusen. Stavning med <ä> ökar i frekvens mot slutet av seklet, men blir aldrig helt dominerande. Man kan således säga att <ä>-stavningen hör samman med senare delen av 1700-talet, men att man inte kan bekräfta att <e>-stavning övergår till <ä>-stavning som hos Salvius. Tabell 5 visar dock att ett skifte mellan de olika stavningarna sker mot mitten av seklet, även om ingen stavningsvariant kan sägas bli den dominerande.

Det finns dock några författare i OSA-korpusen som växlar mellan stavning med <e> och <ä>. Olof von Dalin har 16 belägg med <e> och två med <ä>. De sistnämnda återfinns i Vetenskapsakademiens handlingar, som ju trycktes hos Salvius. Man får därför anta att Dalin själv stavade *menniska* men att Salvius ändrade två förekomster till *människa*.

Ett intressant fall är Johan Henrik Kellgren som totalt sett har 13 belägg i materialet varav sex med <ä>-stavning. Alla beläggen är från samma decennium, men variationen i stavning är stor. Beläggen fördelar sig på följande sätt:

menniska: 1781, 1785, 1787 (4 st), 1793;

människa: 1781, 1782, 1787, 1789;

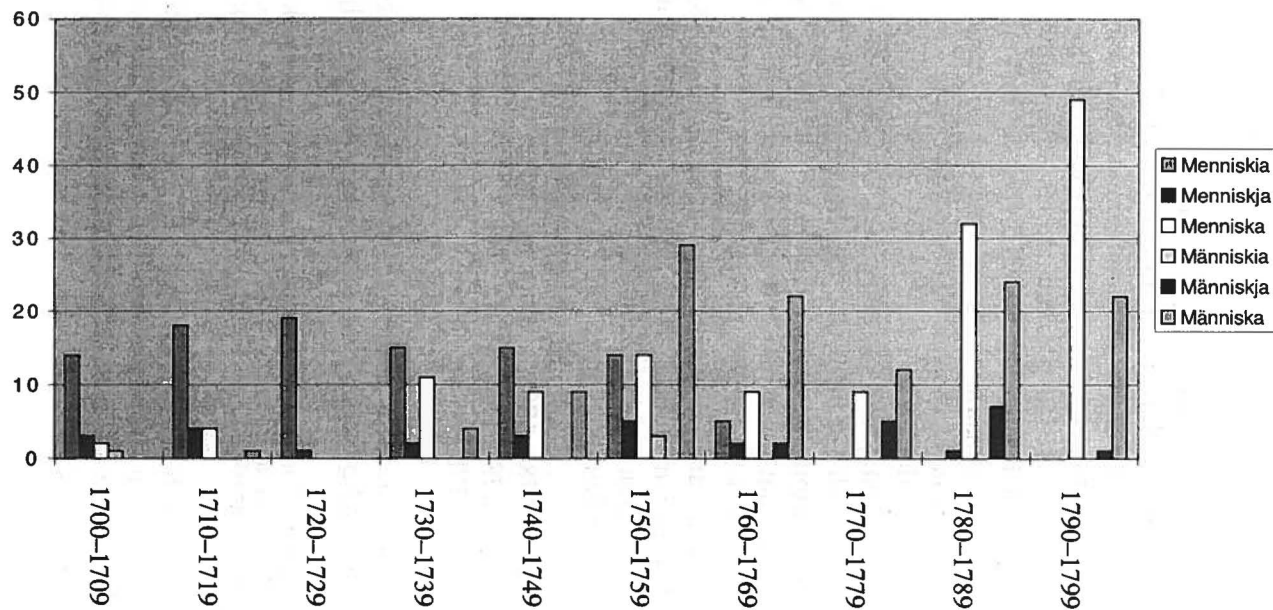
människja: 1782, 1787.

Två av exemplen (*människa* och *människja*) förekommer i samma språkprov och visar varierad stavning av sje-ljudet.

Arbete är ingen olycka, det retar til matlust och underhåller hälsan.
Dessutom är alt reciprokt. Om vi arbeta för människjan, arbetar
och människan för oss. KELLGREN (SVS) 4: 279 (1782).

Det förefaller således som om Kellgren varierar stavningen av såväl /ä/ som /j/, något man skulle kunna undersöka vidare i OSA-korpusen genom att skapa en korpus med språkprov från Kellgrens skrifter.

TABELL 6. Översikt över stavning av sje-ljud och kort ä-ljud.



Höpken verkar gå i motsatt riktning gentemot den antagna utvecklingen genom att han har <ä>-stavning under 1740-talet, men <e>-stavning i de två senare beläggen (*människa*: 1740 (2 st.), 1745 (2 st.); *menniska*: 1755, 1777). Förutom Kellgren och Höpken är det åtta författare som varierar stavningen, vilket inte är särskilt många. Alla har fler exempel på <e>-stavning än <ä>-stavning.

Materialet går således att använda för att se om vissa forskningsresultat rörande en del av epoken är giltiga också för andra skeden, och resultatet av denna undersökning visar att Salvius påverkat stavningen under sin verksamma tid, men att påverkan inte kan sägas ha uppkommit efter hans verksamma tid.

2.3. Lexikon

Många av SAOB:s språkprov är hämtade från andra ordböcker. En sådan källa är Serenius engelsk-svenska ordbok från 1734, som bidragit med över 1 600 av SAOB:s förstabelägg. Serenius ordböcker undersöks i Rogström (1998), där bl.a. förstabeläggen behandlas ingående. Ett skäl till att undersöka dem närmare är för att se vilken typ av ord som kan tänkas införlivas i ett språk med hjälp av tvåspråkiga ordböcker, speciellt om dessa, som Serenius, är de första av sitt slag i ett visst land. Av samma orsak är det viktigt att undersöka om orden verkligen användes av samtidens språkbrukare eller om de bara återfinns i ordböckerna. En dylik undersökning går bara att genomföra med hjälp av en textkorpus från den aktuella tiden, varvid OSA-korpusen passar bra.

En undersökning av detta slag kan också vara en hjälp för det rent redaktionella arbetet på SAOB, genom att man enkelt kan kontrollera om ett visst ord verkligen är det äldsta belägget, eller om det redan finns i ordboken i ett annat, äldre språkprov. Man kan också kontrollera om man varit konsekvent i återgivandet av vissa uppgifter. (I undersökningen ovan framkom att Bælter var en av de författare som växlat i stavningen av *e* och *ä*, men en

kontroll visar att det rör sig om en boktitel, *Guds sons människio-blifvande*, som återgivits på två olika sätt i SAOB.)

Av de ord som SAOB angivit som förstabelägg ur Serenius (1734) har vi undersökt följande (här återgivna med normaliserad stavning): *ambassad*, *adventstid*, *barnbok*, *disk* (i affär), *docka* (för skepp), *flanellskjorta*, *fluglort*, *gräddost*, *harpa* (om kvinna), *hotfull*, *instrumentalmusik*, *intrig* (i pjäs), *kaffeböna*, *knapadel*, *knekt* (i kortspel), *kola av* ('dö'), *krona eller klave*, *köttkniv*, *klöver* (kortspelsfärg), *lilltå*, *lösmynt*, *metafysik*, *mödomshinna*, *republikan*, *ruter*, *rökare*, *skela*, *skogshuggare*, *snabel* (på elefant), *spader*, *sparka fotboll*, *statsbesök*, *storrökare*. Orden som undersökts är med något undantag tämligen vardagliga ord och heller inte speciellt märkliga eller lågfrekventa ord. Av de undersökta orden återfanns endast *ambassad* (1740, 1791), *instrumentalmusik* (1770), *kaffeböna* (1753), *knekt* (1781), *klöver* (1739, 1745 (3ggr), 1781, 1790), *metafysik* (1719, 1790), *ruter* (1745, 1775, 1790), *rökare* (1746) och *spader* (1745, 1790). Det enda ord som har ett äldre belegg än 1734 är *metafysik*, som förekommer redan 1719. Detta belegg avser dock en bok av Aristoteles med titeln *Metaphysica*, och kan därmed inte räknas som ett tidigare förstabelägg på detta ord i svenskan.

Snabel återfanns inte i betydelsen 'elefantsnabel', men väl i flera andra betydelser av 'nos' e.d. på andra djur: bin, kalkon, mullvad och pelikan. *Statsbesök* fanns inte heller, men väl 47 andra sammansättningar med förleden "stat", så *statsbesök* passar väl in i ordbildningstypen.

Det är något förvånande att inte fler ord återfanns i OSA-korpusen. En förklaring kan vara att de kanske är för vardagliga för att finnas med i de källor från denna tid som SAOB utnyttjat. Även om genrespridningen är stor, är det trots allt ändå en brist på texter som behandlar vardagliga, triviala ting, vilket är en följd av källäget. Det saknas helt enkelt dylika texter från äldre tider.

3. Sammanfattning

OSA-databasen är viktig för språkhistorisk, lexikologisk forskning. Genom att konstruera delkorpusar av olika slag kan man utnyttja det rikhaltiga materialet för mer synkrona undersökningar och för pilotundersökningar av skilda slag, vilket detta föredrag velat belysa genom en jämförelse med tre äldre undersökningar av 1700-talssvenska.

Databasen erbjuder dessutom stora möjligheter till kontroller av det rent redaktionella arbetet med ordboken genom att man på ett enkelt sätt kan se hur enskilda uppgifter återgivits och formulerats i de äldre delarna. Man får också en möjlighet att kontrollera att inga äldre belägg av enskilda språkprov förekommer på annan plats i ordboken.

Litteratur

- Cederholm, Yvonne. 1996. "A Historical Lexical Database of Swedish. The O.S.A Project." *EURALEX '96 Proceedings*. S. 65–72.
- Rogström, Lena. 1998. *Jacob Serenius lexikografiska insats*. (Meijerbergs arkiv för svensk ordforskning. 22.) Göteborg.
- Santesson, Lillemor. 1986. *Tryckt hos Salvius*. (Lundastudier i nordisk språkvetenskap. A 37.) Lund.
- Serenius, J. 1734. *Dictionarium Anglo–Svethico–Latinum*. Hamburg.
- Östman, Carin. 1992. *Den korta svenskan. Om reducerade ordformers inbrytning i skriftspråket under nysvensk tid*. (Skrifter utgivna vid Institutionen för nordiska språk vid Uppsala universitet. 30.) Uppsala.

Ulla Clausén

Funktion och attityd i Svensk konstruktionsordbok

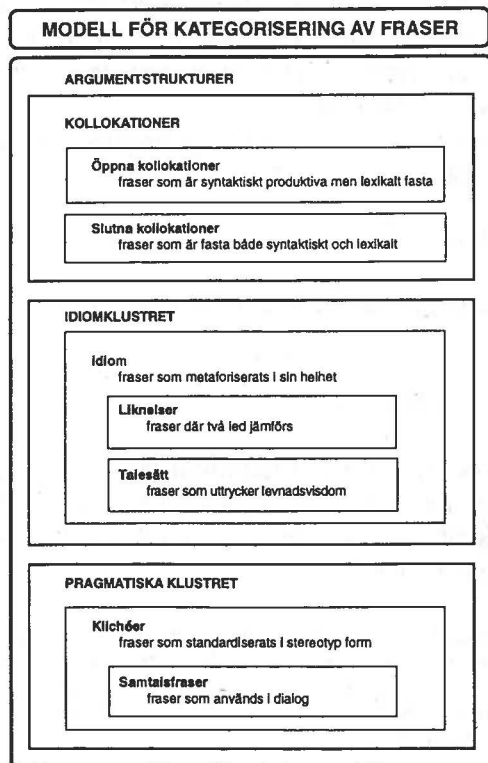
The purpose of this paper is to discuss functional use of the different categories of phrases in the forthcoming phraseological dictionary Svensk konstruktionsordbok. Two primary functions are introduced, *informational function* and *pragmatic function*, and three secondary functions, *comparative function*, *evaluative function* and *modal function*. The paper also deals with the relationship between functions and attitudes. A model for analysis of functions used in the dictionary is introduced.

1. Inledning

Syftet med den här artikeln är att undersöka vilka kommunikativa funktioner som kan förekomma hos de kategorier av fraser som ingår i Svensk konstruktionsordbok, en fraseologisk ordbok som sedan en tid är under arbete vid Svenska språknämndens sekretariat i Stockholm. Samtidigt vill jag utreda sambandet mellan frasernas funktioner och de eventuella attityder som användaren vill uttrycka. Exempel kommer att ges på hur funktioner och attityder markeras och kommenteras i konstruktionsordboken. Jag har tidigare undersökt funktionerna hos den kategori som jag benämner **klichéer** (Clausén 1999). Jag ska nu göra en fördjupad undersökning och utvidga analysen till att omfatta samtliga kategorier av fraser i ordboken.

Materialet till undersökningen har hämtas från olika databaser, bl.a. Språkdatas konkordanser. Detta material har kategoriserats enligt en tidigare utarbetad och nu modifierad modell för kategorisering av fraser i ordboken (figur 1). Av denna modell framgår det att ordboken kommer att innefatta tre huvudkategorier av fraser: **kollokationer**, **idiom** och **klichéer**. Dessa kategorier vilar på en underliggande bas av **argumentstrukturer**, dvs. syntaktiska mönster. Kategorin kollokationer innefattar dels fraser som är syntaktiskt produktiva men lexikalt fasta (**öppna kollo-**

kationer), t.ex. *vackert väder, stänga grinden*, dels fraser som är fasta både syntaktiskt och lexikalt (**slutna kollokationer**), t.ex. *nautisk mil, köra bil*. Idiomen ingår tillsammans med undergrupperna **liknelser** och **talesätt** i **idiomklustret**. Alla fraser som ingår i idiomklustret är i sin helhet metaforiserade, vilket innebär att de ingående ordens konkreta betydelse glidit över till en bildlig nivå (Clausén & Lyly 1995:26f.), t.ex. *ha vind i seglen* i betydelsen 'ha framgång'. Liknelser är fraser där två led jämförs, t.ex. *komma som ett brev på posten* 'komma oundvikligt'. Talesätt är fraser som uttrycker levnadsvisdom, t.ex. *man ska inte låta snålheten bedra visheten*. Kategorin klichéer innefattar fraser som standardiserats i en stereotyp form. De ingår med undergruppen **samtalsfraser** i **pragmatiska klustret** (Clausén 1999).



FIGUR 1. Modell för kategorisering av fraser.

2. Huvudfunktioner

I min föregående undersökning av klichéernas funktioner (Clausén 1999) visade jag att det finns två **huvudfunktioner** hos fraserna: **informativ funktion** och **pragmatisk funktion** (jfr Halliday 1970:140f). Varje fraskategori har *en* huvudfunktion till vilken kommer olika **bifunktioner** (se vidare nedan; jfr Moon 1998:217f.).

2.1. Informativ funktion

Fraser med informativ funktion används för att informera. Fraser med denna funktion integreras i en sats som nominalfras, verbfras eller prepositionsfras. Jag har kommit fram till att argumentstrukturerna, kollokationerna och idiomklustrets fraser har informativ funktion (Clausén 1999:29).

Argumentstrukturen *skära ngt [med ngt]* har sålunda informativ funktion i följande exempel:

Hon skar brödet med en brödsåg.

Också kollokationer har informativ funktion, t.ex. *stänga grinden* i exemplet:

Han stängde grinden efter sig.

Jag anser att också fraserna i idiomklustret har informativ funktion, som t.ex. idiomet *lämna ngn vind för våg* med betydelsen 'lämna ngn åt sitt öde'. Det framgår av följande text som hämtats från Språkdatäs konkordanser:

I själva verket har multikulturalismen helt misslyckats med att göra invandrarna till medborgare. Den har *lämnat* de nyanlända *vind för våg* att klara sig själva och kämpa för att lära sig behärska nya egendomliga seder och bruk och ett förbryllande komplicerat samhälle. (Press 96)

Även liknelser, t.ex. *komma som ett brev på posten* med betydelsen 'komma oundvikligt', har informativ funktion, vilket framgår av följande exempel från Språkdatas konkordanser:

Han har hamnat i ett läge han har mycket svårt att ta sig ur. Vägrar han betala *kommer* straffavgiften från kronofogden *som ett brev på posten*. (Press 97)

Talesätten har också informativ funktion, t.ex. *snålheten bedrar visheten*, en variation av *man ska inte låta snålheten bedra visheten*, vilket framgår av följande text, hämtad från Språkdatas konkordanser:

Frågan är dock om svensken blivit så mycket godare. Kanske har vi bara, i den hårda verklighetens ljus, vänt på det här med att *snålheten bedrar visheten* och insett att ibland är det vist att vara snål. (Press 96)

2.2. Pragmatisk funktion

I min tidigare undersökning av klichéer (Clausén 1999:29f.) kom jag fram till att fraser med pragmatisk funktion inte tillför någon egentlig ny information utan används för att i stereotypa ordalag uttrycka åsikter och känslor. Utmärkande för dem är också att de utgör självständiga enheter som i regel inte integreras i texten utan fungerar som hela block. Min undersökning visade att fraserna inom det pragmatiska klustret uppvisar olika slag av pragmatisk funktion: **kommenterande funktion, strukturerande funktion, samtalsförande funktion, social funktion, influerande funktion** och **expressiv funktion**.

Fraser med kommenterande funktion finns hos den typ av klichéer som jag kallar **klyschor**, dvs. fraser av typen *går det så går det!* De ligger nära talesätten (som har informativ funktion), men skiljer sig från dessa genom att de fungerar som fristående block. Den typ av klichéer som jag kallar **textstrukturerande fraser** har strukturerande funktion, t.ex. *när allt kommer omkring*. Fraser av denna typ bör skiljas från vanliga adverbiala led med informativ funktion, t.ex. *till slut*. Med den textstrukture-

rande frasen sammanfattar användaren sina subjektiva åsikter utan att tillföra någon ny information. En specialgrupp inom det pragmatiska klustret är **samtalsfraser**. De fraser som hör hit är alla interaktiva, vilket förutsätter en närvarande samtalspartner. Samtalsförande funktion finns också hos den typ av samtalsfraser som jag benämner **konversationsfraser**, nämligen fraser av typen *säg som det är!* och *det menar du inte!* Fraser av denna typ tillför ingen ny information utan används i samtalet för att ange hur någon ställer sig till samtalspartnerns meddelande. Social funktion kännetecknar fraser av typen hälsningar och avskedstaganden, tack och ursäkter som *det var roligt att träffas*. Jag kallar dem **hövlighetsfraser**. Influerafunktion finns hos fraser av typen **hotelser**, **varningar** och **förbud**, t.ex. *det där ska du få ångra!* Expressiv funktion har t.ex. svordomar som *djäv-lars blod och änkors klagan!*

2.3. Från informativ funktion till pragmatisk funktion

Min analys av materialet till konstruktionsordboken har visat att fraser med informativ funktion genom en dynamisk process kan övergå till fraser med pragmatisk funktion. Jag har tidigare påpekat (Clausén 1999:32f.) att en kollokation med informativ funktion kan användas som kliché. Kollokationen *ta till överord* kan t.ex. användas som textstrukturerande fras i formen *det är inte att ta till överord*. Fortsatta analyser av frasmaterialet har visat att fraser i idiomklustret kan få pragmatisk funktion. Talesätt med informativ funktion, t.ex. *man ska inte låta snålheten bedra visheten*, kan sålunda övergå till en klyscha med kommenterande funktion. Detta framgår av följande exempel, hämtat från Språkdatas konkordanser:

Det börjar gå lite för långt med de slimmade organisationerna: *man ska inte låta snålheten bedra visheten!* (Expressen 18/12 1996)

Talesättet kan varieras, men har fortfarande pragmatisk funktion:

Självklart ska personer som kräver specialistvård få det. Alla far ju illa, både patienter och personal. Snålheten bedrar visheten.
(Expressen 26/7 1997)

3. Bifunktioner

Min genomgång av fraserna i materialet visar att det finns tre bifunktioner: **jämförande funktion**, **värderande funktion** och **modal funktion**. Utifrån detta resultat kan jag nu lansera en modell för analys av de kommunikativa funktionerna i Svensk konstruktionsordbok (figur 2). Av denna modell framgår att alla tre bifunktionerna kan förekomma hos samtliga kategorier av fraser. Det är viktigt att notera att hos vissa kategorier kan funktionerna vara obligatoriska, men icke-obligatoriska hos andra. När de är icke-obligatoriska står de inom klammer i modellen (figur 2).

3.1. Jämförande funktion

När en fras har jämförande funktion används antingen ett enda ord eller en hel fras bildligt. Jämförande funktion är sålunda en annan benämning på metaforisering. Metaforisering som avser ett enda ord förekommer hos argumentstrukturerna och kollokationerna. Metaforisering som gäller hela frasen förekommer först och främst hos idiomerna, men ibland också hos klichéerna.

3.1.1. Jämförande funktion med avseende på ett enda ord

Genom metaforisering kan ett enstaka ord i en argumentstruktur eller kollokation få bildlig användning. Verbet *skära* i argumentstrukturen *skära ngt [med ngt]* har sålunda konkret betydelse i exemplet:

Hon skar brödet med en brödsåg.

Verbet *skära* har däremot **bildlig** betydelse ('gå igenom') i exemplet:

Bilen skar kurvan i hög fart.

Den jämförande funktionen är här icke-obligatorisk. Också hos kollokationerna kan ett enstaka ord metaforiseras. I frasen *djup snö* har sålunda *djup* konkret betydelse, men bildlig med betydelsen 'ingående' i *djup kunskap*.

3.1.2. Jämförande funktion med avseende på hela frasen

Hos idiomerna innebär den jämförande funktionen att hela frasen är metaforiserad, dvs. de ingående ordens konkreta betydelse har glidit över till en bildlig nivå (Clausén & Lyly 1995:26f.). Frasen *lämna ngn vind för våg* används t.ex. om att lämna någon åt sitt öde (se ovan 2.1. Informativ funktion). De ingående orden får alltså tillsammans en annan betydelse än orden var för sig. En annan och vanligare typ av idiom är den där en kollokation metaforiserats och övergått till idiom genom en dynamisk process. Kollokationen *stänga grinden* anger t.ex. i överförd betydelse att man utesluter el. stänger ute någon, vilket framgår av följande exempel, hämtat från Språkdatas konkordanser:

Vid varje ekonomisk kris i historien har Europa reagerat med att *stänga sina grindar* samt med att massdeportera sin egen underklass. (Press 97)

Den jämförande funktionen är obligatorisk hos fraserna i idiomklustret. Den är explicit hos liknelserna, där två led jämförs med markören *som*, t.ex. *komma som ett brev på posten*. Hos övriga fraser i idiomklustret är den jämförande funktionen implicit.

Som framgår av Modell för analys av funktioner (figur 2) kan också klichéerna ha jämförande funktion, men den är inte obligatorisk (står inom klammer i modellen). Det är t.ex. inte ovanligt att svordomar är metaforiserade och sålunda har jämförande funktion, t.ex. *djävls blod och änkors klagan!* Det som skiljer

klichéerna från idiomerna är alltså huvudfunktionen. Idiomerna har informativ funktion medan klichéerna har pragmatisk funktion.

3.2. Värderande funktion

Jag har kommit fram till att den värderande funktionen, precis som den jämförande, kan förekomma hos samtliga fraskategorier i ordboken. (Se figur 2.) Den är obligatorisk hos fraserna i idiomklustret och pragmatiska klustret. Idiomet *lämna ngn vind för våg* innebär t.ex. att man är kritisk till att någon lämnar någon åt sitt öde. Också hos klichéerna är den värderande funktionen obligatorisk. I en klyscha som *vi ska alla den vägen vandra*, i en konversationsfras som *det menar du inte!* och i en varning som *passa dig jävligt noga!* är den värderande funktionen uppenbar.

Hos argumentstrukturerna och kollokationerna är den värderande funktionen däremot icke-obligatorisk (står inom klammer i figuren). När värderingar förekommer här är de explicita. I t.ex. kollokationerna *vackert väder* och *dåligt väder* framgår värderingarna tydligt av adjektiven *vackert* och *dåligt*.

Det bör emellertid framhållas att en fras som saknar inbyggda värderingar kan vara värderande i kontexten. Jag benämner sådana värderingar **texts specifika** (Clausén 1996:39f.) i motsats till de värderingar som finns inbyggda i frasen, de **lexikon specifika**. Den ursprungligen icke-värderande kollokationen *stänga grinden* har t.ex. värderande funktion i exemplet:

Hon stängde grinden mitt framför näsan på mig.

3.3. Modal funktion

Man brukar tala om två typer av modalitet. Den ena används för att uttrycka hotelser, varningar och förbud. Denna typ av modalitet kallas *deontisk* (från ett grekiskt ord för 'nödvändig'; jfr. Palmer 1986:96f.). Den andra typen används för att uttrycka säkerhet eller tvivel, vaghet och möjlighet. Denna form av mo-

dalitet kallas *epistemisk* (från ett grekiskt ord för 'kunskap'; jfr Palmer 1986:51f., Stubbs 200f.).

Modalitet, dvs. modal funktion, kan vara både lexikonspecifik och textspecifik (Clausén 1996:40f.). Hos argumentstrukturerna och kollokationerna är den textspecifik och kommer fram först i kontexten. Argumentstrukturen *skära ngt [med ngt]* har sålunda modal funktion i exemplet:

Du bör inte skära brödet med en bordskniv.

Även en kollokation som *stänga grinden* kan ha textspecifik modal funktion. Det framgår av exemplet:

Du måste stänga grinden efter dig!

Den modala funktionen är ofta lexikonspecifik hos fraserna i idiomklustret. Här är det först och främst talesätten som uttrycker modalitet. Den modala funktionen är här obligatorisk. I talesättet *man ska inte låta snålheten bedra visheten* är det sålunda *man ska inte* som uttrycker modaliteten explicit. Denna modala funktion är deontisk, dvs. den anger hur man bör bete sig. Talesättet kan varieras som *snålheten bedrar visheten*. Här är modaliteten implicit. En annan typ av talesätt är *många bäckar små gör en stor å*. Här är modaliteten epistemisk, dvs. den anger en möjlighet. Den är implicit. Genom en dynamisk process kan talesätt med deontisk modalitet förlora sin status av explicit modal funktion och på så sätt inte längre tillhöra specialgruppen talesätt bland idiomerna. Modaliteten försvagas och frasen får en annan innebörd. I botten av idiomet finns dock modaliteten. Talesättet *man ska inte låta snålheten bedra visheten* kan t.ex. slopa den explicita modaliteten och fungera som idiom:

Om löntagarorganisationerna skulle ge efter för SAF:s krav på lönestopp eller mycket små löneökningar, får organisationen säkert snart anledning att ångra sig. Effekten skulle bli större bolagsvinster än vad som är acceptabelt och i praktiken försämrade standard för löntagarna. Men även arbetsgivarna skulle finna att *de låtit snålheten bedra visheten*. (Arbetet 26/11 1998)

Lexikonspecifik modalitet finns sålunda alltid hos den typ av idiom som härrör från talesätten. Den är implicit. Liknelserna uttrycker däremot i regel ingen modalitet. Den modala funktionen är sålunda icke-obligatorisk i idiomklustret (inom parentes i modellen; se figur 2).

Jag har kommit fram till att samtliga typer av fraser i pragmatiska klustret har modal funktion. Den är obligatorisk. De textstrukturerande fraserna har sålunda epistemisk modalitet, t.ex. *när allt kommer omkring*, som innebär ett medgivande. Samtalsfraserna med olika undergrupper har alla deontisk modalitet, t.ex. hotelser, varningar och förbud. I en typ av fraser finns både deontisk och epistemisk modalitet, nämligen hos samtalsfraserna, som kan uttrycka både tvång som *håll tyst!* och tvivel som *det kan inte vara sant!*

Eftersom den modala funktionen redan finns i pragmatiska klustret är det lätt att låta talesätt med informativ modal funktion övergå till pragmatisk funktion. Talesätt med modaliteten explicit formulerad används sålunda ofta med pragmatisk funktion, t.ex. *man ska inte låta snålheten bedra visheten*. (Se ovan 2.2. Pragmatisk funktion.) Även talesätt med implicit modal funktion fungerar med fördel som kliché. Talesättet *många bäckar små gör en stor å* kan sålunda fungera som konversationsfras, vilket framgår av följande samtal som avlyssnats i radion:

- Det kan väl inte ge så mycket pengar till kommunerna att sälja ut de här lägenheterna.
- Många bäckar små. (P1 15/2 1999, kl. 7)

4. Modell för analys av funktioner

Utifrån den analys jag gjort av fraser tillhörande olika kategorier, som *skära ngt [med ngt]*, *stänga grinden*, *lämna ngn vind för våg* och *man ska inte låta snålheten bedra visheten* har jag ställt upp en modell för analys av funktioner hos olika kategorier av fraser (figur 2).

MODELL FÖR ANALYS AV FUNKTIONER I SVENSK KONSTRUKTIONSORDBOK				
KATEGORI	HUVUDFUNKTION	BIFUNKTIONER		
ARGUMENT-STRUKTURER	INFORMATIV	[JÄMFÖRANDE] enskilt ord	[VÄRDERANDE]	[MODAL]
KOLLOKATIONER	INFORMATIV	[JÄMFÖRANDE] enskilt ord	[VÄRDERANDE]	[MODAL]
IDIOM	INFORMATIV	JÄMFÖRANDE hela frasen	VÄRDERANDE	[MODAL]
KLICHÉER	PRAGMATISK	[JÄMFÖRANDE] hela frasen	VÄRDERANDE	MODAL

FIGUR 2. Modell för analys av funktioner i Svensk konstruktionsordbok.

5. Funktion och attityd i ordboken

Svensk konstruktionsordbok är avsedd att användas som en produktionsordbok. Därför behövs anvisningar om bruket, dvs. kommentarer om frasernas funktion och om vidhängande attityder.

Den jämförande funktionen markeras alltid i ordboken. Den som innebär metaforisering av ett enstaka ord hos en argumentstruktur eller kollokation markeras med **bildligt** framför exemplet:

bildl. bilen skar kurvan i hög fart.

Den som innebär metaforisering av hela frasen förekommer alltid hos idiomerna. Hos idiom av typen *stänga grinden* markeras då alltid dynamiken mellan kollokationen och idiomet med **överfört**. Detta framgår av följande utdrag ur ordboksartikeln *grind*, där frasen *stänga grinden* ingår:

stänga grinden hon stängde ~en mitt framför näsan på mig äv. överfört om att utesluta el. stänga ute vid varje ekonomisk kris i historien har Europa reagerat med att stänga sina ~ar samt med att massdeportera sin egen underklass

I frasen *stänga grinden* ovan ges först ett exempel på den konkreta betydelsen: "hon *stängde grinden* mitt framför näsan på mig". I exemplet finns visserligen en textspecifik värdering varmed skribenten eller talaren uttrycker en negativ attityd, men den är självklar och kommenteras inte i ordboken. Den jämförande funktionen anges med kommentaren "äv. överfört om att utesluta el. stänga ute" och exemplifieras med "vid varje ekonomisk kris i historien har Europa reagerat med att *stänga sina grindar* samt med att massdeportera sin egen underklass".

Vidare är det viktigt för användaren att känna till vilken attityd som uttrycks med en viss fras, t.ex. om attityden är positiv eller negativ. Min analys av funktionerna hos olika fraskategorier har visat att värderingar alltid är kopplade till den jämförande funktionen och till den modala funktionen (jfr Moon 1998: 244). Genom dessa värderingar ger användarna uttryck för sina attityder. Hos argumentstrukturer och kollokationer är värderingarna explicita, och också attityderna, som hos *vackert väder* och *dåligt väder*. De kräver i regel ingen speciell kommentar.

Hos idiomerna och klichéerna däremot är användarens attityd ofta implicit och behöver lyftas fram. Detta gäller t.ex. den typ av idiom som uppkommit genom en övergång från talesätt till idiom. Hos idiom av denna typ finns nämligen en implicit modalitet. Med denna följer en implicit attityd. Denna attityd måste på något sätt markeras. Genom att dynamiken mellan talesätt och idiom betonas så att idiomerna framstår som en förkortning av talesättet, får användaren klart för sig att den attityd som anges för talesättet också gäller idiomerna. Användaren får också veta att samma attityd också gäller när talesättet fungerar som kliché. Detta framgår av följande exempel från ordboksartikeln *snålhet*, där talesättet *man ska inte låta snålheten bedra visheten* ingår:

man ska inte låta snålheten bedra visheten markerar ngns åsikt att det är oklokt att låta sig ledas av strikt ekonomiska hänsyn, ofta avkortat skulle beslutsfattarna ändå låta ~en bedra visheten och framhärda med Rid-darholmslösningen föreslår vi att de minimerar merkostnaderna; om löntagarorganisationerna skulle ge efter för kraven på lönestopp skulle också arbetsgivarna finna att de låtit ~en bedra visheten; äv. som ironisk kommentar det börjar gå lite för långt med de slimmade organisationerna. Man ska inte låta ~en bedra visheten äv. varierat *snålheten bedrar visheten* kanske har vi i den hårda verklighetens ljus vänt på det här med att ~en bedrar visheten och insett att det ibland är vist att vara snål äv. som ironisk kommentar det blev sorgligt misslyckat när försvaret köpte nya billiga däck till planen. ~en bedrar visheten

Attityden och funktionen hos talesättet *man ska inte låta snålheten bedra visheten* anges med "markerar ngns åsikt att det är oklokt att låta sig ledas av strikt ekonomiska hänsyn". Den avkortade formen som är ett idiom, *låta snålheten bedra visheten*, är den vanliga formen. Den ges i exempelform med kommentaren "ofta avkortat" framför de två exemplen i olika tempus: "skulle beslutsfattarna ändå *låta snålheten bedra visheten* och framhärda med Rid-darholmslösningen föreslår vi att de minimerar merkostnaderna" och "om löntagarorganisationerna skulle ge efter för kraven på lönestopp skulle också arbetsgivarna finna att de *låt tit snålheten bedra visheten*". Talesättets attityd och funktion gäller alltså också för idiomet.

Talesättet kan också ha pragmatisk funktion och fungera som en klyscha. Detta anges med "äv. som ironisk kommentar" vid exemplet "det börjar gå lite för långt med de slimmade organisationerna. *Man ska inte låta snålheten bedra visheten*". Det är också vanligt att talesättet *man ska inte låta snålheten bedra visheten* varierar som *snålheten bedrar visheten*. Detta exemplifieras med "kanske har vi i den hårda verklighetens ljus vänt på det här med att *snålheten bedrar visheten* och insett att det ibland är vist att vara snål vilket kommenteras med "äv. varierat". Här är modaliteten implicit och också attityden, men upprepas inte eftersom detta anges vid det fulla talesättet. Även det varierade talesättet kan ha pragmatisk funktion. Detta anges med "äv. som ironisk kommentar" vid exemplet "det blev sorgligt miss-

lyckat när försvaret köpte nya billiga däck till planen, *snålheten bedrar visheten*".

Litteratur

- Clausén, Ulla. 1996. "Idiom i bruk." *Språket lever! Festskrift till Margareta Westman. (Skrifter utgivna av Svenska språknämnden. 80.)* S. 36-42.
- Clausén, Ulla. 1999. "Pragmatiska fraser i Svensk konstruktionsordbok." Peter Slotte, Pia Westerberg, Eva Orava (utg.): *Nordiska studier i lexikografi. 4.* Helsingfors: (Skrifter utgivna av Nordiska föreningen för lexikografi. 5.) S. 27-35.
- Clausén, Ulla & Lyly, Erika. 1995. "Idiom och lexikografi." *Språkvård* 1995:1. S. 23-30.
- Halliday, M.A.K. 1970. "Language Structure and Language Function." *New Horizons in Linguistics*. Edited by John Lyons. S. 140-165.
- Moon, Rosamund. 1998. *Fixed expressions and idioms in English*. Oxford: Clarendon Press.
- Palmer, F.R. 1986. *Mood and modality*. Cambridge: Cambridge University Press.
- Stubbs, Michael. 1996. *Text and Corpus Analysis*. Oxford: Blackwell.

Automatisk extrahering av ordleder

A typical trait of Swedish word formation are words which are productive in word formation. They are not bound morphemes because they are either root morphemes or composites. They are not lexemes because they are used in a similar way as bound morphemes. These bound lexemes, as they are called here, constitute a crucial part of a compounding dictionary, particularly of a bilingual one. Potential bound lexemes are traced and automatically selected from a large text corpus of Swedish. Two parameters are established to make the extraction precise: compounding ability and compounding productivity. The first one provides measure of the relation between occurrences of a word as compound component and as an independent word. The second one is based on how many collocates it cooccurs with in compounds.

Inledning

Den nedan beskrivna korpusundersökningen av svenska texter har främst ett praktiskt syfte – att extrahera material till ett aktivt sammansättningslexikon. Det övergripande målet är att underlätta för även icke infödda talare av svenska att förstå, översätta och skapa nya sammansättningar. Arbetet ingår i projektet ”Svenska sammansättningar och deras ekvivalenter i polska”, som huvudsakligen behandlar verbala sammansättningar – dessa saknas nästan helt som produktiva mönster i slaviska språk. Projektet finansieras av HSFR och bedrivs på Institutionen för slaviska språk vid Göteborgs universitet.¹

Vid extraheringen användes texter från Språkbanken vid Göteborgs universitet (www.svenska.gu.se): 42 580 498 löpord, varav 33 245 519 från dagstidningar, och 9 337 979 annan prosa. Extraheringen genomfördes helt automatiskt med LexWare, ett korpusverktyg utvecklat av LexWare Labs i Göteborg (www.lexwarelabs.com). LexWare identifierar morfosyntaktiska

¹ Prof. Roman Laskowski är projektledare, forskningsassistenter är fil. dr Elżbieta Dura och fil. dr Maria Toporowska-Gronostaj, samt under sista halvåret fil. mag. Morgan Nilsson.

enheter i den givna texten och väljer sedan ut de som efterfrågas. Dess textanalys bygger bl.a. på: över 70 000 uppslagsenheter¹, 6 tilläggsordlistor², 800 000 grammatiska former, 400 ordbildningsregler, och 500 regler för kontextbaserad disambiguering. Tester på en testkorpus (Stockholm-Umeå-Corpus) visar att man måste räkna med över 2% fel i LexWares morfosyntaktiska identifiering. Det innebär att beräkningarna inte skall betraktas som exakta, utan enbart som approximationer. Man bör även räkna med en viss övertolkning av sammansättningar p.g.a. felstavningar i de undersökta texterna, t.ex. analyseras felstavning av *betraktas* som *bet-rakats*.

I slutet av 1999 kommer de erhållna urvalen av ordleder att publiceras på projektets hemsida: *bratislava.slav.gu.se/LexWare/sms*. Här presenteras enbart korta utdrag.

1. Ordleder och bundna lexem

Huvuddelen i ett sammansättningslexikon är sammansättningar, huvuddelen i ett aktivt sammansättningslexikon är produktiva ordleder. De senare särskiljs som två ordledstyper i *Nyord i svenskan* (1986): tidstypiska ordleder, t.ex. *tv*, både som för- och efterled, och de som är mera bestående, vars produktivitet förklaras med att "en viss bildning med ordet i fråga som förled eller efterled väckt sådan anklag att den kommit att stå som mönster för efterföljare. Ofta rör det sig om bildlig användning av ord" (*Nyord i svenskan* 1986:6). Det senare är en del av ett välkänt fenomen i svenskan: en operationalisering av sammansättningsleder. Detta kan illustreras t.ex. med: *riksdagsman* som avser inte bara *män* utan även kvinnor, *bilburen* som inte betyder att en bil bär något, *blixtfull* där det inte är någon blixt inblandad, *barnvänlig* där ingen vänlighet förväntas av saker som är *barnvänliga*.

¹ Uppslagsord har hämtats mestadels från Språkbankens lexikaliska databas, affix från Liljestränd (1975) och Thorell (1981).

² Tilläggsordlistor innefattar bl.a. egennamn, datatermer, latinska eller engelska ord.

”Många av våra ordbildningsmedel, [...] intar en mellanställning mellan sammansättningsled och suffix”, skriver Söderbergh (1967:30); ”på detta sätt har vi fått ett suffix som *-lig*, vilket utvecklats ur den fornsvenska sammansättningsleden *-liker*”. Även Liljestrand (1975:45) tar upp vad han betecknar som övergång från självständiga morfem till avledningsemorfem, ordleder som är övergångsformer mellan sammansättning och avledning. Enligt honom

finns det ett stort antal produktiva för- och efterleder, vars klassificering som sammansättningsleder eller som avledningselement man kan vara tveksam om. Särskilt förefaller det vara fallet när det gäller efterleder. Det utmärkande för dessa efterleder är att de dels förekommer som självständiga ord, dels som delar av sammansättningar. Betydelsen i båda användningarna är densamma. Men efterhand som efterleden börjar uppträda i allt fler sammansättningar börjar den förlora sin ursprungliga, fixerade betydelse och får en allt allmännare innebörd.

Svårigheter med att behandla denna gråzon i svensk ordbildning gör den inte mindre viktig i språkbruket. Övergång från ord till ordleder ”har skett oupphörligt under språkets historia” (Söderbergh 1967:30), och ord som är operationaliserade som ordleder har alltid funnits i det svenska lexikonet: de är inte affix och bundna morfem, utan ordleder och bundna lexem. Det krävs en separat artikel för avgränsning av bundna lexem, här karakteriseras de enbart utifrån kriterier som används i den aktuella korpusundersökningen.

Som lexem räknas här lexikonenheter definierade i tre parametrar: form, betydelse och funktion i ord- och satsbildning. Definitionen följer Mel’čuks tredelade definition av en lingvistisk symbol: *signifiant*, *signifié*, *syntactics* (1982:24–28). Det är alltså inte enbart lexikalisering av form och betydelse som kan skilja lexem åt, utan även lexikalisering av funktion. T.ex. är *ha* minst tre olika lexem: *ha* som perfektoperator, *ha* som stödverb och *ha* som innehållsord. Som bundna lexem räknas de ord som svårigen kan klassificeras som fria morfem därför att de används på liknande sätt som bundna morfem, och som inte heller kan klassificeras som bundna morfem därför att de är antingen rotmorfem eller flermorfemiga.

De enheter som eftersökes i extraheringen är alltså inte enbart de med en särskild ordledsform, som t.ex. *justitie-*, *general-* eller *-bent*. Även när de inte har någon särskild ordledsform betraktas de som bundna lexem när en särskild betydelse är knuten till användningen som sammansättningsled, t.ex. är *vänlig* två lexem: det fria lexemet *vänlig* och det bundna lexemet *-vänlig*.¹

Övergångsformer utgör ett problem för statiska ordböcker, samtidigt som det är särskilt viktigt att de som lär sig svenska som andraspråk får veta att ett ord har operationaliserats som ordled. En infödd talare undviker intuitivt kompositionell sammansättning där ett ordmönster föreligger och det är just den typen av lexikala kunskaper hos en infödd talare som behöver återspeglas i en aktiv ordbok: t.ex. att *fågelfri* tolkas snarare som 'fri från fåglar' än 'fri som fågeln' (som annars är en helt vanlig fras). Det faktum att adjektivet *fri* även förekommer som huvudled i fraser med liknande betydelse som den som realiseras av ordleden *-fri* minskar inte dess roll som en produktiv ordbildningskomponent, och den aktiva ordboken behöver återspegla detta på något sätt. I synnerhet gäller detta de lexem som är produktiva i ordbildning och vars ordmönster kan beskrivas med selektionsrestriktioner, och inte enbart med kollokationsrestriktioner.

2. Ordleder i ordböcker

Någon transfer av sammansättningskunskaper kan inte väntas från slaviska språk och därför ställs speciella krav på att ordleder i en tvåspråkig ordbok med svenska som källspråk görs åtkomliga. Varken ordleder eller någon vägledning till ordmönster finns att hitta i svensk-polska ordböcker. Sammansättningar finns däremot representerade, ibland som uppslagsord och ibland som användningsexempel. Ett av syftena med projektet är att fylla den luckan. Den svenska delen i den största svensk-polska ordboken

¹ Det faktum att vissa bundna och fria lexem associeras tydligt till en gemensam rot behöver också återspeglas i ordboken, det är dock alltid enklare att slå samman enheter som har särskiljts än tvärtom.

(Kubitsky 1988) följer mestadels *Svensk ordbok* (1986), utom just när det gäller ordleder. Om de överhuvudtaget får egna uppslag presenteras de som vanliga ord. T.ex. *-faldig*, *-årig*, *-makeri*, *ärke*- finns inte alls, *betonad*, *riktig*, *sinnad*, *tålig* finns enbart som vanliga adjektiv.

Många ordleder finns i svenska ordböcker, men de kan vara svåra att komma åt. Den 12:e upplagan av *Svenska Akademiens ordlista* lägger ännu mer möda än de tidigare upplagorna på att återspegla hur orden är uppbyggda, men den är huvudsakligen en lista över ord och inte ordleder. Förleder går att hitta i ordlistan, men efterleder finns enbart utspridda som exempel: *-fri*, *-säker*, eller *-vänlig* hittas enbart under sammansättningar: *svavelfri*, *tjuvsäker*, *barnvänlig*, m.m. I *Svensk ordbok* och dess efterföljare *Nationalencyklopedins ordbok* (1995) är ordleder rätt välrepresenterade som uppslagsord. När de är egna uppslagsord får de beteckningen *förled* eller *slutled*, annars markeras funktionen med en kommentar. Hos t.ex. adjektivet *fri* i betydelsen 'fri från' finns kommentaren "ofta i sammansättningar"; adjektivet *special* är markerat med "nästan enbart i sammansättningar" men hos adjektivet *säker* markeras inte alls dess efterledsfunktion. Man kan inte vara säker på vad avsaknad av kommentar eller användningsexempel innebär, huruvida ordet antas vara frånvarande i ordbildning eller inte.

Det är alltså ingalunda en regel att ord som förekommer enbart eller företrädesvis i sammansättningar får egna uppslag som sådana, inte ens i svenska ordböcker. Många finns enbart som användningsexempel. Annorlunda är det i *Nyord i svenskan* (1986), där just ordleder fokuseras. Ordboken upprättades som resultat av en undersökning av svenska texter från 1940 till 1980. Den har 7 500 enheter, däribland många produktiva ordleder som inte tas upp i de ovannämnda uppslagsverk, t.ex. *-bank* (*blodbank*), *mjuk-* (*mjukstart*), *när-* (*närdemokrati*), *snabb-* (*snabbval*), m.m. Men vissa ordled som är högre frekventa i den undersökta korpusen visar sig saknas i *Nyord i svenskan*, t.ex. *-fylld* (*motsatsfylld*), *-laddad* (*ångestladdad*).

I ryskan är behovet av en speciell ordbildningsordbok inte enbart erkänt utan även tillfredsställt. En ordbok som ägnas ute-

slutande åt ordbildning kom ut 1996: *Tolkovyj slovar' slovo-obrazovatel'nych edinic russkogo jazyka* (Rysk ordbildningsordbok). Den har 1 875 välbeskrivna uppslagsenheter. Författarinnan T.F. Efremova skriver att idén är c:a 100 år gammal, förverkligandet har dock dröjt. Ett utkast till ett polskt ordbildningslexikon finns hos M. Majewska (1998): *O potrzebie rejestrowania hasel morfemowych* (Om behovet av att registrera morfemiska uppslagsenheter). Artikeln innehåller även en tillbakablick som visar att ordbildning brukade ta upp mycket plats i äldre polska uppslagsverk. Författarinnan argumenterar för behovet av att upprätta en speciell ordbok ägnad uteslutande åt ordbildning, eller åtminstone en lika systematisk och heltäckande förteckning av morfem som av lexem i en gemensam ordbok. Egentligen borde det vara självklart, skriver Majewska. Om man finner uppslagsordet *auto* 'bil' i en ordbok, varför skulle man inte finna *auto-* 'själv-'? Båda kan förekomma även i ordbildning och behöver separeras i synnerhet då det föreligger homofoni, t.ex. *autoalarm* 'billarm' och *autobiografia* 'självbiografi'.

3. Ordleder i texter

Två korpusundersökningar utfördes inom projektet: den första med fokus på nya verbala sammansättningar som helhet, den senare på leder i dessa; enbart den senare beskrivs här. Med verbala sammansättningar avses här inte enbart de med finita verb som huvudled utan även med particip, både perfekt och presens.¹ Med nybildningar avses de ord som inte finns representerade som uppslagsord hos LexWare. LexWares uppslagsenheter överensstämmer i stort med de i *Nationalencyklopedins ordbok*, vilket innebär att de sammansättningar som väljs ut som nya inte finns bland uppslagsorden i *Nationalencyklopedins ordbok*, men de kan vara upptagna i t.ex. *Nyord i svenskan* eller *Svenska Akademiens ordlista*. Undersökningen koncentrerar sig på nybildningar av två skäl: extraheringen väntas

¹ Även deverbala substantiv undersöks i projektet.

utgöra ett komplement till och inte en upprepning av det som redan finns i ordböcker, den skall förse oss med ordmönster som verkligen är produktiva.

89 664 grammatiska ord¹: infinitiv, finita verb, presens och perfekt particip, finns belagda med 7 357 817 löpord i korpusen på över 42 miljoner löpord. 32 372 av dessa ord utgörs av nybildningar, 57 292 utgörs av redan kända ord. Antalet förekomster i löpord anges nedan parallellt med antalet förekomster i grammatiska ord. Ett urval av de mest frekventa orden i verbala sammansättningar presenteras nedan med ett exempel på en sammansättning som de funnits belagda i. Vad gäller efterleder omfattar antalet förekomster både finita verb, infinitiver och particip. I vissa fall är det inte particip utan endast participiell form; så motsvaras *färgad* som efterled i de flesta fall av *att ha färg*, inte av *att färga*.

TABELL 1. *Utdrag ur de mest frekventa förlederna.*

förled	ord	löpord	exempel
<i>ny-</i>	854	4 613	<i>nyöversätter</i>
<i>väl-</i>	568	2 462	<i>välöverbägt</i>
<i>svår-</i>	304	925	<i>svåröverbakade</i>
<i>special-</i>	205	639	<i>specialvävda</i>
<i>färdig-</i>	270	631	<i>färdigåkt</i>
<i>små-</i>	209	451	<i>småöva</i>
<i>hög-</i>	107	451	<i>högväxte</i>
<i>stor-</i>	97	430	<i>störväxte</i>
<i>själv-</i>	196	413	<i>självöppnande</i>
<i>halv-</i>	192	396	<i>halvöppnade</i>

¹ Grammatiskt ord och löpord används som i Svenska Akademiens grammatik (1999:204).

TABELL 2. *Utdrag ur de mest frekventa verben i efterleder.*

verb i efterled	ord	löpord	exempel
<i>fylla</i>	354	1 236	<i>adrenalinfylld</i>
<i>bygga</i>	152	1 030	<i>amatörbyggda</i>
<i>spela</i>	197	879	<i>bollspelande</i>
<i>äga</i>	192	820	<i>vitägda</i>
<i>skada</i>	240	717	<i>ögonskadade</i>
<i>gå</i>	78	678	<i>östgående</i>
<i>likna</i>	439	667	<i>öliknande</i>
<i>färga</i>	305	600	<i>ölfärgad</i>
<i>betala</i>	76	575	<i>efterbetalt</i>
<i>köra</i>	133	548	<i>amatörkörd</i>

Många sammansättningar skapas tillfälligt, och därför finns det en risk att de återspeglar den dominerande texttypen (tidnings-text) snarare än de dominerande ordmönstren i språket generellt. Frekvens behöver kompletteras med andra, mera lexikaliskt inriktade parametrar för att uppnå större träffsäkerhet. Två sådana har definierats och använts i extraheringen: sammansättningsbenägenhet och sammansättningsproduktivitet.

4. Sammansättningsbenägenhet

För alla lexikonverb som finns belagda i korpusen har alltså dessas sammansättningsbenägenhet beräknats enligt följande:

förekomster av verbet i sammansättningar /
samtliga förekomster av verbet

Måttet anger hur ordets användning i ordbildning är relaterad till dess användning i satsbildning. Om textunderlaget antas vara tillräckligt stort, kan sammansättningsbenägenhet beräknas även i antal grammatiska ord och inte bara i löpord. Den kan beräknas för verb som lexikonord samt för olika verbformer separat.

Nedan presenteras den genomsnittliga sammansättningsbenägenheten hos verb. Beräknad i förekomster blir den genomsnittliga sammansättningsbenägenheten låg p.g.a. högfrekventa hjälpverb. Beräknad i grammatiska ord är den däremot mycket hög, inte bara hos particip utan även hos finita verb former och infinitiv. Om de verb som inte alls belagts i sammansättningar inte tas med i beräkningen av den genomsnittliga sammansättningsbenägenheten, blir den mycket högre för de verb som överhuvudtaget förekommer i ordbildning. Även om ett par procent räknas bort som fel, visar resultaten ändå tydligt att sammansättning utgör en betydande andel i nybildning även när det gäller verb.

TABELL 3. *Genomsnittlig sammansättningsbenägenhet hos verb.*

	samtliga verb, dvs. sammansättnings- benägenhet ≥ 0		enbart verb med sammansättnings- benägenhet > 0	
	löpord	ord	löpord	ord
samtliga former	1.08	36.10	2.25	55.35
finita och infinitiv	0.24	22.92	0.62	45.31
presens particip	15.89	58.31	24.00	78.81
perfekt particip	3.74	40.47	4.71	56.03

Verb som visar sig helt saknas i ordbildning är som väntat hjälpverb: *bli, bör, lär, kunna, måste, skall, vilja*. Vissa verb förekommer inte i sammansättningar som finita former men gärna som presens particip, t.ex. *behöva*. Det är en intressant och öppen fråga varför vissa innehållsverb med mycket hög frekvens i de undersökta texterna inte alls förekommer i sammansättningar. Nedan presenteras en rätt brokig lista av verb som har hög frekvens men är helt sammansättningsobenägna, åtminstone som finita former eller perfekt particip.

<i>befara</i>	<i>betyda</i>	<i>fortsätta</i>	<i>föredra</i>
<i>förefalla</i>	<i>försöka</i>	<i>gilla</i>	<i>gälla</i>
<i>hänvisa</i>	<i>hävda</i>	<i>inleda</i>	<i>innebära</i>
<i>innehålla</i>	<i>inse</i>	<i>instämma</i>	<i>kosta</i>
<i>meddela</i>	<i>medföra</i>	<i>medverka</i>	<i>minska</i>
<i>påpeka</i>	<i>riskera</i>	<i>räcka</i>	<i>sakna</i>
<i>ske</i>	<i>skylla</i>	<i>slippa</i>	<i>sluta</i>
<i>tillhöra</i>	<i>undvika</i>	<i>understryka</i>	<i>upphöra</i>
<i>uppmåna</i>	<i>uppmuntra</i>	<i>utesluta</i>	<i>utgöra</i>
<i>uttrycka</i>	<i>veta</i>	<i>våga</i>	<i>värna</i>
<i>ytra</i>	<i>ägna</i>	<i>övertaska</i>	<i>övertyga</i>

Det är dock inte verb med låg, utan med hög sammansättningsbenägenhet som är mest relevanta för ett sammansättningslexikon. Nedan presenteras de ordnade enligt frekvens och sammansättningsbenägenhet. Sammansättningsbenägenheten har räknats fram på samma sätt som genomsnittet för alla verb.

TABELL 4. *Verb med hög frekvens och hög sammansättningsbenägenhet.*

efterled	benägenhet	löpord
<i>fylla</i>	17.42	1 236
<i>bygga</i>	5.58	1 030
<i>spela</i>	2.82	879
<i>äga</i>	11.48	820
<i>gå</i>	0.60	678
<i>skada</i>	11.65	717
<i>betala</i>	3.50	575
<i>likna</i>	19.87	667
<i>köra</i>	29.47	529
<i>skriva</i>	2.12	534

5. Sammansättningsproduktivitet

Sammansättningsproduktivitet är en parameter som skapats för att ytterligare höja träffsäkerheten i urvalet av ordleder. Den är särskilt inriktad på att välja de ordleder som allra troligast kan utgöra bundna lexem. Parametern fungerar som ett komplement till både frekvens och sammansättningsbenägenhet och tillför den dimension som lexikaliskt sett verkar viktigast.

För de lexikonenheter som identifierats i sammansättningar har dessas sammansättningsproduktivitet beräknats enligt följande:

lexikonenheter förekommande tillsammans med
den valda enheten /
samtliga sammansättningar med den valda enheten

Produktivitet kombinerad med frekvens och sammansättningsbenägenhet ger en mycket värdefull sortering. De listor av för- och efterleder som presenteras nedan visar detta tydligt.¹

TABELL 5. *Högproduktiva förleder.*

förled	produktivitet (%)	antal ord	antal kollokat	exempel
<i>ny-</i>	42	854	358	<i>nybildade</i>
<i>väl-</i>	45	568	258	<i>välskötta</i>
<i>färdig-</i>	61	270	166	<i>färdigbehandlad</i>
<i>svår-</i>	53	304	162	<i>svårtolkat</i>
<i>små-</i>	71	209	149	<i>smådjävlat</i>
<i>lätt-</i>	65	214	139	<i>lättanvända</i>
<i>själv-</i>	70	196	138	<i>självkrönt</i>
<i>halv-</i>	62	192	119	<i>halvciviliserade</i>
<i>snabb-</i>	69	133	92	<i>snabbinkallad</i>
<i>fel-</i>	53	172	91	<i>felklädd</i>
<i>hård-</i>	61	143	88	<i>hårdbantad</i>

¹ Med *kollokat* avses endast tillsammans förekommande lexikonenheter och inte skillnaden mellan *bas* och *kollokat*.

TABELL 6. *Högproduktiva efterleder.*

efterled	produktivitet (%)	antal ord	antal kollokat	exempel
<i>-liknande</i>	99	439	433	<i>dvalliknande</i>
<i>-fylld</i>	58	354	207	<i>adrenalinfylld</i>
<i>-baserad</i>	68	250	169	<i>bildbaserad</i>
<i>-färgade</i>	55	305	168	<i>blodfärgade</i>
<i>-klädda</i>	66	255	168	<i>affärsklädda</i>
<i>-inriktad</i>	59	269	160	<i>barinriktad</i>
<i>-styrd</i>	65	217	142	<i>bildstyrd</i>
<i>-formad</i>	59	236	140	<i>bananformade</i>
<i>-intresserad</i>	55	210	116	<i>fotbollsintresserad</i>
<i>-relaterade</i>	68	161	109	<i>inkomstrelaterade</i>

6. Lexikologisk relevans hos urvalen

En intelligent korpusundersökning kan vara till hjälp i lexikologiskt arbete på många olika sätt.

Den mest uppenbara nyttan med att extrahera nybildningar är att ens lexikon kan kompletteras med nytillkomna lexikaliseringsringar. Vissa nya sammansättningar verkar förtjäna egna uppslag i ordboken, t.ex. *intelligensbefriad*, *kritikerrosad*, *långtgående*, *pådrivande*, *räntebärande*, *synliggöra*, *återuppstå*, m.m. Även bland icke-sammansatta nya verb finner vi möjliga uppslagsord, som t.ex. *destabilisera*, *desinformera*.¹

För en aktiv ordbok är dock de nya sammansättningar som följer ett produktivt ordmönster mest intressanta. Om korpusen t.ex. visar att de högfrekventa efterlederna *-fylld* och *-laddad* utgör ett produktivt ordmönster bör dessa tas upp som uppslagsenheter och förses med både betydelse och form som är specifik för användningen som efterled. Korpusundersökningen visar en påtaglig homogenitet i användning av *-fylld* och *-laddad*. I få fall är det fråga om att något verkligen 'fylls' eller 'laddas'

¹ I *Nationalencyklopedins ordbok* finns t.ex. endast *destabiliserande*, *desinformation*.

bokstavligen, även om sådana fall förekommer, t.ex. *Libero-laddad*. De flesta belägg avser en bildlig användning, som även finns i sådana konventionaliserade sammansättningar som *färo-fyllad* och *känsloladdad*; t.ex. *prestigefyllad*, *lustfyllad*, *skräckfyllad*, *kampfyllad*, *motsatsfyllad*, *prestigeladdad*, *ångestladdad*.

Bland leder som utgör ordmönster utan att någon metafor är inblandad men som ändå är specialiserade som ordled återfinns *hand-*, *ny-*, *special-*, vilka egentligen betyder 'manuellt', 'nyligen', och 'särskilt'. Att *hand-* som en verbal förled kommer av frasen *för hand* och inte av substantivet *hand*, visas i urvalet av dess sammansättningar:

<i>handblåst</i>	<i>handbroderat</i>	<i>handbyggd</i>	<i>handdrejade</i>
<i>handdrivna</i>	<i>handfärgade</i>	<i>handknutna</i>	<i>handknypplade</i>
<i>handkolorerade</i>	<i>handritad</i>	<i>handrullad</i>	<i>handsmida</i>
<i>handsnidad</i>	<i>handstickad</i>	<i>handsydd</i>	<i>handsätta</i>
<i>handtextade</i>	<i>handtryckta</i>	<i>handvevad</i>	<i>handvirkad</i>

En rangordning av verb enligt produktivitet fokuserar de verb som är produktiva efterleder och vars kollokationsmönster i ordbildning är välrepresenterade i texter. En extrahering av kollokat är särskilt viktig i det aktuella projektet då det skall kartlägga hur verbets valensram realiseras i sammansättningar. Ett exempel på ett urval av olika förleder till ett specifikt verb presenteras nedan:

<i>behovsstyrda</i>	<i>centralstyrda</i>	<i>datorstyrda</i>	<i>detaljstyrda</i>
<i>föräldrastyrda</i>	<i>högerstyrda</i>	<i>indiskstyrda</i>	<i>lagstyrda</i>
<i>lättstyrda</i>	<i>maffiastyrda</i>	<i>maktstyrda</i>	<i>militärstyrda</i>
<i>miljöstyrda</i>	<i>moderatstyrda</i>	<i>målstyrda</i>	<i>planstyrda</i>
<i>radarstyrda</i>	<i>regelstyrda</i>	<i>resultatstyrda</i>	<i>rollstyrda</i>
<i>räntstyrda</i>	<i>röststyrda</i>	<i>självstyrda</i>	<i>statsstyrda</i>
<i>svårstyrda</i>	<i>terminstyrda</i>	<i>toppstyrda</i>	<i>trådstyrda</i>

En värdefull upplysning om egenheter hos ordleder får man genom att extrahera och jämföra närsynonyma ordleder som förekommer med en och samma ordled. T.ex. används endast *väl* i förled som en allmän positiv bestämning vid particip, det

närbesläktade *bra* förekommer inte alls i den rollen. Jämför man användning av *färdig-* och *slut-*, samt *hel-* och *total-*, visar det sig att *färdig-* och *hel-* förekommer mest med perfekt particip, medan *slut-* och *total-* även förekommer med finita verb.

TABELL 7. *Färdig- och slut- med samma verb i efterleder.*

färdig-	slut-	verb i efterleder
<i>färdigbehandlad</i>	<i>slutbehandlad</i>	<i>behandla</i>
<i>färdigbesiktigad</i>	<i>slutbesiktigas</i>	<i>besiktiga</i>
<i>färdigförhandlad</i>	<i>slutförhandla</i>	<i>förhandla</i>
<i>färdigräknade</i>	<i>sluträknad</i>	<i>räkna</i>
<i>färdigspelad</i>	<i>slutspelad</i>	<i>spela</i>

TABELL 8. *Hel- och total- med samma verb i efterleder.*

hel-	total-	verb i efterleder
<i>helstoppad</i>	<i>totalstoppa</i>	<i>stoppa</i>
<i>helmissade</i>	<i>totalmissa</i>	<i>missa</i>
<i>helreoverade</i>	<i>totalreovera</i>	<i>reovera</i>

På liknande sätt kan man jämföra antonyma ordleder som förekommer med en och samma ordled. Att förse dessa med ekvivalenter i ett annat språk hjälper även till att avgöra vilka som är viktigast att ta upp som uppslagsord. T.ex. kräver sammansättningar med *stor-* och *små-* egna ekvivalenter, medan *lätt* och *svår* verkar vara kompositionella även i engelskan.

TABELL 9. *Stor- och små- med samma efterleder.*

stor-		små-	
<i>storförtjust</i>	mighty delighted	<i>småförtjusta</i>	rather delighted
<i>storskrattade</i>	laugh loudly	<i>småskrattade</i>	chuckle
<i>störväxte</i>	tall of stature	<i>småväxte</i>	short of stature

TABELL 10. Lätt- och svår- med samma efterleder.

lätt-		svår-	
<i>lättavlästa</i>	easy to read from	<i>svåravläst</i>	difficult to read from
<i>lättadministrerad</i>	easy to administrate	<i>svåradministrerad</i>	difficult to administrate
<i>lättanalyserat</i>	easy to analyse	<i>svåranalyserat</i>	difficult to analyse

7. Slutord

Bundna lexem är både svårast och viktigast att ta upp i tvåspråkiga ordböcker. En klar vägledning behövs för en icke-svensk både vad gäller lexikaliserade sammansättningar och sammansättningsmönster med bundna lexem. Bundna lexem är viktiga även i svenska ordböcker. Korpusundersökningens resultat tyder på att sammansättning är så pass produktiv i författande av skriven svenska att bara detta motiverar ett modernt aktivt lexikon ägnat enbart åt ordbildning. Alla produktiva för- och efterleder skulle få egna uppslag i en sådan ordbok. Man skulle ha möjlighet att slå upp en enskild ordled på samma sätt som man slår upp ett ord i en ordbok, och finna alla relevanta uppgifter om dess form, betydelse och användning samlade på en plats, som t.ex. i den ovannämnda ryska ordboken.

Den typ av intelligent extrahering som presenteras ovan är värdefull i lexikologiskt arbete. Med ett tillräckligt brett textunderlag kan huvudfunktionen hos en enhet spåras i texter. Med tanke på att övergångsformer inte bara är enheter utan även en process, behövs den typen av extrahering från nya texter nästan jämt. Med denna som hjälp kan en statisk ordbok förses med ett dynamiskt komplement, där uppdateringar automatiskt skulle följa nytillkomna texter.

Litteratur

- Efremova, T.F. 1996. *Tolkovyj slovar' slovo-obrazovatel'nych edinic russkogo jazyka*. Moskva: Russkij jazyk.
- Hellberg, Staffan. 1978. *The Morphology of Present-Day Swedish*. Stockholm: Almqvist och Wiksell.
- Johannisson, Ture. 1964. "Tendenser i nutida svensk ordbildning." I: *Nutidssvenska*. Stockholm: Almqvist & Wiksell.
- Josefsson, Gunlög. 1997. *On the principles of word formation in Swedish*. Lund: Lund University Press.
- Kubitsky, Jacek. 1988. *Słownik szwedzko-polski. Svensk-polsk ordbok*. Stockholm: Natur och Kultur.
- Liljestrand, Birger. 1975. *Så bildas orden*. Lund: Studentlitteratur.
- Majewska, Małgorzata. 1998. "O potrzebie rejestrowania hasel morfemowych w słownikach języka polskiego." I: *Poradnik językowy* 8-9. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego.
- Malmgren, Sven-Göran. 1993. "Sammansättningsmorfologi och lexikografi." I: *Nordiske studier i leksikografi*. 2. (Skrifter udgivet af Nordisk Forening for Leksikografi. Skrift nr. 2.) København.
- Mel'čuk, I.A. 1982. *Towards a Language of Linguistics*. München: Wilhelm Fink Verlag.
- Noreen, Adolf. 1906. *Vårt språk*. 7. Lund: CWK Gleerups Förlag.
- Nationalencyklopedins ordbok*. 1995. Höganäs: Bra Böcker.
- Nyord i svenskan från 40-tal till 80-tal*. 1986. Stockholm: Esselte Studium.
- Språkbanken. 1999. www.svenska.gu.se
- Stockholm-Umeå-Corpus. 1999. www.ling.su.se/DaLi/Projects/SUC/
- Svensk ordbok*. 1986. Stockholm: Esselte Studium.
- Svenska Akademiens grammatik*. 1999. Stockholm: Norstedts ordbok.
- Svenska Akademiens ordlista över svenska språket*. 1998. Stockholm: Norstedts Ordbok.

- Söderbergh, Ragnhild. 1968. *Svensk ordbildning*. Stockholm: Nämnden för svensk språkvård.
- Teleman, Ulf. 1972. *Om svenska ord*. Lund: Gleerups.
- Thorell, Olof. 1981. *Svensk ordbildningslära*. Stockholm: Esselte Studium.

Øystein Eek

Markedsorientert lemmaseleksjon: kriterier og prioriteringer

The article deals with criteria for the selection of lemmas in bilingual dictionaries from a commercial point of view. The author is an Editorial director at Kunnskapsforlaget, a Norwegian publishing company. As a publisher of dictionaries of various sizes and for various market segments, he discusses which criteria are most important for the various dictionary series. In general, they are the combined *target group x use* criteria, *frequency* and *update*, and also *encyclopedic representativity*, *culture-specific considerations* and *gender neutrality* – their importance varying according to the type of dictionaries, whether they are dictionaries for the general market or school dictionaries for different levels. The article also includes a brief discussion of *lexicographic genres* within the family of dictionaries.

1. Bakgrunn og introduksjon

Dette foredraget er sprunget ut av diskusjoner om produktutvikling i Kunnskapsforlagets ordbokredaksjon, altså en annen type miljø enn de akademiske leksikografiske miljøene de fleste av konferansens foredragsholdere er knyttet til. Forlagets formål med å utgi ordbøker er kommersielt: Vi skal tilby ordbøker som skaffer forlaget best mulig inntekter (og vi tror vi oppnår dette best ved å tilby markedet best mulig ordbøker). Også forskningsmiljøene bør være interessert i hvilke ordbokutgivelser som er kommersielt aktuelle, og hvilke retningslinjer for det leksikografiske arbeid forlagene er opptatt av.

Vi kan ikke se at det er noen motsetning mellom den akademisk-leksikografiske interesse for lemmaseleksjon, og den interesse som også har en markedsmessig motivasjon, men for et forlag er det selvsagt særlig viktig å arbeide ut fra kriterier og prioriteringer som fungerer i et bestemt marked og i en tidsbestemt konkurransesituasjon. Når vårt forlag for tiden satser særlig tungt på produktutvikling innenfor ordboksektoren, er det ikke først og fremst fordi vi plutselig har gjennomgått en leksikogra-

fisk bevisstgjøring, men fordi vi har møtt økt konkurranse. I tillegg bygger vi på en mer ideell forpliktelse til å drive kontinuerlig produktutvikling til beste for våre kunder, og vi ønsker å samarbeide med de beste norske leksikografer ved å engasjere dem som forfattere og konsulenter. Flere av dem er til stede her på konferansen.

Diskusjonen om lemmaseleksjon er langt mer kompleks når det gjelder allmennordbøker enn fagordbøker, og siden Kunnskapsforlaget arbeider mest med allmennordbøker, skal jeg konsentrere meg om disse. For å snevre temaet ytterligere inn, vil jeg også holde meg til tospråklige allmennordbøker, der vi har de fleste titlene, og der behovet for produktutvikling for tiden er størst.

I Kunnskapsforlaget pågår en kontinuerlig prosess hvor vi prøver å analysere og beskrive vårt utgivelsesprogram langs flere akser enn seriene. Vår term for dette er *linjestrategi*.

Linjene defineres ut fra

- *språk*, f.eks. engelskordbøker;
- *målgruppe*, f.eks. elever i videregående skole;
- *medium/publiseringssplattform*, f.eks. elektroniske ordbøker på CD-ROM eller on-line; her hører også de tradisjonelle seriene hjemme;
- *tematisk bredde*, f.eks. allmennordbøker vs. fagordbøker;
- *ettspråklighet vs. tospråklighet*.

Et biprodukt av dette arbeidet vil etter hvert bli mer detaljerte utvalgskriterier og redigeringsprinsipper for titlene i hver serie, og en sterkere styring av prosjektene fra forlaget side vis à vis forfatterne. Dette er én av flere grunner til at forlaget vil ha en mer bevisst holdning til lemmaseleksjonen enn det som har vært tilfelle tidligere. I samme retning peker kravet om raskere fremdrift av prosjektene, som fører til at stadig flere ordbøker blir laget av forfatterteam. På dette punkt er vi kommet noe sent i gang med en utvikling som er kjent fra mange ordbokforlag i utlandet.

For tiden er vi foruten lemmaseleksjonen intenst opptatt av hvordan ordboksartiklene skal organiseres og presenteres ved hjelp av typografiske og andre visuelle virkemidler. De som sammenligner den tyske blå ordboken som kom i 1998, med tidligere utgaver, vil se at brukergrensesnittet er i ferd med å bli fornyet. I år kommer flere titler hvor dette arbeidet er ført videre.

Det er et viktig poeng for oss at den fornyelsen vi foretar, skal være lett synlig for dem som skal velge en ordbok for sine elever, eller kjøpe en ordbok i bokhandelen. Ordboken skal oppfylle og helst overoppfylle de forventninger kunden har, og dette positive inntrykket skal forsterke seg gjennom bruk, slik at en tittel fra samme forlag blir foretrukket neste gang kunden er på ordbokjakt.

Forlagets utfordring blir å definere og innprente forfatterne og de øvrige medarbeiderne en miks av seleksjonskriterier som fungerer optimalt for flest mulig kunder innenfor bokens/seriens målgruppe. Retningslinjene bør være så presise at de kan gi forfatterne konkrete holdepunkter når de skal avgjøre om et bestemt ord eller en gruppe av ord skal være med eller ikke.

I tillegg skal kriteriene fungere som retningslinjer i et arbeid som vanligvis foregår slik at man har tilgang til et eldre materiale som skal suppleres. I noen tilfeller er seleksjonskriterier bak dette eldre materialet dokumentert gjennom forfatterens forord, men ofte i så generelle vendinger at det har liten praktisk verdi for dem som skal forestå revisjonen.

Det finnes sikkert mange sett av generelle kriterier for lemma-seleksjon i ordbøker, jeg har forholdt meg til Bo Svenséns systematikk i *Handbok i leksikografi*. Jeg har prøvd å formulere noen kriterier uavhengig av Svensén, styrt at hva jeg opplever som mest presserende i vårt praktiske arbeid, og så relatere dette til Svenséns sett av kriterier.

Det vi er ute etter når vi stikker hodet frem på denne konferansen, er selvsagt å få leksikografisk begrunnede tilbakemeldinger på vårt markedsorienterte resonnement, slik at vi kan oppnå ytterligere kvalitetsforbedring av produktene.

2. Kunnskapsforlagets allmennordbok-serier

Vi utgir for tiden følgende serier:

Serie	Antall oppslagsord	Målgrupper
Lommeordbøker	20 000–30 000	1) Norske turister 2) Fremmedspråklige turister 3) Andre som trenger en enkel, rimelig og kompakt ordbok
Skoleordbøker	30 000–35 000	1) Elever i grunnskolen 2) Elever som har elementær undervisning på videregående skole
Blå ordbøker	80 000–100 000	1) Elever i videregående skole 2) Segment i forbrukermarkedet som trenger en standardordbok, dvs. en passe stor, passe avansert og passe kostbar ordbok
Store ordbøker (få titler)	130 000 og over (gjelder de to engelsk-ordbøkene til sammen)	1) Profesjonelle brukere 2) Studenter 3) Segment i forbrukermarkedet som trenger en større ordbok enn standardordboken

Historisk bygger lommeordbøkene og skoleordbøkene for engelsk, tysk, fransk og spansk på samme ordtilfang, noe som er en utfordring når det gjelder tilpasning av lemmaseleksjonen i produktutviklingen.

3. Kriterier for lemmaseleksjon

De kriterier jeg vil behandle, er følgende:

- (1) Brukergruppe x bruksmåte
- (2) Frekvens
- (3) Aktualitet
- (4) Encyklopedisk representativitet
- (5) Kulturspesifikke hensyn
- (6) Kjønnsnøytralitet
- (7) Oppdragelse av brukeren
- (8) Behandling av sammensatte ord
- (9) Leksikografisk genre

En annen og mer fullstendig og generell systematikk finnes i Bo Svenséns *Handbok i leksikografi* (1987:28–40).

3.1. Brukergruppe x bruksmåte

Det første kriteriet, kombinasjonen av brukergruppe og bruksmåte, vil jeg belyse med lommeordbøkene som eksempel. Vi tenker oss at de først og fremst kjøpes av nordmenn som skal til land der vedkommende språk blir brukt, og i noen grad av fremmedspråklige turister i Norge. Men hva skal f.eks. en italiensk lommeordbok, som altså ikke er en parlør, men en vanlig ordbok i miniformat, brukes til?

Den italiensk-norske delen kan ikke brukes til å forstå muntlig italiensk med mindre man stadig avbryter den italiensktalende og går over til en slags metakonversasjon. Derimot egner denne ordbokdelen seg godt til å undersøke betydningen av ord man ser på skilt, eller korte tekster i brosjyrer og lignende, f.eks. på hoteller, når man benytter transportmidler, besøker butikker, er på turistsentra eller severdigheter. For meg er idealbruken av en lommeordbok å finne ut meningen med mest mulig av kortfattede tekster som er synlige på offentlige steder i Italia.

Å tilpasse ordforrådet til den antatt viktigste brukergruppen og bruksmåten vil være å sørge for at mest mulig av det man regelmessig kan se på trykk på offentlige steder, er med. Hovedvekten vil ligge på praktisk informasjon, både rettet til turister og til landets egne innbyggere (ikke minst det siste!) samt på forhold som har med kulturminner og andre severdigheter å gjøre..

Den norsk-italienske delen vil bli brukt f.eks. til fremvisning av ord fra ordboken til italiensktalende som man prøver å kommunisere med. De noe mer språkkyndige vil kunne bruke den til å supplere sitt ordforråd når de prøver å tale italiensk. Vi er usikre på hvor stor andel av lommeordbok-kjøperne som har tilstrekkelig språklig kompetanse til slik bruk. Uansett må lemmaseleksjonen fange opp mest mulig av slikt det er vanlig å gjøre og befatte seg med under turistopphold i Italia, og den bør fange opp alle aldersgrupper fra og med tenåringer, alle reiseformer, individuelle

og kollektive, alle årstider og landsdeler. Termer fra vintersport kan f.eks. være svært aktuelle.

Hvis lommeordboken ikke er fyldig på dette området, vil jeg hevde at den ikke er særlig god eller formålstjenlig.

I dag lager vi ordbøker hovedsakelig for norske brukere. I fremtiden vil vi måtte vurdere om det er riktig å tilpasse ordforrådet til italienske turister i Norge, og ta med ord som er typiske for det italienske turister kommer bort i her i landet. Kanskje vil markedsundersøkelser vise at denne brukergruppen er betydelig, i så fall må seleksjonen justeres.

Er da lommeordboken for turistmarkedet en temaordbok, altså en "Turistordbok" e.l.? Mitt svar er nei, fordi lommeordboken også skal inneholde et kjerneordforråd fra både fremmedspråket og norsk. Men hvordan disse to elementene skal vektet kvantitativt i forhold til hverandre, har jeg foreløpig ikke noe presist svar på.

Man kan nå straks innvende at jeg har gjort det lett for meg selv ved å bruke akkurat lommeordbokserien som eksempel på bruk av dette kriteriet, og det var også meningen. Bruken av den blå serien er nok langt mer differensiert. Men også her må vi definere en eller flere typiske brukssituasjoner knyttet til skolearbeid, og analysere hvordan lemmaseleksjonen fungerer for dette formål i denne målgruppen. Hvis vi f.eks. kommer til at én typisk brukssituasjon for norsk-engelsk del av engelskordboken er å få hjelp til å skrive brev til jevnaldrende venner i Storbritannia og USA, må vi ha en formening om hva som er de mest karakteristiske temaer for slike brev, og inkludere ordforråd som gir best mulig hjelp til slik brevskrivning.

3.2. Frekvens

Frekvenskriteriet er velkjent i den leksikografiske diskusjon, og frekvensmålinger i tekstkorpora har vært et gjennomgangstema også på denne konferansen. Ettersom alle norske forlag er små aktører, har vi ikke resurser til å bygge opp store korpora, og det er derfor problematisk for oss å bygge på frekvenskriteriet på en

vitenskapelig måte. Vi håper at vi i løpet av noen år kan ha systemer for rutinemessig bruk av korpora iallfall som korrektiv i vår lemmaseleksjon og at det også skal bli mulig for våre ordbokforfattere å kvalitetssikre den norsk-fremmedspråklige delen av sine ordbøker med støtte i store og gode korpora, for eksempel i prioriteringen av ekvivalenter og betydningsgrupper. Inntil videre er vi dessverre som regel henvist til forfatterens skjønn, dvs. subjektive kompetanse, på dette felt. Våre forfattere trenger nok opplæring i de metoder vi har fått beskrevet i andre foredrag på konferansen, bl.a. utnyttelse av språkresurser på Internett.

3.3. Aktualitet

Alle som utgir ordbøker, har vel ambisjoner om å få med en rimelig andel av ord som er av nyere dato. Hva som er en rimelig andel, er avhengig av hvor sterkt man vektet aktualiteten i forhold til andre kriterier. Men aktualitet er også noe mer; den kan dreie seg om hvor stor andel av ordforrådet som knytter seg til virksomhet som er typisk for samtiden.

Når forlagene i dag i forordet ofte betoner at det er lagt stor vekt på ordforråd fra moderne teknologi og lignende, kanskje endog sterkere enn det er dekning for i lemmaseleksjonen, er det ut fra en forvissning om at det er et godt salgsargument. I tillegg kan det være i forlagets interesse at kundene opplever ordbøker som ferskvare, at de om noen år vil bli ansett som foreldet, slik at man må kjøpe en ny ordbok. Men det er usikkert om et flertall av kundene opplever dette som positivt eller som negativt. For forlaget vil det kreve større redaksjonelle resurser å operere med betydelig gjennomtrekk av stoff.

Ikke bare skal ord inn, men ord skal også ut i forbindelse med en revisjon. Vi har tidligere på konferansen hørt at ordbokforfatterne jevnt over har vært for tilbakeholdende med dette, slik at det historiske innslaget i dagens ordbøker nok kan bli i største laget.

Denne problemstillingen er svært aktuell i forbindelse med våre blå ordbøker. Den typiske kunde er skoleelev 16–19 år,

som skal ta eksamen i et fremmedspråk og skal ha et hjelpemiddel til å forstå pensumtekstene og til å produsere de nødvendige mengder idiomatisk riktig tekst på fremmedspråket. Vi tror denne brukergruppen er tjent med at det finnes en forholdsvis stor mengde aktualitetsord, ettersom det faglige innholdet i fremmedspråkundervisningen i Norge har et langt større innslag av samtidsorientering og dagligliv enn for en generasjon siden. Tilsvarende tones det Svensén kalles gamle ord ned, dvs. ord som brukerne normalt kun støter på i eldre tekster eller i omtale av eldre forhold. Også de nye blå ordbøkene skal etter min mening kunne brukes ved lesning av sentral, eldre litteratur, men kolleger jeg har snakket med i senere tid, mener bestemt at dette ikke er mulig innenfor de snevre rammer som en mellomstor ordbok gir, dersom ordboken samtidig skal ha et sterkt aktualitetspreg. Jeg vil iallfall håpe at vi i tospråklige, mellomstore ordbøker for fremmedspråk kan inkludere ordforrådet i de eldre tekster som oftest leses i videregående skole.

3.4. Encyklopedisk representativitet

I Kunnskapsforlaget er det et uttalt mål å oppnå synergi mellom leksikonarbeid og ordbokarbeid. Vi utgir flere encyklopediske verk, og har også valgt å ha en fagredaksjon med stor encyklopedisk bredde. Leksikonredaktørene trekkes nå for første gang hos oss inn i ordbokarbeidet for å kvalitetssikre ordforrådet innen sine felt. Det må være riktig at ordforrådet skal gjenspeile et encyklopedisk mangfold, og at fordelingen av ordforrådet i større allmennordbøker ikke skal avvike nevneverdig fra den relative fordelingen av det terminologiske ordforrådet mellom de ulike fagområdene i leksika. Vi har et svært avansert fagkategoriseringssystem til rådighet, og vil etter hvert ta dette i bruk også i arbeidet med allmennordbøkene. Vi ønsker også etter hvert å knytte leksikon og ordbok sammen på den måten at kjøperne av Store norske leksikon får fremmedspråklige ekvivalenter til leksikonordforrådet tilgjengelig i elektronisk form.

Et viktig skritt på veien vil være å utstyre flest mulig oppslagsord i de elektroniske ordbokmanus med fagkategorisering, for å kunne studere kvaliteten på de enkelte delordforråd i ordbøkene, og forbedre disse på en sikrere måte enn før. Som et biprodukt vil vi også kunne tilby kjøpere av elektroniske ordbøker emneordlister som er så store at de har praktisk bruksverdi til mer enn skolestiler.

Alle vil nok være enige om at encyklopedisk representativitet er et relevant kriterium for lemmaseleksjonen i virkelig store ordbøker. Vi har vurdert om behovet for balanse og allsidighet i fagspråket kan løses på annen måte på mellomstort nivå, f.eks. med spesialordbøker tilpasset de enkelte faglige studieretninger i videregående skole. Men vi er kommet til at dette blir for komplisert redaksjonelt og for usikkert kommersielt. Vi kan heller ikke fortsette å benytte det gamle tricket med å prioritere visse fagområder i hver ny utgave av allmennordbøkene og bruke dette i forord og markedsføring for å demonstrere styrke; tvert imot blir tidligere skjevheter da bare forskjøvet til nye områder. Vår konklusjon er dermed at encyklopedisk representativitet er like relevant på mellomstort, dvs. blått, nivå, som i de største ordbøkene.

Det gjøres for tiden et stort og ambisiøst arbeid i forlaget med å samle inn termer fra en mengde forskjellige fagområder som supplement til leksikonbasen, for utnyttelse i ett- og tospråklige ordbøker, og med et stort antall informanter.

3.5. Kulturspesifikke hensyn

3.5.1. I fremmedspråket

En ordbokutgiver kan ønske å gi et rikest mulig bilde av den kultur og det/de samfunn det aktuelle fremmede språket benyttes i. F.eks. skal engelskordbøker da inneholde et ordforråd knyttet til britiske og amerikanske, australske og irske, indiske osv. samfunnsforhold og kulturtradisjoner. Dette vil gjøre ordforrådet i engelske, tyske, franske osv. ordbøker ulike ikke bare i språk, men også i begrepsinnhold. Jo større vekt utgiveren legger på

dette kriteriet, jo mindre synergi kan han oppnå mellom titlene i serien.

Som nevnt er det en sammenheng mellom vektlegging av dette kriteriet og ønske om å lage en mest mulig formålstjenlig ordbok, for å bruke Svenséns term. Jeg har imidlertid sympati for ordbokforfattere som legger større vekt på dette kriteriet enn egentlig nødvendig ut fra de fleste brukernes behov, fordi ordboken er ett av de få steder man vil ha tilgang på denne type informasjon i et privat bibliotek, uansett om man har jevnlig bruk for den eller ikke.

3.5.2. I norsk

I løpet av den siste generasjonen har et stort antall innvandrere fra andre kulturer slått seg mer eller mindre permanent ned i vårt land, og de vil lære fremmedspråk ved hjelp av lærebøker og ordbøker utgitt i Norge. For disse må det være viktig at deres egen kulturbakgrunn er så fyldig som mulig representert i disse læremidlene. Jeg har vanskelig for å tro at dette kriterium er satt i system av ordbokforfattere så langt. Tvert imot tar vel forfatterne sitt eget og fremmedspråkets kulturgrunnlag som en selvsagt referanseramme rundt det egne språkets ordforråd.

Det er en uoverkommelig oppgave å innlemme et større antall ord fra alle innvandrerklulturene i den norsk-fremmedspråklige delen av våre ordbøker, men det vil etter min oppfatning være riktig i dag å ta med noen viktige ord fra noen viktige innvandrerklulturene i ordbøker som skal brukes av elever fra mange forskjellige kulturer i norsk skole, ettersom disse innvandrerklulturene nå er en viktig del av norsk hverdag.

3.6. Kjønnsnøytralitet

Alle vil i dag forvente at moderne ordbøker avspeiler et samfunnstrinn med full likestilling, slik at ordforråd og eksempler yter begge kjønn størst mulig rettferdighet og slik at man ikke kritikkkløst, under henvisning til at man skal spare plass, videre-

formidler mannssamfunnets ordforråd og yrkesliv. I praksis er dette ikke lett, og da en ivrig medarbeider søkte å gjennomføre dette prinsipp i et av våre mindre ordbøker, tok jeg meg selv i å synes at det ble pedantisk. Det er nok ikke så dumt å bruke konsulenter som har dette som spesialområde.

3.7. Oppdragelse av brukeren

Svensén omtaler i avsnittet om det sosiale aspektet ved lemmasелеksjonen ønsket om at den passive ordboken skal inneholde flest mulig ord også fra sfærer som er mindre sosialt akseptable, men at det i den aktive ordboken neppe er like stort behov for oversettelse til fremmedspråket av mer spesielle slanguttrykk.

Særlig i ordbøker som brukes av yngre skoleelever, kan være ønskelig at ordboken avspeiler de verdier skolen bygger på. Ordbøkene skal formidle sunne, gode verdier ved å prioritere ord knyttet til positiv, byggende virksomhet, og ikke fokusere på områder som rusmidler, kontroversielle kulturformer og sex. Ordforrådet skal være pensumrelatert, og i tillegg skal man ta hensyn til at ordbøkene ikke bare brukes til oppslag, men i noen grad leses. Disse ordbokbrukerne bør ikke oppdras til å bruke ord som er sterkt kontroversielle i den eller de kulturer der fremmedspråket brukes.

3.8. Behandling av sammensatte ord

Det finnes gode argumenter for å ha med i tospråklige ordbøker mange oppslagsord der konstruksjon av ekvivalenten egentlig er selvnynlysende. Det gir fremfor alt brukeren større trygghet i omgangen med det fremmede språket enn når leksikografen begrenser seg til å ta med det ordmateriale som er leksikalisert. Brukere med svak språklig kompetanse vil nemlig i mange tilfeller ikke kjenne seg trygge på at de bruker det fremmede språket riktig ved å oversette ledd for ledd. (Se også Elzbieta Duras artikkel i denne boken.)

Vi er etter hvert så vant med å høre og lese engelsk at vi trolig får svakere evne til å skjelve mellom akseptert norsk språk og med eller mindre improviserte oversettelseslån. Ved å ha med norske ekvivalenter til mange engelske sammensatte ord som ikke er leksikalisert, vil ordboken gi kjærkommen navigasjons-hjelp i et vanskelig farvann.

3.9. Leksikografisk genre

Våre tospråklige ordbøker er tradisjonelle språkkordbøker, der bare unntaksvis realopplysninger inngår i ordbokartiklene. Artikkelinnholdet i våre allmennordbøker konsentrerer seg om grammatiske opplysninger, ekvivalenter, språklige forklaringer, eksempler og faste uttrykk, men vi går nok ikke av veien for realopplysninger der dette letter forståelsen, særlig i de ettspråklige ordbøkene. Spørsmålet om skillet mellom ordbok og encyklopedi har ikke først og fremst med lemmaseleksjonen å gjøre, men ettersom det vanligvis er langt flere egennavn i en encyklopedi eller en ordbok, og man har en mellomform som kalles encyklopedisk ordbok, der også innslaget av egennavn er større enn i en ren ordbok, velger jeg å si noen ord om vår behandling av egennavn her.

I våre interne retningslinjer for engelskordbøkene sier vi at "kjente egennavn (proprier), altså navn på land, byer, elver, guder, bibelske skikkelser, høytider, stjernebilder og lignende, tas med hvis de har ulik form på engelsk og norsk, f.eks. *Firenze/Florence*. Noen stedsnavn tas med fordi uttalen er vanskelig, for eksempel *Edinburgh*. Vi prøver også å få med egennavn det er vanlig å alludere til, f.eks. *Ascot, Wall Street*." Hva som er "kjente egennavn" vil så variere med seriens størrelse. Jeg antar at det blir nødvendig å søke opp alle slike ord i et sent stadium av redigeringen, og vurdere utvalget. Seleksjonen av egennavn i den norsk-fremmedspråklige delen bygger på at faste kategorier skal være med, slik at det er forutsigbart for brukeren hva han kan finne.

4. Avslutning

Dette har vært en rask gjennomgang av noen kriterier som jeg finner særlig relevante for vårt praktiske ordbokarbeid. Hvor forfattere og andre medarbeidere kan ha 4–5 av disse kriteriene in mente under arbeidet, og bruker dem til å begrunne sine avgjørelser, er jeg sikker på at det vil gi resultater. I våre serier vil jeg anse *brukergruppe x bruksmåte*, *frekvens* og *aktualitet* som de som bør prioriteres høyest uansett serie, *encyklopedisk representativitet* er svært viktig i de to største seriene, *kulturspesifikke hensyn* er viktigst i de to største og lommeordbøkene, men er også – med et lavere ambisjonsnivå – aktuelt for de små skoleordbøkene. *Kjønnsnøytralitet* er etter min mening viktigst i skoleordbøkene og de blå, *oppdragelse av brukeren* i skoleordbøkene, mens *behandling av sammensatte ord* og *leksikografisk genre* er noe vi må se mer på fremover. Men i tillegg til at det er viktig å benytte de riktige kriterier på riktig måte, er det viktig at de som utvikler teksten, gjør klart for utgiver og brukere hvilke kriterier de i praksis har lagt særlig vekt på.

Litteratur

- Axelsen, Jens. 1995. "Lemmaseleksjon i en dansk–engelsk ordbog." Ásta Svavarsdóttir m.fl. (red.): *Nordiske studier i leksikografi*. 3. Reykjavík. (Skrifter utgitt av Nordisk forening for leksikografi. Skrift nr. 3.) S. 19–26.
- Bergenholtz, Henning og Sven Trarp. 1994. *Manual i faglexikografi*. Herning: Systime.
- Duvå, Grete og Anna-Lise Laursen. 1995. "Tekstbasert lemmaseleksjon." Ásta Svavarsdóttir m.fl. (red.): *Nordiske studier i leksikografi*. 3. Reykjavík. (Skrifter utgitt av Nordisk forening for leksikografi. Skrift nr. 3.) S. 19–26.
- Helleland, Botolv. 1992. "Proprium og leksikografi." Ruth Vatvedt Fjeld (red.): *Nordiske studier i leksikografi*. Oslo. (Skrifter utgitt av Nordisk forening for leksikografi. Skrift nr. 1.) S. 190–199.

- Henriksen, Petter. 1992. "Forholdet mellom encyklopedi og leksikografi." Ruth Vatvedt Fjeld (red.): *Nordiske studier i leksikografi*. Oslo. (Skrifter utgitt av Nordisk forening for leksikografi. Skrift nr. 1.). S. 219–227.
- Svensén, Bo .1987. *Handbok i lexikografi*. Stockholm: Esselte.

Guðrún Kvaran

Tanker om en islandsk fremmedordbog

The topic of this paper is the definition of the term 'dictionary of foreign words' in HSK and NLO. The definitions of the terms *fremmedord*/*Fremdwort* and *låneord*/*Lehnwort* in some German and Danish dictionaries are also considered. At the end of the paper, the discussion turns to the possibility of making a dictionary of foreign words for Icelandic. One term, *tökuorð* ('loan word') is used both for foreign words and loanwords, and the conclusion is that for Icelandic the concept *dictionary of loanwords* is more to the point than the concept of a *dictionary of foreign words*. Such a dictionary should contain information both on actual loanwords, i.e. words that have been adapted to Icelandic phonology and morphology, etc., and truly foreign words, which have not done so.

Indledning

Jeg har valgt at skrive om temaet *fremmedordbog* og vil her diskutere hvad der hører med i sådan en bog og hvad man bør udelade hvis man tog fat på at skrive en ordbog over ord af fremmed oprindelse i det islandske sprog. Det kan næppe siges at man har vist udforskningen af låneord og deres indflydelse på ordforrådet nogen større interesse på Island i tidernes løb, med undtagelse af enkelte artikler, som især behandler høj- og lavtyske, saksiske, frisiske eller franske låneord i oldislandsk. Frank Fischer udgav 1909 en bog om låneordene i det gamle norrøne sprog og Christian Westergaard-Nielsens vigtige arbejde om låneordene i det 16. århundredes trykte islandske litteratur fra 1946 bør også nævnes og Veturliði Óskarsson, lektor i Uppsala, arbejder på en ordbog over låneord i islandske diplomer indtil 1500. Om nutidssproget findes der ingen større undersøgelser. Ganske vist behandlede Alexander Jóhannesson en god del yngre låneord i tillægget til sin etymologiske ordbog fra 1956 men største delen af hans lemmaer var ældre låneord og hele materialet blev behandlet ud fra en etymologisk synspunkt.

Om nutidssproget findes enkelte artikler, blandt andet om engelske låneord og deres tilpasning til sproget, f.eks. af Bruno Kress (1966) og Eyvindur Eiríksson (1975). Min kollega, Ásta Svavarsdóttir, og jeg har deltaget i et projekt om engelsk indflydelse på seksten europæiske sprog, især i de sidste årtier, og resultatet bliver udgivet i en ordbog hos Oxford University Press, med oversigtsartikler og bibliografiske referencer.

Ingen har taget fat på at samle til og udgive en islandsk fremmedordbog over nutidssproget, imens der er en rig tradition for den slags bøger i Tyskland og i Skandinavien, og opfattelsen og indstillingen til fremmede ord er nogenlunde den samme for dansk og tysk. I Frankrige og inden for det anglosaksiske sprogområde spiller fremmedordbøger ingen betydelig rolle.

Det er på tide at begynde at arbejde på en fremmedordbog i Island. For at udføre det arbejde må vi undersøge hvad vi kan lære af vore naboers erfaringer, men vi må først og fremmest gøre os klart hvad man forstår med ordet *fremmedordbog* og se om definitionen også gælder for islandsk. Det er også vigtigt at undersøge hvad man mener med termerne *låneord* og *fremmedord*? I et forsøg på at finde svar på disse spørgsmål vil jeg se på nogle definitioner af begreberne i danske og tyske almenordbøger hvorefter jeg vil undersøge hvorledes de bliver behandlet i nogle *fremmedordbøger*. Men først vil jeg slå op og se hvad der står i de tilgængelige leksikografiske håndbøger.

Hvordan defineres en fremmedordbog?

I opslagsbogen *Wörterbücher*, der blev udgivet hos de Gruyterforlaget 1990 i serien *Handbücher zur Sprach- und Kommunikationswissenschaft* (HSK 5.2:1168), er en *fremmedordbog* (Fremdwörterbuch) skildret som en ordbog der indeholder hele det fremmede ordforråd i et sprog. Derefter er lemmaer i tyske fremmedordbøger taget som eksempel. Til tyske fremmede ord hører så på den ene side låneord fra andre sprog (Wortentlehnungen) og på den anden side låneorddannelser (Lehnwortbildungen), d.v.s. ord som til dels eller helt er dannet ved hjælp af

låneord eller ordbildningselementer forekommende i låneord, men har intet direkte forbillede i eller overensstemmelse med det fremmede sprog. Videre står der, at den traditionelle definition af et fremmedord, d.v.s. ord som er lånt fra et fremmed sprog uden at være helt assimileret, skildrer kun en del af de ord som forekommer i tyske fremmedordbøger, d.v.s. kun låneordene men ikke låneorddannelserne. Låneord, som er helt tilpasset, som f.eks. tysk *Wein*, *Fenster*, *schreiben* får ikke plads i en fremmedordbog og heller ikke oversættelseslån (Lehnübersetzungen) f.eks. *Mitlaut* for *Konsonant* eller betydningslån (Lehnbedeutungen) f.eks. *Heiland* for latin *salvator*. Denne beskrivelse gælder i hovedsag de gængse tyske fremmedordbøger.

I *Nordisk leksikografisk ordbog* (NLO) beskrives en fremmedordbog således at den „behandler normalt dels ældre lånord der det udenlandske ophavet fremdeles er mærkbart”, og *flannel* og *karakter* er nævnt som eksempler, og „dels nylig innlånte ord med klart fremmed utseende og iblant også fremmed uttale” (1997:125) og her er eksemplerne *skateboard* og *ayatolla*. Videre står at man ikke plejer at tage med ældre låneord som *kirke*, *pærest*, *vin*, *brev* som er blevet helt tilpasset. En fremmedordbog indeholder efter denne beskrivelse både *låneord* og *fremmedord*.

I Bo Svenséns *Handbok i leksikografi* fra 1987 er skildringen af *Ordböcker över „främmande ord”* (23) helt identisk.

Låneord – fremmedord

Men hvilken forskel gør man så på *låneord* og *fremmedord*? I NLO er *fremmedord* ikke et opslagsord, men derimod *låneord*, som defineres som et ord der er blevet optaget fra et andet sprog og har tilpasset sig så godt at det med hensyn til udtale, retskrivning og bøjning bliver anset som en integreret leksikalsk enhed.

Pia Jarvad, som har skrevet en bog om nye ord i dansk, og nu for nylig har udgivet en ordbog over nye ord i dansk fra 1955–1998, mener at den klassiske definition af *låneord* og *fremmedord* i dansk er følgende: „låneord er ord som er lånt, som i en overgangsperiode har været følt fremmede, for derefter at være

akklimatiseret i en sådan grad at de ikke mere føltes fremmede, og som sådan ligestillede med sprogets oprindelige ord. Fremmedord er ord som er optaget fra et andet sprog, men bevarer deres fremmede ydre” (1995:61–62). Hun peger også på at i nogle sproghistoriske beskrivelser er ordet *låneord* brugt om „alt hvad der kommer udefra, og der skelnes ikke mellem fremmedord og låneord” (1995:62). Som eksempel nævner hun den sproghistoriske oversigt i *Nudansk Ordbog* hvor termen *fremmedord* nævnes som et synonym til låneord. Hun mener at man i almindelig sprogbrug i dag skelner anderledes imellem de to termer end før: „fremmedord betegner stort set ord som er vanskelige at forstå, uanset oprindelse eller deres mere eller mindre fremmedartethed eller tidspunkt for indlånet Det vil sige at efter den sprogbrug som er gængs i dag, er fremmedord det samme som sværtforståelige ord, og er ord som oprindeligt hører til et fagområde som fx teknik eller medicin” (1995:62). Efter Pia Jarvads mening er *låneord* derimod „nye ord som har deres udgangspunkt i et fremmed sprog. Låneord er både ord der fra starten hører til almenordforrådet og ord som har deres udgangspunkt i fagområder” (1995:62). Her er det det synkroniske synspunkt som bestemmer definitionen i stedet for det diakroniske i den klassiske definition.

Vi kan se på definitioner i tyske og danske monolingvale ordbøger: Ifølge den monolingvale *Duden* (1978) er *fremmedord* (Fremdwort) et ord som er optaget fra et andet sprog uden at være helt akklimatiseret hvad gælder udtale, skrivemåde eller fleksion. Som eksempel nævnes ordet „interessant”. En *fremmedordbog* er ifølge definitionen en ordbog som beskriver de fremmedord som bruges i sproget.

Låneord beskrives som ord som er optaget fra et fremmed sprog men har i udtale, skrivemåde og/eller fleksion tilpasset sig modtagerssproget. Som eksempel nævnes *Mauer* ’mur’ fra latin *murus*. Ifølge denne definition hører låneord ikke med i en fremmedordbog.

Definitionen af begge termer er stort set den samme i *Wahrig* (1975).

Ifølge *Nudansk ordbog* (1990) bruges *fremmedord* især om låneord, der vedblivende føles fremmede. Som eksempel nævnes ordet *cape*. Låneord er derimod „ord, der optages i et sprog fra et andet (modsat 'arveord'), spec. om ord, der i den grad er gået ind i sproget, at de ikke længere føles som fremmede (modsat 'fremmedord')”

Denne beskrivelse støtter sig til *Ordbog over det danske sprog* (ODS) fra 1923 hvor der står at et fremmedord er et „ord, der er optaget fra andet sprog; især om ord der føles som fremmede.”

Danske og tyske fremmedordbøger

Hvis vi nu ser på nogle danske og tyske fremmedordbøger med disse definitioner i tankerne er det klart at grænsen mellem de to typer ord er flydende. I Munksgaards gennemreviderede *Fremmedordbog* fra 1997 har man ikke forsøgt at definere hvad hører til i en fremmedordbog men det henvises til definitionen af fremmedord i ODS. I Kunnskapsforlagets *Fremmedord blå ordbok* er der heller ikke en definition af fremmedord.

I fortalen til Gyldendals *Fremmedordbog* blev der citeret fra Sven Brüels forord til 8. udgave hvor han definerer fremmedord som „låneord, der i form ikke har tilpasset sig arveordene (*diakon* føles fremmedartet, *degn* ikke, men de har samme græske oprindelse)” (1993:[5]). Hans oprindelige formål med bogen var at hjælpe læseren med at „finde ud af hvad mange nydannede ord, der endnu ikke kan være optaget i ordbøgerne, betyder.” Hvis man sammenligner den nye udgave af fremmedordbogen med *Nudansk ordbog* ses at ganske mange ord, som forekommer i den førstnævnte også hører hjemme i en almenordbog.

I Gjellerup og Gads *Fremmedordbog* fra 1988 er der en kort fortale hvor der er gjort et forsøg på at definere *fremmedord*. Udgiverne kommer ind på udtrykket „føles fremmede” og mener at det er vanskeligt at få tag på, og at de, i mangel af mere præcise kriterier, vil definere fremmedord som „ord, hvis udseende, udtale, bøjning eller brug virker fremmedartet på mange men-

nesker” (5). De indrømmer at denne definition fører til at en fremmedordbog kommer til med at indeholde de ord, som dens forfattere selv finder fremmedartede, d.v.s. valget kan kun være subjektivt. Eksempel på et ord, som de mener at ikke hører til i en fremmedordbog er *traktor* fordi det helt selvfølgelig indgår i hverdags sproget. Derimod skal f.eks. ordet *bælg* med i bogen fordi forfatterne mener at mange yngre sprogbrugere er usikre over for det. I selve ordbogen har *bælg* tre forskellige betydninger: 1) frøkapsel, 2) skindet på et mindre dræbt dyr, 3) rør, kasse eller sækkelignende beholder af blødt materiale. Her ville jeg tro at man havde med et arveord at gøre, som måske var gået lidt ud af brug, men ikke et fremmedord efter den gængse opfattelse. Her ville det være svært at trække grænsen.

I Dudens *Fremdwörterbuch* (1997:7) findes der en artikel om fremmedordets historie og funktion, hvor det indrømmes at der findes ingen pålidelige kriterier på hvad et fremmedord er. Fire punkter nævnes som man kunne støtte sig til :

- (1) Orddannelsen, hvor bestemte præ- eller suffixe stammer fra fremmede sprog, f.eks. *Kon-* og *-tion* i *Konfrontation*.
- (2) Udtalen, hvor den afviger den gængse tyske.
- (3) Skrivemåden, hvor den afviger fra den vanlige tyske ortografi.
- (4) Ordet forekommer sjældent i det daglige sprog og føles derfor som fremmed. De eksempler, som her nævnes er fleste taget til låns fra latin eller græsk.

Forfatterne lægger dog vægt på at disse kriterier kan kun betragtes som en mulighed og hjælp til at identificere fremmede ord, ikke som regler.

En eventuel islandsk fremmedordbog

Men nu vil jeg gå over til at diskutere en eventuel islandsk fremmedordbog og spekulere på hvilke oplysninger den må give

for at være til mest mulig nytte. Det gælder såvel den almindelige sprogbruger som sprogforskere, som vil undersøge fremmedordenes tilpasning til sprogets fonologiske og morfologiske system eller ordforrådets historiske udvikling.

På grund af den store rolle sprogpurismen har spillet på Island i det sidste halvandet århundrede var der længe relativt få fremmede ord i skriftsproget men en god del gamle låneord har levet videre i talesproget, foruden at nye stadig kommer til. Termen *tökuorð* for *låneord* er blevet brugt i det mindste i de sidste 80 år. Den defineres i *Íslensk orðabók*, en ensproglig ordbog over det islandske sprog, som et ord som et sprog har taget op fra et andet, d.v.s. den bliver brugt både for låneord og fremmedord ifølge definitionen i de danske og tyske bøger jeg slog op i. Termen *framandorð* for *fremmedord* er derimod ganske ny og jeg kender ingen definition af den på islandsk. Så vidt jeg ved blev ordet *framandorðabók* først brugt i NLO som et islandsk ord for *fremmedordbog*.

Vore naboers forhold til fremmedordbøger kan, så vidt jeg kan se, ikke uden videre overføres til en islandsk ordbog. I islandsk har man ikke skelnet mellem fremmedord og låneord, men begrebet *tökuorð* omfatter både assimilerede og ikke helt assimilerede ord. Inddelingen i to grupper tager ikke hensyn til de aspekter som gælder ved brugen af enkelte ord. Men hvis man vil indskrænke betydning af *låneord* til de ord, som er helt assimilerede og accepterede af alle, så hører dertil ord som f.eks. *prestur* 'præst', *biskup* 'biskop', *altari* 'alter', *kirkja* 'kirke', *skammel* 'skammel', *orgel* 'orgel', *kex* 'kiks', *blússa* 'bluse', *fægiskófla* 'fejebblad', *grænsápa* 'grønsæbe', *brúnsápa* 'brunsæbe', *viskastykki* 'viskestykke' og mange mange flere.

Fremmedordene må man derimod dele i to grupper. Til den første hører de ord som „føles fremmede” i den betydning, som jeg nævnte før, d.v.s. de har for det meste føjet sig efter sprogets regler men er endnu ikke blevet accepteret af alle. Disse ord bliver undertiden kaldt *aðskotaorð* 'uvedkommende ord' eller måske 'snylteord'. Til den anden gruppe hører al slags ord som bliver brugt i uformel tale, hvor man for bekvemmelighedens skyld eller måske for at slå om sig med kundskaber i fremmede

sprog, bruger fremmede ord uden at de er blevet assimileret i nogen grad. De er på islandsk kaldt *sletta* som betyder 'klat, stænk'. Ingen af disse ord kan dog bruges som lingvistiske termini, dertil er de for negativt belastede.

Baldur Jónsson, professor ved universitetet i Island, har i en artikel gjort opmærksom på problematikken i terminologien inden for det fremmede ordforråd. Han foreslog hvilke krav man skulle stille et fremmedord hvis det skal accepteres som en del af det islandske ordforråd (1997:16). Han opstiller fem kriterier:

- (1) Trykket må være på første stavelse.
- (2) Et fremmedord må følge islandske fonologiske regler.
- (3) Et fremmedord må følge islandske fonotaktiske regler.
- (4) Et fremmedord må tilhøre en bøjningsklasse, der allerede eksisterer i sproget, følge dens bøjningsregler og andre specielle regler.
- (5) Et fremmedord må skrives med islandske bogstaver således at regelmæssig forbindelse findes mellem skrivemåde og udtale.

Hvis et ord opfylder disse betingelser kan det betragtes som et låneord, men andre ord, som ikke opfylder disse betingelser, men er alligevel brugt i sproget, dels assimileret, dels slet ikke, vilde så falde i gruppen *fremmedord*. Lad os se lidt nærmere på disse kriterier.

Det er en gældende regel i islandsk at tryk ligger på første stavelse. Vi siger '*hótel* men ikke *hó'tel*, '*píanó* men ikke *pí'anó*, '*sardína* men ikke *sar'dína*, '*kartafla* men ikke *kar'tafla*. Ligger trykket på et andet sted i ordet så går det ikke som låneord, det er da et uassimileret fremmedord.

Hvad de fonologiske regler angår så må et fremmedord kun bestå af fonemer som eksisterer i sproget. Det kan ikke medbringe nye fonemer. Baldur Jónsson tager *typa* 'type' og *typískur* 'typisk' som eksempel. Udtalen af *y* som rundet vokal er længst forsvundet i sproget og faldet sammen ved et urundet *i* eller *í*. Ordet *typa* med *y*-udtale kan derfor ikke regnes som tilpasset til islandsk selvom det bruges af ganske mange. Det kan

derimod *típa* og *típískur* som man ville sætte i gruppen assimileret fremmedord.

Af Baldur Jónssons artikel fremgår at det er svært at definere hvad er en „islandsk” fonotaktisk regel. Hvis man ser kun på arveordenes opbygning så ville man udelukke en mængde låneord som blev optaget i sproget før år 1500. Med gamle assimilerede fremmedord har man allerede taget op fonologiske strukturer som ikke fantes i arveordene. Det gælder f.eks. ord hvor dobbelkonsonant følger et *ú*, f.eks. *skúffa* 'skuffe', *grúppa* 'gruppe', *tútta* 'sut', *skrúbbur* 'skrubbe'. Disse ord er blevet accepteret men føles dog af mange som fremmede.

Hvad bøjningen angår så må et fremmedord falde i en bøjningsklasse hvis det skal kunne accepteres som en del af ordforrådet. Som bekendt er substantivbøjningen i islandsk ganske kompliceret. De fleste substantiver falder ganske let i en bøjningsklasse efter genus. Det samme kan siges om verberne. De får alle infinitivendelsen *-a* i stedet for et fremmed *-e* eller ingen endelse, f.eks. *grilla* 'grille', (egl. 'grill'), *blöffa*, egl. 'bluff'. Det samme gælder ikke adjektiverne. En del af dem er ganske vist assimileret hvad bøjningen angår, f.eks. *stabíll*, *pósitífur*, andre er kun delvis assimileret, f.eks. *smart*, *flott*, *röff* (e. *rough*), *töff* (e. *tough*).

Den sidste regel om skriveformen er også ganske svær. Som Jón Hilmar Jónsson (1980:62) har peget på har man holdt fremmedordene uden for retskrivningen. De er hverken taget med i islandske retskrivningsregler eller lærebøger og det er derfor vanskeligt at danne sig et klart billede af den skrivemåde der gælder i fremmedord i islandsk. Andet som gør undersøgelsen vanskelig er at selvom fremmedord er ganske almindelige i talesproget vil de fleste undgå dem i skriftsproget. Men hvis man bruger dem overhovedet prøver de fleste at skrive ordene efter islandsk udtale, f.eks. *fila* 'feel', *frika* 'freak', *djús* 'juice', *séni* 'geni' og *teip* 'tape'. I andre tilfælde findes både den fremmede skrivemåde og den islandske. F.eks. forekom ordene *beat* og *bít* i den samme artikel i en islandsk avis for nylig og et andet eksempel er *pítsa* hvor den mere almindelige skrivemåde er *pizza*. Det

senere er et uassimileret fremmedord, det første et assimileret ord som „føles fremmed”.

Før man begynder at arbejde på en islandsk fremmedordbog er det vigtigt at man er nogenlunde enig om definitioner af termene *låneord* og *fremmedord* og hvor langt man vil gå i begge retninger. Hvis vi forstår termen *framandorðabók* på samme måde som man gør i dansk (og tysk) *fremmedordbog* så ville helt assimilerede låneord som er accepteret af alle men er alligevel af fremmed oprindelse ikke høre dertil. Jeg er næsten sikker på at den islandske sprogbruger ville forvente at en god del sådanne ord også var med i bogen. Men det ville være alt for stort et arbejde at tage med alle låneord i islandsk. Man kunde starte enten omkring år 1900 eller i midten af det 19. århundrede og tage med alle de ord af fremmed oprindelse som er integreret i sproget til dags dato. Ganske mange fremmedord skal også med i sådan en bog. Hvis man arbejder efter Baldur Jónssons regler kan man ret nemt bestemme hvad et nogenlunde assimileret fremmedord er, selvom man vil stå overfor mange vanskelige problemer. Andre fremmedord skulle man udelade og vente på at de indordner sig også under sprogets regler. Måske er termen *tökuorðabók*, d.v.s. låne- og fremmedordbog, nærmere det vi behøver.

Bibliografi

- Duden* 1997 = *Duden. Das Fremdwörterbuch*. 6., auf der Grundlage der amtlichen Neuregelung der deutschen Rechtschreibung überarbeitete und erweiterte Auflage. Duden Band 5. Mannheim–Leipzig–Wien–Zürich: Dudenverlag.
- Eiríksson, Eyvindur. 1975. ”Beyging nokkurra enskra tökuorða í nútímaíslensku.” *Mímir, blað félags stúdenta í íslenskum fræðum* 23:55–71.
- Fischer, Frank. 1909. *Die Lehnwörter des Altisländischen*. Palaestra 85. Berlin: Mayer & Müller.
- Fremmedordbog* 1988 = *Gjellerups fremmedordbog*. Af Fleming Albertus. København: Gjellerup & Gad.

- Fremmedordbog* 1993 = Sven Brüel og Niels Åge Nielsen. *Fremmedordbog*. Gyldendals røde ordbøger. 11. udgave, 5. oplag ved Lilian Plon. København: Gyldendal.
- Fremmedordbog* 1997 = Karl Hårbøl, Jørgen Schack, Henning Spang-Hanssen. *Dansk fremmedordbog*. 15. udgave. Munksgaards ordbøger. København: Munksgaard.
- Bjarne Berulfsen og Dag Gundersen. 1999. *Fremmedord blå ordbok*. Femtende udgave. Oslo: Kunnskapsforlaget.
- Íslensk orðabók handa skólum og almennungi. 1983. Red. Árni Böðvarsson. Reykjavík: Bókaútgáfa Menningarsjóðs.
- Jarvad, Pia. 1995. *Nye ord – hvorfor og hvordan?* Gyldendal: København.
- Jarvad, Pia. 1999. *Nye ord. Ordbog over nye ord i dansk 1955–1998*. København: Gyldendal.
- Jóhannesson, Alexander. 1951–1965. *Isländisches etymologisches Wörterbuch*. Bern: Franke.
- Jónsson, Baldur. 1997. "Zur formalen Anpassung von Fremdwörtern im Isländischen." *Skandinavistik* 27(1):15–23.
- Jónsson, Jón Hilmar. 1980. "Om skivemåte og bøyning av fremmedord i islandsk." *Språk i Norden 1980*. Årsskrift for språknemndene i Norden og Nordisk språksekretariat, 61–67.
- Kress, Bruno. 1966. Anglo-Amerikanisch und Isländisch. *Nordeuropa. Jahrbuch für Nordische Studien* 1:9–22.
- NLO 1997 = *Nordisk leksikografisk ordbok*. Red. Ruth Vatvedt Fjeld [et al.] (*Skrifter utgitt av Nordisk forening for leksikografi*. Skrift nr. 4.) Universitetsforlaget, Oslo.
- Nudansk ordbog* 1990. = *Politikens nudansk ordbog*. 14. udgave. København: Politikens forlag.
- ODS 1919–1954 = *Ordbog over det danske sprog*. I–XXII. København: Gyldendalske boghandel, Nordisk forlag.
- Svensén, Bo. 1987. *Handbok i lexikografi*. Stockholm: Esselte Studium.
- Wahrig 1974 = Wahrig, Gerhard. *Deutsches Wörterbuch*. Gütersloh, Berlin, München, Wien: Bertelsmann Lexikon-Verlag.

Westergaard-Nielsen, Christian. 1946. *Låneordene i det 16. århundredes trykte litteratur. (Bibliotheca arnamagnæana. 6.)* København: Munksgaard.

WSK 1990 = "Das Fremdwörterbuch." *Wörterbücher, Dictionaries, Dictionnaires*. Ein internationales Handbuch zur Lexicographie. 5.2:1168–1178. Berlin, New York: Walter de Gruyter.

Åsa Holmér

Vad ska *chatta* heta på finska?

Efcot-projektet, att bygga upp ett europeiskt forum för dataterminologi

In Sweden today almost everyone comes into contact with computers. However, when speaking or writing about computers, English terms are often used for lack of Swedish ones. In order to give recommendations on which Swedish computer terms to use, the Joint Group for Swedish Computer Terminology, JOGSCOT, was established in 1996. JOGSCOT includes national expertise both in computer technology and in terminology, general language planning, standardization and translation – a fact which helps to pave the way for the recommendations given. Before being published on the JOGSCOT web site, which is open and free of charge for all Internet users, the suggested recommendations are discussed mainly via e-mail.

The MLIS programme (Multilingual Information Society) of the the European Commission is now financing a project called WebIT/EFCOT. The EFCOT part of this project is based on the work carried out within JOGSCOT and aims at exporting its method of working to Finland, Norway and Greece. A second aim within EFCOT is to offer a European forum for discussions on computer terminology – to begin with as a joint web site, the EFCOT site.

För tre år sedan var jag med om att bilda Svenska datatermgruppen här i Sverige. Gruppens arbete blev nästan omgående uppmärksammat, och vid ett nordiskt möte inom Nordterm (samarbetsorganet för terminologer i Norden diskuterade vi därför hur vi skulle kunna upprätta liknande grupper i övriga nordiska länder. Endast ett halvår senare uppenbarade sig en möjlighet att åtminstone delvis finansiera detta. En ansökningsinbjudan kom från MLIS-programmet (Multilingual Information Society), som är ett EU-program med syfte att främja mångspråkighet i samhället genom att ge pengar till projekt som bygger upp informationsteknisk infrastruktur. Vi (TNC) sände in ett projektförslag och fick det antaget.

Projektets officiella namn är WebIT/EFCOT. Projektet är nämligen tvådelat. WebIT-projektet har sin bas på Irland och har

som mål att samla dataterminologi inom informationsteknik och programvaruproduktion i en webbaserad flerspråkig databas.

EFCOT (European forum for computer terminology) som jag ska berätta om här i dag har två syften:

- Det första är att bygga upp datatermgrupper i de deltagande länderna. För en dryg månad sedan inleddes arbetet i Finland, och för några dagar sedan hölls det första mötet i en kommande norsk datatermgrupp.
- Det andra är att bygga upp en samordning av dataterminologi över nationsgränserna, i form av ett europeiskt forum för diskussioner.

TNC står som samordnare och våra systerorganisationer i Finland respektive Norge, TSK och RTT, är partner. Dessutom finns ytterligare en partner: det grekiska företaget Archetypon (som har stor erfarenhet från arbete med programvaror och språkteknologi).

Bakgrunden till det hela är egentligen Svenska datatermgruppen, så därför tänker jag börja med att beskriva hur Svenska datatermgruppen fungerar.

Svenska datatermer i stället för engelska

För tre år sedan hade tidningarna precis börjat skriva om företeelser som Internet, world wide web, web sites, e-mail, web master. De språkvårdande institutionerna TNC och Svenska språknämnden hade känt av en ökning av frågor om dataterminologi. Och frågorna kom inte bara från specialister inom området, utan från översättare och tekniska skribenter liksom från journalister som behövde skriva och tala om den nya tekniken.

De ville inte använda citatord från engelskan; de frågade oss efter svenska motsvarigheter. Detta ledde till att Svenska språknämnden kallade till möte: vi ville samla personer med intresse och kunskande inom detta område för att diskutera formerna för ett nätverk för datatermer på svenska.

Svenska datatermgruppen bildas

Intresset visade sig vara mycket stort, och vid det första mötet (den 7 maj 1996) bildades Svenska datatermgruppen. Det är en brett sammansatt grupp med företrädare för databranschen, medierna och språkvården.

I gruppen ingår både översättningsföretag som producerar handböcker (Interverbum, WM-data, Wordwork, Translator teknikinformation), tillverkare, som har sista ordet när program och handböcker ska översättas (IBM, Microsoft, Apple, Sun Microsystems) och datatidningar, där datoranvändarna ofta första gången läser om nya produkter och företeelser och vad de kallas (datatidningsförlaget IDG och ytterligare några datatidningar).

Dessutom finns representanter för språkvården (TNC, Svenska språknämnden, Svenska Akademien, Informationstekniska standardiseringen, Svenska språkbyrån i Finland, Norsk språkråd) men också för medierna (översättningsföretaget Svensk text, Hufvudstadsbladet, ett par reklambyråer samt språkvårdarna på de stora dagstidningarna, på Sveriges Radio och TT) högskolor och en del andra större institutioner och företag (Posten, Telia, Ericsson). Som samordnare fungerar TNC och Svenska språknämnden.

Så fungerar gruppen

Redan vid det första mötet gick diskussionens vågor höga, så vi beslöt rätt så omgående att organisera oss i mindre grupper, för att göra arbetet mer effektivt. Svenska datatermgruppen består i dag därför av en koordinationsgrupp, en sammanträdesgrupp och en remissgrupp.

Koordinationsgruppen är ytterst ansvarig för att formulera och publicera termgruppens rekommendationer. Gruppen, som träffas ungefär en gång i månaden, består för närvarande av åtta personer, med en jämn fördelning mellan datatekniska och språkvårdande representanter. Till varje enskilt möte kan olika per-

soner med relevanta kunskaper om de aktuella begreppen bjudas in.

Sammanträdesgruppen träffas tre gånger per år. Den består egentligen av representanter för de företag och organisationer som bjöds in till det första mötet, liksom koordinationsgruppens medlemmar. Sammanträdesgruppen diskuterar i första hand riktlinjer, organisatoriska och administrativa frågor samt särskilt viktiga eller svårhanterliga begrepp. I mån av tid diskuteras också de begrepp som koordinationsgruppen för tillfället arbetar med. Syftet med gruppens möten är dessutom att erfarenheter och tankar ska utbytas över bransch- och företagsgränserna.

Remissgruppen träffas bara i cyberrymden, via e-post. Den fungerar som remissinstans för de rekommendationsförslag som formuleras av koordinationsgruppen. Det är alltså i den gruppen som själva diskussionen om databegreppen äger rum. I remissgruppen ingår samtliga deltagare i sammanträdesgruppen, men också andra personer som har särskilda kunskaper eller på annat sätt kan bidra med goda synpunkter. Hit välkomnar vi hela tiden nya medlemmar.

Det viktigaste i arbetsmodellen är arbetsgången där de representativa och inflytelserika personer från det breda spektrum av företag och institutioner som ingår i Svenska datatermgruppen får säga sin mening, vilket leder till att rekommendationerna får en bred förankring.

Vilka begrepp tar gruppen sig an?

Svenska datatermgruppen har som mål att ta sig an begrepp inom vad man skulle kunna kalla det datatekniska allmänspråket, dvs. sådan terminologi som vi alla kommer i kontakt med när vi använder våra datorer. Däremot skall det vi rekommenderar vara tillräckligt väl förankrat i fackområdet för att en ämnesexpert också ska kunna följa rekommendationen i fråga. Vi har avgränsat ämnet till att gälla termer inom Internetvärlden i synnerhet. Termerna ska helst vara relativt aktuella, dvs. inte ännu ha en etablerad svensk form. I andra hand tar vi upp termer och be-

grepp som är mer tekniska eller som är etablerade, men som många människor är osäkra över (t.ex. när man ska välja *data-* respektive *dator-* som förled i sammansättningar).

Egen webbplats

Redan efter några månader presenterade vi våra första rekommendationer på en egen webbplats. Version 14, som publicerades för ett par veckor sedan, innehåller 116 termartiklar där runt 150 begrepp diskuteras.

Rekommendationer

En termartikel har följande format (se exempel nedan): Överst anges den eller de termer vi rekommenderar. Sedan följer ett slags definition följt av den egentliga rekommendationen. Till sist har vi också möjlighet att i ett kommentarfält ge fler synpunkter på rekommendationen:

bärbar dator

dator som är utformad så att den är lätt att bära med sig och som i regel är försedd med egen strömförsörjning

Använd *bärbar dator* för eng. *laptop computer*, *notebook computer* och andra engelska benämningar för detta begrepp.

Kommentar

Det svenska uttrycket *portföljdator* förekommer ibland för detta begrepp.

Det finns en öppen dörr mot användaren. Det innebär att vi emellanåt inte ger ett enda förslag till rekommenderad term, utan vi tillråder användaren att skapa egna ord. Ett exempel är här rekommendationen för svensk term till *help desk*:

hjälpcentral

funktion i ett företag eller en organisation med uppgift att hjälpa anställda eller kunder att lösa problem med framför allt datorhantering eller databehandling

Eng. *helpdesk* kan översättas med *hjälpcentral*. Man kan också namnge hjälpcentralen mer precist alltefter arten av hjälp som tillhandahålls, t.ex. *användarhjälp*, *användarstöd*, *informationstjänst*, *kundservice*, *kundtjänst*, *språkrådgivning*, *termtjänst*, *datorhjälp*, *datorjour*, *dataakut*.

Förutom rekommendationerna har vi andra webbsidor på vår webbplats: Vi har en presentation av vår verksamhet. Vi beskriver också principerna för termarbetet. Vi har en avdelning med frågor och svar där mer obearbetade terminologiska problem tas upp till diskussion och en avdelning för skrivregelsfrågor. Dessutom har vi en intern webbplats (som man kan komma till om man har ett lösenord) från vilken vi kan förmedla information som framför allt koordinationsgruppen behöver.

Olika sändlistor

Kommunikationen med Svenska datatermgruppen och mellan medlemmar i de olika undergrupperna av Svenska datatermgruppen sker nästan uteslutande via e-post. Svenska datatermgruppen har därför ett antal sändlistor som kanaliserar e-postkommunikationen både inom gruppen och mot allmänheten. Det gör att kommunikationen fungerar lätt och smidigt.

Förutom sändlistor till koordinationsgruppen, sammanträdesgruppen och remissgruppen finns följande sändlistor:

Datatermnytt

Via denna sändlista meddelar vi nyheter, t.ex. då nya rekommendationer blivit klara, eller om det skett några andra förändringar

på vår webbplats. Sändlistan är alltså till för att sprida kortfattad information från Datatermgruppen till deltagarna, inte för diskussioner mellan deltagarna.

Anmälan till datatermnytt

Hit skickar man sin e-postadress och blir då uppsatt på sändlistan datatermnytt.

Allmänhetens adress till gruppen (går till koordinationsgruppen)

Hit kan vem som helst skicka in kommentarer till våra rekommendationer eller önskemål om angelägna begrepp som gruppen borde ta upp i framtiden.

Respons utifrån via e-brev och sökstatistik

Vi får en signal om vilka begrepp vi ska ta oss an genom de e-brev vi får till gruppen. Dessutom kan vi med hjälp av en funktion som registrerar de sökningar som görs ta ut listor över vilka uttryck som folk mest sökt på och på så sätt se vilka begrepp vi ännu inte har behandlat.

Stor uppmärksamhet

Som jag sade inledningsvis märkte vi redan i början av gruppens arbete att många har uppmärksammat våra rekommendationer. Åtskilliga tidningar har publicerat artiklar om gruppens arbete. Representanter för gruppen har blivit intervjuade i radio och för utbildningsprogram i tv etc. Medarbetarna hos de institutioner och företag som deltar i gruppen har uppmanats att följa rekommendationerna (t.ex. alla medarbetare inom koncernen Sve-

riges Radio och Television och inom Posten), och många har skapat länkar till gruppens webbsidor från sina egna webbplatser. Det har tillsammans med den uppmärksamhet gruppen fått resulterat i att antalet anrop till Svenska datatermgruppens webbsidor redan från början varit högt (senaste statistik från mars månad 1999 är 22 725 anrop eller 800–1 000 per vardag), och reaktionerna har varit övervägande positiva.

Vi har ju undrat över denna uppmärksamhet. Här är några möjliga skäl, av vilka de två första är väldigt viktiga:

1. Fackmän inom det datatekniska området är med i gruppen och påverkar arbetet. Dessutom tar vi hänsyn till bruket, i den mån det finns något bruk.
2. Vi kan sprida rekommendationerna genom de inflytelserika medlemmarna, däribland journalister, en mycket viktig grupp att ha med: de som skriver och talar om fackområdet i medierna.
3. Kommunikationen sker via e-post, och vi har en webbplats som inte kostar något att använda. Detta gör att vi snabbt och effektivt kan diskutera frågor och smidigt kan sprida rekommendationerna. Just dataterminologin passar för detta, men självklart kan arbetsmetoden också utnyttjas för andra ämnesområden.
4. Det ligger i tiden! Numera har många persondatorer, och i och med detta är dataterminologi i allmänt bruk. Om vi hade startat verksamheten för tio år sedan hade arbetet förmodligen inte väckt samma uppmärksamhet.

Olika språk, snarlika lösningar

Jag tror att vi inom detta ämnesområde oavsett språk har mycket gemensamt. Det finns därför goda skäl att ha ett forum där vi kan diskutera, prata, chatta om terminologi med varandra över språkgränserna.

I Finland och Norge har redan arbetet med nationella datatermgrupper inletts. Och vid en konferens nyligen inom för-

eningen EAFT (European Association for Terminology) fick jag reda på att man i Spanien inom organisationen Termcat arbetar på ett likartat sätt med att rekommendera katalanska datatermer med hjälp av e-post och en webbplats. Snart är ett europeiskt forum ett faktum.

Dimitrios Kokkinakis
Sofie Johansson Kokkinakis

Sense-Tagging at the Cycle-Level Using GLDB

This report describes a large-scale attempt to identify automatically the appropriate sense for content words taken from Swedish open-source texts. Sense-tagging, 'the process of assigning the appropriate sense from some kind of lexicon to the (content) words in a text', is a difficult and demanding task in Natural Language Processing and researchers have been engaged in finding a suitable solution to the problem for a very long time. The usefulness of automatically assigning each word in unrestricted text with its most likely sense is necessary for a great spectrum of applications. The sense-tagger described here has been tested both on a random sample of content words, as well as on a large population of a single ambiguous entry. In the first case, the achieved precision was 84,21%, and in the second 82,75% respectively. Evaluation was made against manually sense-annotated texts.

1. Introduction

One of the many problems encountered in Natural Language Processing is that of semantic lexical ambiguity. This means that deciding which meaning of a word is intended in a given utterance or discourse is a very difficult task, that humans usually perform without even consciously noticing that the ambiguity exists. While a native speaker of Swedish can almost immediately recognize that in the following four examples the verb *handla* refers to four different meanings, i.e. 'to take action', 'to trade', 'to buy', 'to deal', the same task in computer processing is a major headache:

- (1) *Polisen handlade snabbt denna gång.*
'The police acted quickly this time.'
- (2) *EU har åter börjat handla med Kina.*
'EU has started trading with China again.'
- (3) *John handlade mat för 1000 dollar.*
'John bought food for 1000 dollars.'
- (4) *Filmen handlade om en ung mans väg till framgång.*
'The film was about a young man's way to success.'

The idea of performing sense disambiguation is a controversial matter in many respects, and it has been discussed and sometimes criticized with respect to whether *how* and *if* it can possibly be done. Criticism has been directed against several attempts at automatic sense disambiguation, and there are no simple answers to the otherwise well-justified questions associated with the issue whether it is feasible or not to make clear sense distinctions. Some of the criticism is based on the term 'sense' itself, which is not a well-defined concept; problems referring to the fact that humans cannot agree on what sense is appropriate for the words in a given sentence; sense distinctions are interpreted differently by different researchers, following different approaches to the disambiguation problem; and finally, that dictionaries differ substantially regarding the different sets of senses for the same word. Despite the criticism, we regard sense-tagging as a very important process and component within a wider and deeper text-processing architecture.

2. Sense and Semantic Tagging

Many words in the dictionaries have multiple senses or meanings, while, when a word is actually used in a context, just one of these meanings generally applies. By the term *sense*, we here mean *dictionary sense*. For instance, *development* may be a highly ambiguous word in English, but for photographers it refers unambiguously to *processing a film*, and for an architect it refers to *building*. Sense-tagging should not be confused with *semantic tagging*, which is a more general case, in which the labels assigned to the words in a text are broad semantic categories, or clusters of semantically related concepts. Semantic categories may be labels of the form ANIMATE, ARTIFACT, LOCATION or HUMAN; or WordNet synsets (Fellbaum 1998), such as *Life and Living Things* and *Food, Drink and Farming*. Despite their subtle differences, both sense and semantic tagging aim at the resolution of lexical ambiguity, either on a small or large scale. Furthermore, the more general term of

Word Sense Disambiguation or WSD, is used in the context of both. Both terms, *sense-tagging* and *WSD*, will be used interchangeably.

WSD tries to solve lexical ambiguity, which in turn is closely related to two lexical semantics concepts, that of *polysemy* (related word senses) and *homonymy* (unrelated word senses). There is no clearcut border between these two concepts. From the point of view of WSD the difference between these two concepts is not a controversial issue, lexical ambiguity, in the context of automatic means of WSD, refers to both.

3. Background

The different approaches to lexical disambiguation that will be discussed here are classified according to the major source of information that researchers have used for the WSD task. Different classification schemes can be found in Wilson & Thomas (1997), in which they distinguish between manual, computer-assisted and fully automatic methods to WSD; Fujii (1998), in which he distinguishes between qualitative and quantitative approaches; and Sanfilippo *et al.* (1998 §5.3.2), in which they distinguish between knowledge-based, corpus-based and hybrid approaches.

3.1. WSD Using an Explicit Dictionary

Kelly & Stone (1975) manually developed a dictionary for approximately 2,000 words by studying senses from a word corpus of half a million words and writing disambiguation rules for each multi-sense word. For their help they had key-word-in-context (KWIC) concordances and the 1966 *Random House Unabridged Dictionary*. Their work was labour-intensive and manual, and the local context was their main source of information used. The manually-written disambiguation rules,

each corresponding to a sense in the dictionary, were used by an algorithm achieving 90% accuracy.

The use of standard dictionaries is contributed to Lesk (1986). He was the first to suggest the use of dictionary definition overlap for WSD. Lesk proposed that a sentence could be disambiguated relative to a dictionary by choosing the configuration of senses that maximizes the number of words which are common to the textual definitions and the context of the word to be disambiguated. The quality of the results on his experiments lie within the 50–70% correct sense distinction.

Cowie *et al.* (1992), applied the simulated annealing technique to WSD. They were the first research team that used this method for WSD, in conjunction with the LDOCE, reporting 72% correct assignment of senses.

Wilks & Stevenson (1998) also used LDOCE for disambiguation. Their approach was based on combining different knowledge sources for achieving qualitatively better results, than merely using the definitions. The algorithmic details behind their high figures on precision rely on the use of an optimized version of the simulated annealing technique. Wilks & Stevenson are one of the very few research teams that have attempted sense disambiguation on all content words in a text, achieving 94% correct sense assignment.

3.2. WSD Using Thesauri and Ontologies

A few knowledge bases often discussed in connection with semantic disambiguation, such as the *WordNet* and the *Roget's Thesaurus*, The Princeton WordNet, Miller *et al.* (1990), and the *EuroWordNet*, Peters *et al.* (1998), are the most commonly used networks for disambiguation considering the relevant bibliography.

WordNet has a predominant position since it is publicly available, and has been extensively studied for quite some time by a number of different research teams. Miller *et al.* (1993) used WordNet for linking content words from a text to their appropri-

ate sense in the lexicon. This was viewed by the Miller group either as a corpus, in which words have been tagged semantically, or as a lexicon in which example sentences can be found for many definitions.

3.3. WSD Using Information from Corpus

Some of the reasons in favour of using corpora for the WSD, and not using a dictionary or a semantic net, are the following: that copyright constraints are usually associated with the dictionaries; that dictionary descriptions are just a static view of a language in a particular time frame, while language is a dynamic system in constant change that can be better and more easily captured by monitoring text corpora; and finally, that imperfections and incomplete coverage are usually tied to the lexical resources.

The approaches can be divided into three different methods. (i) Supervised methods use as their primary source of information a *disambiguated corpus*. The annotated corpus is then used for the supervision and induction of rules, which are fed into stochastic models, and which can predict the correct sense of words in new contexts (*cf.* Yarowsky 1994). (ii) Restricted Supervised Methods based on bilingual texts, usually aligned; see Brown *et al.* (1991). (iii) Unsupervised Methods rely on raw, *unannotated corpus* and, in few cases, on the content of a machine-readable dictionary. One of the motivations behind the use of these methods is the fact that it is very difficult to find domain-dependent lexical knowledge sources. On the other hand, the major drawback of using unsupervised methods is that no fine-grained distinctions between senses can be made.

3.4. WSD and Swedish

The only known attempt to word-sense disambiguate Swedish on a large scale is a project undertaken at Språkdata (financed by 'The Swedish Council for Research in the Humanities and

Social Sciences' (HSFR)). The project is entitled *Lexikalisk betydelse och användningsbetydelse (SemTag)*, i.e. 'Lexical Sense and Sense in Context'. The sense annotation is carried out interactively through a concordance-based interface, interacting with GLDB, figure 1. All words are sense-tagged. The corpus used in SemTag is the SUC corpus (Ejerhed *et al.* 1992).

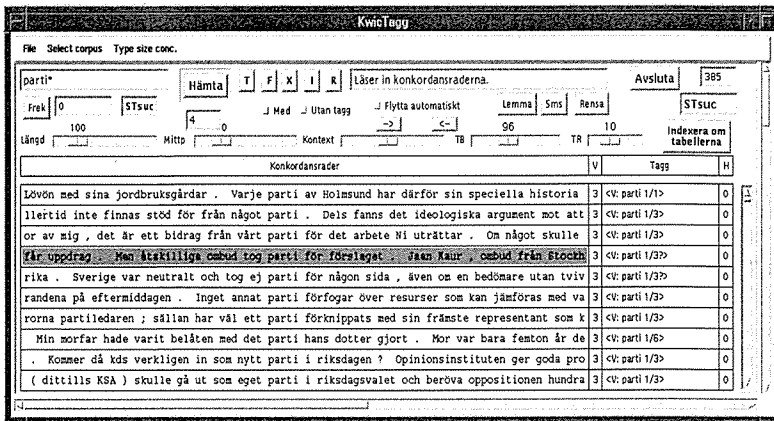


FIGURE 1. *The KwicTagg Interface.*

4. The Usefulness of WSD and Some Potential Applications

The lack of high quality as well as the slow progress within MT has been blamed on word-sense ambiguity. It is wellknown that a single non-ambiguous word in a source language might be translated by a number of different words or expressions in the target language (translational or transfer ambiguity), and a source word can have more than one sense (monolingual ambiguity) (Hutchins & Somers 1992).

In IR it is necessary to disambiguate content words in the queries sent to knowledge bases or free-text search; it is also useful for the purposes of text categorization or indexing, and thus for deciding whether a document is relevant for a particular application or not, by reducing the noise produced due to poly-

semy; *cf.* Schütze & Pedersen (1995). Sense-tagging can improve the performance of Information Extraction (IE). Despite the fact that in IE domain-specific ontologies are already employed, methods for large-scale WSD might improve the IE system's performance even more; *cf.* Kilgarriff (1997a), Chai & Bierman (1997). This can be accomplished by triggering patterns to perform extraction only of relevant senses.

Corpus-based lexicography would benefit from automatic means of identifying the appropriate senses of the words in large corpora for the sake of facilitating and qualitatively improving the information already present in dictionaries. This could be accomplished, by sorting out thousands of concordance lines of irrelevant text with senses not valuable for a specific lexicographic assignment, or by arranging the definitions in the lexicon according to frequency of use, in such a way that the most common senses precede the least common. This is of course a matter dependent on the size and the representativity of the corpus we use, but it is not totally unfeasible.

5. The Critics

The identification of the right meaning of a word, regardless if it is taken from a dictionary or a semantic net is controversial in many respects. Some of the criticism of WSD is concentrated onto three points. First there is the criticism related to the fact that humans cannot agree on which sense is appropriate for the words in a given sentence; *cf.* Kilgarriff (1997b) for a survey. Then there is the fact, that sense distinctions are interpreted differently by different researchers, following different approaches to the disambiguation problem. In this respect, researchers are divided into those who use coarse-grained sense distinctions, or 'lumpers', and those who use fine-grained sense distinctions, or 'splitters'; for an interesting discussion on the matter see Kilgarriff (1997b) and Wilks (1997) published in the same volume. Finally, the fact that different dictionaries differ substantially

regarding the different sets of senses that they associate with the same word complicates the problem even more.

This last claim is also strengthened by the fact that dictionaries tend to be incomplete, both with respect to coverage and content; cf. Boguraev (1995), for a discussion of the use of dictionaries in computational linguistic research. Finally, Wilks (1995) discusses in detail the question whether it is possible to sense-tag on a large-scale and systematically, or not. He examines and attacks two extreme views. According to Wilks, these two views are both misleading claims and are widely believed, though not simultaneously: 'sense-tagging has been solved' or 'it cannot be done at all'. His conclusion is that the field of sense-tagging is still open to further development and that dictionary-based and (un-annotated) corpora-based efforts are equally useful for practical applications.

6. The Chosen Approach

The method chosen here is dictionary-driven and relies on an existing lexical resource for modern Swedish, structured as a relational database, i.e. the *Gothenburg Lexical Database* or GLDB. The content in GLDB has been used for the production of standard contemporary Swedish lexica, for instance the three-volume *Dictionary of the National Encyclopedia*, NEO (1996). The method is based on the *simulated annealing* technique (SA), which has been used for quite some time for sense disambiguation of English words, using a standard machine-readable dictionary, (namely LDOCE).

The idea behind SA is to perform enough exploration of the whole search space early on, so that the final solution is relatively intensive to the starting state. SA is often used to solve problems in which the number of moves from a given state is very large, and it has been applied to the *travelling salesman problem*, in which space is the different paths through the cities that the salesman must visit in an optimal way without visiting the same city twice.

7. Knowledge Sources

7.1. The Structure of GLDB

The work on the GLDB was started 25 years ago by Professor Sture Allén and his research group at Språkdata. The underlying linguistic, theoretical model of GLDB is the *lemma-lexeme* model, Allén (1981). The lemma comprises formal data such as part of speech and inflection(s). The lexemes (or numbered senses) are in turn divided into two categories, a compulsory kernel sense and a non-compulsory set of one or more sub-senses, called the *cycles*. GLDB contains a description of 61,050 lemmas, and 67,785 senses, while 19,082 lemmas contain valency information. GLDB has the advantage of covering the 'whole' language and not just a small subset. A particularly interesting feature of GLDB is the fact that metaphors, though not *dead* ones, are encoded as separate sub-senses of a lemma, usually preceded by the key-word *överfört*, i.e. 'transferred'. A number of printed Swedish monolingual, defining dictionaries have been generated from the GLDB; see Malmgren (1992).

7.2. The Information Used

For enhancing the performance of the sense-tagger we must be able to use as much as possible of the available information in GLDB. The following information seemed an adequate, necessary starting point for the sense-tagging: Definitions and definition extensions; Morphological examples; Syntactic examples; Deverbal Nouns and Valencies. Compared to LDOCE, the GLDB's definitions are much shorter, and the head-word entries usually contain fewer example samples.

8. Data Preparation

For enhancing the lexical disambiguation result using the available resources, it is necessary to perform pre-processing both in the resources and the text to be sense-tagged. This is motivated by the fact that by making certain normalizations and simplifications in the resources, such as lemmatization, we contribute to the production of qualitatively better results.

There have been several reasons that motivated the use of pre-processing. Some of these have been: (i) the fact that not all entries in the GLDB are relevant during the sense annotation of a "normal-length" newspaper text. This means that we extract only a subset of the GLDB depending on the unique occurrences of word forms in the text to be processed; and (ii) not all entries in GLDB consist of a single entry form (lemma). This is the case with 914 phrasal verbs consisting of two or three units, 115 multi-word nouns, and 5 multi-word adjectives. Moreover, it is absolutely necessary to reduce the complexity of the matching process by operating onto base forms by conducting lemmatization, both in the text and the information in the lexical entries, especially the GLDB definitions. Using base forms reduces the complexity and time required for the calculation of the overlap between the resources. Part-of-speech tagging is also an important aspect, since it eliminates *accidental* homography (*länkar* 'links, chains, guides' as verb, *länkar* 'links, chains' as noun). Another aspect in favour of the pre-processing required has to do with the productivity of the Swedish language in creating new compound words, especially nouns. New content words, not present in the lexicon, must be identified and possibly assigned a sense, based on entries with similar defining criteria, i.e. in this study, by using the definition of the last part of the compound.

8.1. Multi-Word Units (in Text and GLDB)

Multi-word expressions cannot be properly understood if they are not recognized as unique units. There are a number of diffe-

rent types of units recognized: *phrasal verbs, idioms, lexical and grammatical collocations.*

There are 914 phrasal verbs explicitly given a separate entry in GLDB, such as *brinna av:1/1* 'to go off, to explode', *dela ut:1/1* 'to distribute' and *ställa ut:1/1,1/2* 'to exhibit'; in 223 of these, the Swedish third-person reflexive pronoun *sig* 'himself/herself/themselves' is the last part of the unit; i.e. *bekanta sig:1/1* 'to acquaint oneself', *dra på sig:1/1* 'to put (pull) on' and *ställa in sig:1/1* 'to intend to'.

There are two issues which need special attention with respect to phrasal verbs in Swedish. One is that they can be discontinuous in the text, and the other that in some cases it is only the intonation and/or extended context that can decide whether a verb is a phrasal or not. Consider for instance the verb *köra* 'to drive, to run, to force, to convey' combined with the token *på* 'on/at/into', which in the first example below is a particle, while in the second a preposition: (i) *marknader kör på som om inget hänt* 'markets keep on as if nothing has happened'; (ii) *att köra bil på nätter*, 'to drive a car at night' .

8.2. Known Compounds, Morphological Examples (in GLDB)

The morphological examples in GLDB are simply compound nouns in which the lemma-entry participates as the main information carrier of the compound. For example, the first sub-sense of the second sense of the noun entry *avgång* 'wastage, retirement, resignation' contains two morphological examples (compounds) :

avgång 1/2/a: **avgångsbetyg** 'leaving certificate', **avgångsklass** 'final class'

All the compounds have been automatically split into their respective parts, by identifying the lemma or part of it in the

compound. The morphological examples of the noun *avgång* are actually used by the sense-tagger as:

avgång 1/2/a: avgångsbetyg, betyg, avgångsklass, klass

After the automatic segmentation, the split compounds were automatically post-validated for erroneous splitting, or for completing the produced segments. We anticipate that a small number of segmentation errors might be present in the morphological examples.

8.3. Unknown Compounds (in Text)

For the cases where no entries in GLDB cover these compounds, the application of heuristic compound segmentation is performed.

Previous attempts to segment compounds without the help of a lexicon are described in Brodda (1979), and Klenk & Langer (1989). The segmentation algorithm we use proceeds by scanning unidentified word forms from left to right, trying to identify grapheme combinations which feel *unnatural* or simply unallowed as non-compound forms in the Swedish language, and which carry information of potential token boundaries. The heuristic method behind the segmentation of compounds in our method is based on producing 3-gram and 4-gram character sequences of several hundreds of non-compound lemmas, and then generating 3-grams and 4-grams that are not part of the lists produced, some manual adjustments being also imposed. Furthermore, 4-grams with 4, 3 or 2 vowels, and 3-grams with 3 or 2 vowels were not used, except in the cases with two similar consecutive vowels, such as *ii* and *ee*. Ambiguities are unavoidable, although the heuristic segmentation has been evaluated for precision, and over 90% accuracy was measured, a more thorough evaluation is beyond the scope of this report, and that is why we concentrated on precision alone, which is easier to estimate than coverage.

Consider for instance the following examples:

TABLE 1. *Splitting of Unnatural Grapheme Combinations.*

<i>"Unnatural" Grapheme Combinations</i>	<i>Splitting Point i.e. ('!')</i>	<i>Examples</i>
<i>ivb</i>	<i>iv b</i>	<i>skriv bord</i> (writing desk), <i>kollektiv boende</i> (collective housing)
<i>ktm</i>	<i>kt m</i>	<i>kontak tman</i> (contact person), <i>makt mmissbruk</i> (power abuse)
<i>ksf</i>	<i>ks f</i>	<i>olycks fall</i> (accident), <i>Danmarks färjan</i> (Denmark ferry)
<i>tss</i>	<i>ts s</i>	<i>rätts salen</i> (court room), <i>arbetslöshets siffrorna</i> (unemployment figures)
<i>gsk</i>	<i>gs k</i>	<i>växlings kontor</i> (exchange office), <i>tillverknings kapacitet</i> (manufacturing capacity)
<i>ngss</i>	<i>ngs s</i>	<i>forsknings skola</i> (research school), <i>bantnings studie</i> (slimming study)

8.4. Idiomatic Expressions (in Text and GLDB)

Idiomatic expressions are recognized and are not treated for WSD. A list of over 4,900 idiomatic expressions has been extracted from GLDB and implemented as a finite-state recognition machine, used during the pre-processing stage. The original list of all the idioms in GLDB has been expanded to about 6,500, since we had to cope with the expansion of parenthetical and shorthand information for different variations of an idiom. For instance:

GLDB: *ben: (han är) bara skinn och &*, i.e. 'he is) nothing but skin and bone', has been encoded as:

(i') *han är bara skinn och ben*

(i'') *skinn och ben*

8.5. Deverbal Nouns (in Text and GLDB)

Nominalized verbs are yet another problematic set of items that have to be processed in a specific manner, since these are not treated as separate lemmas in the database, but are (usually) encoded under the verb entry. In Swedish, deverbal nouns are usually constructed by means of the morphemes: *~(n)ing*, *~ande*, *~ende* and *~nde*. Some of these nouns are very productive, while some are only theoretically possible or less frequent. The method we chose to deal with these cases is first to identify them in the text, and then mark them accordingly, so that they will be analyzed during the WSD procedure, depending on the corresponding verbal entries. Notice, however, that *there are* separate lemmas in GLDB ending in: *~(n)ing*, *~ande*, *~ende* and *~nde*, that originate from verbs. The criterion for having such entries in GLDB is based on the fact that these nouns have very specific, and concrete meanings, (e.g. *kaffeservering:1/1* 'café, coffee-room', *pristävling:1/1* 'prize competition'), and no longer denote a verbal action of some kind.

8.6. Part-of-Speech Tagging and Lemmatization (in Text and GLDB)

The application of part-of-speech tagging is carried out for sorting out non-relevant definitions of homograph tokens during the extraction of a relevant subset of the GLDB with respect to the text that is analyzed, and for making easier the lemmatization, which is applied to tagged tokens. Brill's (1994) rule-based tagger, is used for part-of-speech tagging, trained on Swedish texts; Johansson Kokkinakis & Kokkinakis (1996).

From a computational and performance point of view, it is attractive and desirable to operate on the roots (or stems) of words, both in the text and the lexical resources. For that reason we have lemmatized the definitions in GLDB and we use the lemmatizer prior to processing a text by the sense-tagger. The

lemmatizer is applied to part-of-speech annotated texts, which enhances the quality of the stemmed results.

9. Computer Processing and GLDB – some Problems

The experience gained through working with GLDB and the sense-tagging task has proved that GLDB *is* an adequate resource for the WSD process, although the evaluation needs to be extended over a larger sample of the language than the one we have used so far (see next section). Nevertheless, there were a few occasions where the structure of GLDB lacked consistency. Of course, we do not disregard the fact that the GLDB's organizational structure is made *by* humans *for* humans and not for any particular computer processing, and that explicit encoding of all occasional word forms (especially compound nouns), or phrasal verbs would have led to an unmanageable explosion of the entries in GLDB. A final point regards the way the valencies and the deverbal nouns are described within the different sub-senses (cycles) in the database. The problem associated with this issue is that this information is *not* explicitly denoted for the individual cycles. The valency and deverbal information descriptions are given only for the lexemes, and implicitly for all the different sub-senses of a sense. This encoding methodology makes the accurate identification of the sub-senses much more difficult.

10. Evaluation

For the evaluation part of this study we have manually sense-tagged different text samples. The evaluation is performed after the texts have been tokenized, the idioms, deverbal nouns and multi-word expressions identified, then part-of-speech tagged, lemmatized, and content words identified and marked appropriately. The manual annotation has proven to be a very labour-intensive but challenging and necessary process. Two experiments were conducted. In the first case, we randomly extracted

short newspaper samples and sense-tagged the verbs, nouns and adjectives. The evaluation was carried out in two different ways: (i) WSD *only* by using the definition and definition extensions of the lexical entries in the GLDB; and (ii) WSD by using the definitions supplemented with definition extensions, morphological and syntactic examples, and even the typical prepositions, (valencies), for each entry. In the second case-study, we extracted 60 concordance lines in which a single ambiguous verb, the verb *handla* 'to deal, to trade, to take action, to buy', was the object of the investigation, and was then sense-tagged.

10.1. Results

Table (2) shows the results of the evaluation, using only the definitions and definition extensions, and the definitions and definition-extensions extended with other knowledge, such as the morphological and syntactic examples (All Material). Here the metric *precision* is defined as the percentage of the sense-tagged words that were found correct: (relevant hits/all hits). The manual annotation was far from straightforward.

TABLE 2. *Sense-Tagging Evaluation Results (Case Study 1).*

<i>Part-of-Speech</i>	<i>Occ.</i>	<i>Definitions & Extensions</i>	<i>All Material</i>
Adjectives (20), Nouns (88), Verbs (44) (32 sentences)	152	<i>Precision</i> 37,5%	<i>Precision</i> 84,21%

As shown in table 2 the performance of the WSD with the use of only the definitions is, as one might have expected, very much lower than when considering the second case in which the definitions have been supplemented with a lot of other information. The qualitative improvement between the two cases is very high. If the definitions in GLDB had been richer as to information, the performance might have improved even more. Note, however, that Wilks & Stevenson (1998) observed that by using

simulated annealing, the longer definitions in LDOCE tended to win over the shorter ones, since the length of the definitions varied considerably between different entries and thus influenced the software towards an erroneous solution. Accordingly, they elaborated a method to penalize the longer definitions, something that we did not consider in our study.

Table 3 shows the evaluation figures for a single entry, the verb *handla*, at the sense level, using example sentences randomly extracted from various sources.

TABLE 3. *Sense-Tagging Evaluation Results (Case Study 2).*

<i>Case Study (2)</i>	<i>Occ.</i>	<i>Definitions & Extensions</i>	<i>All Material</i>
		<i>Precision</i>	<i>Precision</i>
< <i>handla</i> >	60*	41,37%	82,75%

(*Since two of the occurrences were phrasal verbs not covered by GLDB, the evaluation was calculated on 58 examples.)

Large annotated data for single ambiguous words would probably reveal interesting groups of patterns and clear differences between such groups and entries, this is left for future research.

11. Conclusions

We have ported a large-scale sense-tagger, originally developed for the sense-tagging of English, to Swedish. We used one of the most comprehensive lexical resource available, the *Gothenburg Lexical Database* and tested the performance on open-source (newspaper) texts. The methodology behind the approach implemented in the sense-tagger follows the simulated annealing technique, used recently for the sense-tagging of English texts with the *Longman Dictionary of Contemporary English* (LDOCE). The combination of simulated annealing with a machine-readable dictionary has outperformed all other approaches based on dictionaries, achieving very high accuracy on the sense-tagging of *all* the content words in a text and not only a very small set, as is usually the case described in WSD-related literature.

Evaluation of WSD for Swedish gave good evidence that the task is feasible and the precision figures were very encouraging. The tagger will be used in the context of a larger architecture, for the acquisition of lexical semantic knowledge from open-source texts. The sense-tagger will contribute very important information for the precise assignment of lexical semantic knowledge to polysemous and homonymous content words. In this respect, we regard sense-tagging as a very important process and component when it is seen in the context of a wider and deeper text-processing architecture. Of equal importance is the way that the results returned by the sense-tagger can be used for the resolution of the preposition phrase attachment problem. GLDB contains information (typical prepositions) associated with the different senses of verbs, nouns and even a large number of adjectives. Once the right sense for a token is identified, the information found in the valency slot can more efficiently guide an algorithm to make the right decision as to whether a prepositional phrase functions as an argument or adjunct.

This issue will be investigated in the near future, as well as the evaluation of the sense-tagger on a larger scale.

Acknowledgements

This research was *partially* supported by the EU Language Engineering programme LE1-2238/AVENTINUS. We would like to thank Yorick Wilks and Mark Stevenson (University of Sheffield) for giving us the opportunity to test the optimized version of the simulated annealing software with Swedish resources. We are also grateful to Jerker Järborg for manually annotating the samples tested and for comments on a previous draft, as well as to Sven-Göran Malmgren for enabling the use of the GLDB.

References

- NEO 1996 = *Nationalencyklopedins ordbok*, volumes 1–3, Språkdata & Bra Böcker AB.
- Roget's Thesaurus = http://humanities.uchicago.edu/forms_unrest/ROGET.html
- Allén, S. 1981. "The Lemma-Lexeme Model of the Swedish Lexical Database." In: Rieger, B. (ed.): *Empirical Semantics*. Bochum. Pp. 376–387. (Reprinted in: Allén, S. 1999. *Modersmålet i fäderneslandet. Ett urval uppsatser under fyrtio år av Sture Allén. (Meijerbergs Arkiv för svensk ordforskning. 25.)* Pp. 268–278.)
- Boguraev, B. 1995. "Machine-Readable Dictionaries and Computational Linguistic Research." In: Walker D., Zampolli A. and Calzolari N. (eds.): *Automating the Lexicon. Research and Practice in a Multilingual Environment*. Oxford University Press. Pp. 301–336.
- Brill, E. 1994. "Some Advances In Rule-Based Part of Speech Tagging." In: *Proceedings of the 12th AAI '94*, Seattle Wa.
- Brodda, B. 1979. "Något om de svenska ordens fonotax och morfotax: iakttagelse med utgångspunkt från experiment med automatisk morfologisk analys." (PILUS 38). Institutionen för lingvistik, Stockholms universitet.
- Brown, P.F., et al. 1991. "Word-Sense Disambiguation Using Statistical Methods." In: *Proceedings 29th ACL*. Berkeley, California. Pp. 264–270.
- Chai, J.Y. and Bierman, A.W. 1997. "The Use of Lexical Semantics in Information Extraction." In: Vossen, P. et al. (eds.): *Proceedings of Automatic Information Extraction and Building of Lexical Semantic Resources Workshop*. Madrid, Spain
- Cowie, J., Guthrie, J. and Guthrie, L. 1992. "Lexical Disambiguation Using Simulated Annealing." In: *Proceedings of 15th COLING*, Vol. 1. Nantes. Pp. 359–365.
- Fr. Dolan, W.B. 1994: "Word Sense Ambiguity: Clustering Related Senses." In: *Proceedings of the 15th COLING*, Vol. II. Kyoto, Japan. Pp. 712–716.

- Ejerhed, E. *et al.* 1992. "The Linguistic Annotation of the Stockholm-Umeå Corpus project." Technical Report No. 33, Univ. of Umeå.
- Fellbaum, C. (ed.) 1998. *WordNet, an Electronic Lexical Database*. MIT Press.
- Fujii, A. 1998. *Corpus-Based Word Sense Disambiguation*, PhD thesis, Tokyo Inst. of Computer Science, Japan.
- Hutchins, W.J. and Somers, H.L. 1992. *Introduction to Machine Translation*. Academic Press.
- Johansson-Kokkinakis, S. and Kokkinakis, D. 1996. "Rule-Based Tagging in Språkbanken." Research Reports from the Department of Swedish, Göteborg University, GU-ISS-96-5.
- Kelly E., and Stone, P. 1975. *Computer Recognition of English Word Senses*. North-Holland Linguistic Series.
- Kilgarriff, A. 1997a. "Foreground and Background Lexicons and Word Sense Disambiguation for Information Extraction." In: *Proceedings of the Lexicon Driven Information Extraction Workshop*, Frascati, Italy.
- Kilgarriff, A. 1997b. "I Don't Believe in Word Senses." In: *Computers and the Humanities*, Vol. 31:2.
- Klenk, U. and Langer, H. 1989. "Morphological Segmentation Without a Lexicon." In: *Literary and Linguistic Computing*, Vol. 4:4. Oxford University Press. Pp. 247-253.
- Krovetz, R. 1997: "Homonymy and Polysemy in Information Retrieval." In: *Proceedings of the joined 35th ACL and 8th EACL*. Madrid, Spain. Pp. 72-79.
- Lesk, M.E. 1986. "Automatic Sense Disambiguation Using Machine Readable Dictionaries: How to Tell a Pine Cone from an Ice-cream Cone." In: *Proceedings of the ACM SIGDOC Conference*. Toronto, Ca. Pp. 24-26.
- Malmgren, S.G. 1992. "From Svensk ordbok ('A dictionary of Swedish') to Nationalencyklopediens ordbok ('The Dictionary of the National Encyclopedia')." In: Tommola H. *et al.* (eds.): *Proceedings of the EURALEX '92*. Vol. 2. Pp. 485-491.

- Miller, G.A. (ed.) 1990. "WordNet: An on-line Lexical Database." In: *International Journal of Lexicography*, 3(4), Special Issue.
- Miller, G.A., Leacock, C., Teng, R. and Bunker, R.T. 1993. "A Semantic Concordance." In: *ARPA Human Language Technology Workshop*. New Jersey, USA. Pp. 303–308.
- Nag, H. T. 1997. "Exemplar-Based Word Sense Disambiguation: Some recent Improvements." In: Cardie, C. & Weischedel, R. (eds.): *Proceedings of the 2nd Conference on Empirical Methods in NLP*. Rhode Isl., USA. Pp. 208–213.
- Peters, W., Peters, I. and Vossen, P. 1998. "Automatic Sense Clustering in EuroWordNet." In: *Proceedings of the LREC*, Vol. 1. Granada, Spain. Pp. 409–416.
- Riggs, F. 1993. "Social Science Terminology: Basic Problems and Proposed Solutions." In: Sonneveld, H. and Loening, K. (eds.): *Terminology, Applications in Interdisciplinary Communication*. J. Benjamins Publ. Co. Pp. 195–222.
- Sanfilippo, A. (chair) 1998. "EAGLES: Preliminary Recommendations on Semantic Encoding." Interim Report, The EAGLES Lexicon Interest Group. (<http://www.ilc.pi.cnr.it/EAGLES96/>, site visited 10 Feb. 1999)
- Schütze, H. and Pedersen, J.O. 1995. "Information Retrieval Based on Word Senses." In: *Proceedings of the 4th Annual Symposium on document Analysis and Information Retrieval*. Las Vegas, NV. Pp. 161–175.
- Wilks, Y. 1995. "Texts and Senses." Memoranda in Computer and Cognitive Science, CS-95-23. Univ. of Sheffield.
- Wilks, Y. 1997. "Senses and Texts." In: *Computers and the Humanities*, Vol. 31:2.
- Wilks, Y. and Stevenson, M. 1998. "Word Sense Disambiguation Using Optimised Combinations of Knowledge Sources." In: *Proceedings of the COLING/ACL '98*. Montréal, Canada.
- Wilson, A. and Thomas, J. 1997. "Semantic Annotation." In: Garside, R. et al. (eds.): *Corpus Annotation. Linguistic Annotation from Computer Text Corpora*. Longman. Pp. 54–65.

Yarowsky, D. 1994. "A Comparison of Corpus-Based Techniques for Restoring Accents in Spanish and French Text." In: *Proceedings of the 2nd Workshop on Very Large Corpora*, Kyoto, Japan. Pp. 19–32.

Rörande polysem belägenhet

In my work with Svensk Konstruktionsordbok, the verbs *sitta* ("sit"), *stå* ("stand") and *ligga* ("lie") have shown to be problematic to handle because of the many ways they are used in the many phrases they appear in. In this paper, where I focus on the verb *ligga* ("lie"), I discuss the possibility of using the term *polysemy* only in two strict senses; extended meanings and metaphorically transferred meanings. By rather strict constraints (function and productiveness) of these types of meaning variation, I try to achieve a systematic and efficient tool for structuring the descriptions of these verbs. In the end of this discussion I suggest a polysemic model for these verbs, where the extended meanings form a base from which metaphorically transferred meanings may emerge.

1. Inledning

Polysemi är ett vitt begrepp som ofta används för olika sorters flertydighet på olika språkliga nivåer. I det följande resonerar jag kring hur man skulle kunna avgränsa och systematisera begreppet med särskilt avseende på utvidgade respektive överförda betydelser av de s.k. belägenhetsverben, *sitta*, *stå* och *ligga*. Detta som ett teoretiskt försök att strukturera beskrivningarna av dessa verb i Svensk Konstruktionsordbok (SKO). I den här framställningen koncentrerar jag diskussionen kring verbet *ligga*.

2. Bakgrund

I arbetet med SKO har vi till uppgift att presentera och beskriva de konstruktioner och fraser uppslagsorden ingår i. Med detta menas såväl beskrivning av ordens argumentstrukturer¹ och öppna kollokationer som lexikaliserade fraser, d.v.s. slutna kollo-

¹ I lexikografiska sammanhang talas ibland om valens, men i enlighet med alla hittillsvarande beskrivningar av SKO använder jag termen argumentstruktur.

kationer, idiom och pragmatiska fraser (jfr den principiella modell Ulla Clausén presenterar i sitt bidrag i denna volym.) Det är alltså inte vår ambition att definiera varje uppslagsords särskilda betydelse på ordnivå. Grundtanken är i stället att vi förklarar betydelsena hos de uttryck uppslagsorden ingår i. Det gör vi i anslutning till de konstruktioner och fraser som beskrivs med betydelspecificationer vid de språkliga exemplen. Däremot om ett uppslagsord är homonymt eller polysemt definierar vi betydelsen även på ordnivå.

När det gäller det vi kan kalla belägenhetsverben, d.v.s. *sitta*, *stå* och *ligga*, handlar det om väldiga mängder information som ska presenteras inom ramen för varje artikel. I fråga om verbet *ligga* rörde det sig vid en genomräkning om omkring 160 fraser. Då är alla uttryck med partikelverb, reflexiva partikelverb och andra reflexiva konstruktioner borträknade, eftersom alla sådana presenteras separat, om än som biuppslagsord till huvudverben. Antalet varierar naturligtvis lite beroende på vad man räknar som ett eget uttryck och inte, men det är hur som helst fråga om mycket material att strukturera och sortera så att man får en överskådlig och lättillgänglig presentation. Ett av de verktyg som står till buds är uppspjälkningen av uppslagsord i s.k. betydelsemoment, eller, om man så vill, polysema varianter (för enkelhets skull kallar jag dem hädanefter polysemer).

2.1. Tidigare lösningar

SKO är tänkt att ersätta *Svensk handordbok* (SHO), och även om SKO kommer att se väldigt annorlunda ut, delas ambitionen att så långt det är möjligt presentera alla de mer och mindre fasta uttryck uppslagsorden förekommer i. Därför finns det all anledning titta på de lösningar som presenteras just där. Även andra ordböcker är generösa med exempel på mer eller mindre fasta uttryck, t.ex. *Ord för ord* (OfO) och *Svensk ordbok/Nationalencyklopedins ordbok* (SOB/NEO). Jag har också tittat på de mindre detaljerade beskrivningarna i *Bonniers svenska ordbok* (BSO) och LEXIN.

De betydelsemoment de olika ordböckerna föreslår varierar väldigt (se appendix). I SHO har man delat in t.ex. *ligga* i 41 (!) numrerade betydelsemoment sorterade i 9 grupper. (Partikelverb och reflexiva konstruktioner behandlas för sig och är alltså inte medräknade.) I OfO delas *ligga* in i 7 betydelsemoment medan SOB/NEO delar in *ligga* i 6 betydelsemoment, i de senare anger man också en del s.k. betydelsenyanser. BSO gör inga numrerade indelningar, men skiljer hos *ligga* ändå ut två betydelser och en ”friare” användning. LEXIN, slutligen, delar inte upp *ligga* i några moment alls.

Om man så tittar närmare kan man konstatera att alla ordböcker har betydelsen ’befinna sig i vågrät ställning’ med en eller annan formulering först. Hos SOB/NEO kommer ’vara belägen’ i det andra betydelsemomentet, liksom hos OfO, men i SHO kommer den först i grupp 4 och har även en motsvarighet i grupp 5. I LEXIN nämns inte denna betydelse alls och i BSO tas den med som den andra av tre (onummerade) betydelser. Övriga betydelsemoment stämmer så illa överens att man inte kan redogöra för dem med mindre än att man räknar upp allihop.

Man kan alltså slå fast att det inte finns någon vidare enighet i hur många eller vilka betydelsemoment man bestämmer sig för. Och om man tittar ännu lite noggrannare ser man att bland de moment som ändå någorlunda överensstämmer mellan böckerna, rankas de inte lika.

2.2. Problem

Att betydelsemomenten har givits så olika innehåll, extensioner och rang, tror jag beror dels på att böckerna har olika syften och dels på att de är skrivna vid olika tidpunkter under närmare fyrtio års tid. SHO, som gavs ut för mer än trettio år sedan, kan väl sägas tjäna som ett slags ”skräckexempel” med de 41 betydelserna för verbet *ligga*. Denna överväldigande mängd definitioner är ett problem i sig med avseende på artikelns genomträngbarhet. Beskrivningen blir oerhört detaljerad och därmed svår att hantera, inte minst om man tänker sig numreringen som

en sorts hierarkisk struktur (jfr exv. Svensén 1987 och Malmgren 1992); hur avgör man t.ex. att *ligga* 'om ljussken, mörker o.d.' (jfr SHO:IV-3 i appendix) är överordnat *ligga* 'om vinden' (jfr SHO:IV-7 i appendix)? Dessutom definieras betydelsemomenten ibland på olika beskrivningsnivåer (och i enstaka fall inte alls).

En viktig orsak till att man i SHO får så här många och många olika betydelser att hantera är att man definierar verbets betydelse utifrån argumentens, och ibland hela uttryckens, lexikala betydelse¹. Exempelen med ljusskenet, mörkret och vinden ur SHO ovan är särskilt tydliga, men samma analysätt kan anas i t. ex. SOB/NEO, där man bitvis har en sorts konceptuell gruppering av fraser i betydelsemoment (se SOB/NEO 4-6 i appendix). Man kan ändå konstatera att de modernare ordböckerna (BSO, LEXIN och SOB/NEO) presenterar betydligt färre betydelsemoment och artiklarna i dessa böcker är något mer välstrukturerade och därmed lätta att navigera i. Men uppenbarligen är inte heller de alls överens om betydelsemomentens innehåll, extension eller rang. Indelningarna förefaller i stället, åtminstone vid en sådan här jämförelse, vara subjektiva (därmed inte sagt ogrundade) bedömningar av vilka uttryck som är de mest centrala i användningen av verben.

Även om indelningen av betydelsemoment fungerar ganska väl i sina sammanhang (i alla fall i de modernare ordböckerna), undrar jag om det inte måste till en annan sorts systematisk metod att isolera de polysemer jag har att definiera bland t.ex. de 160 fraserna i artikeln *ligga* i arbetet med SKO.

3. En annan väg

I beskrivningar av polysemi skiljer man ibland på utvidgade, bildliga, överförda och andra speciella betydelser (jfr t.ex. Malmgren 1992). Bildliga och andra (enstaka eller s.a.s. tillfälliga)

¹ Ytterligare en möjlig orsak är att man blandar metaforiska användningar med konkreta betydelser, men jag ska inte fördjupa mig i det här.

speciella användningar av ord presenterar vi i samband med beskrivningen av de vanliga användningarna av ordets konstruktioner och fraser (jfr avsnitt 2. ovan). Den sortens betydelser ligger alltså inte till grund för indelningen av betydelsmoment, eller polysemer, i SKO. Med utvidgade och överförda betydelser förhåller det sig annorlunda. Utvidgad betydelse kan man tala om när betydelsen av ett ord används i mer och mer vid, eller allmän, bemärkelse. Man kan se det som att den grundläggande betydelsen tunnas ut och blir mer allmän, men det är fortfarande just den grundläggande betydelsen man laborerar med. Överförd betydelse kan man tala om när ett ord "adopteras" till en helt annan innebörd via en eller flera lexikaliserade metaforiska fraser. Den nya betydelse av ordet som därigenom uppstår har alltså inte lika mycket med den ursprungliga betydelsen att göra.

Dessa begrepp används ibland på lite olika sätt (jfr t. ex Malmgren 1994), men om vi just nu bestämmer oss för ovanstående tolkningar, ger de en användbar möjlighet till åtskillnad mellan olika typer av polysemer; om de uppstår på olika sätt kan man förvänta sig att de har olika sorters egenskaper och därför kan isoleras på olika sätt.

3.1. Utvidgade betydelser

För att komma åt ett verbs (eventuella) utvidgade betydelser verkar det rimligt att utgå från dess grundbetydelse och titta på hur den modifieras. Verbet *ligga:s* mest grundläggande betydelse är ungefär 'befinna sig i vågrätt läge', det är förmodligen de flesta överens om (även de ordböcker som diskuterats tidigare i denna framställning). Det handlar alltså om att en kropp har en viss fysisk position.

Hur avgör man då hur denna grundbetydelse modifieras? Att luta sig mot argumentens eller hela frasernas lexikala betydelser, så som i SHO, leder till en osmidig beskrivning; mängden möjliga lexikala betydelser hos argument och fraser är i princip större än själva antalet fraser vi ska hantera. Ett annat sätt vore att generalisera betydelserna hos argumenten genom att ta fasta på

att olika sorters betydelser har olika funktion i förhållande till själva verbet. Vi kan, kanske enklare, försöka se det från andra hållet: Ett verb har en viss betydelse, och kombineras det med något (ett argument) blir den betydelsen modifierad, och betydelsen modifieras på olika sätt beroende på vad argumentet har för funktion.

Jag vill pröva hur långt resonemanget leder med begreppet funktion i bemärkelsen satsdelsfunktion. I det material jag samlat in¹ består *ligga*-fraserna av antingen V, V+adj, V+adv, V+PP, V+NP eller V+NP+PP. (För att rationalisera beskrivningen här skriver jag inte ut subjekten.) Prepositionsfraserna och adverbena fungerar som adverbial och det är typiskt fråga om rums- eller sättsadverbial. Adjektiven fungerar antingen som subjektiva predikativ och nominalerna fungerar antingen som subjektiva predikativ eller som objekt.

LIGGA		
FRASENS FORM	ARGUMENTETS FUNKTION	EXEMPEL
V	-	<i>ligga</i>
V +adv	sättsadvl	<i>ligga skavfötters</i>
V +PP	sättsadvl	<i>ligga på mage</i>
V +PP	rumsadvl	<i>ligga vid torget</i>
V +adv	sättsadvl	<i>ligga sist</i>
V +PP	sättsadvl	<i>ligga i skilmässa</i>
V +NP +PP	obj +sättsadvl	<i>ligga ngn till last</i>
V +adj	subj. predikativ	<i>ligga död</i>
V +NP	subj. predikativ	<i>ligga lik</i>
V +NP +PP	subj. predikativ +sättsadvl	<i>ligga granne med ngn</i>
V +NP +PP (<i>det</i> -konstr.)	eg. subj. +rumsadvl	<i>det ligger sanning i ngt</i>

Bland de adverbiala argumenten återfinns rums- och sättsadverbial. De syftar på olika sätt på liggandet som sådant, men låter sig i stort beskrivas i termer av belägenhet. I de uttryck vilkas argument som fungerar som predikativ är verbets roll väldigt

¹ Huvudsakligen från Språkdata och Presstext.

underordnad och kan närmast sägas ha en kopulativ roll (jfr Ekberg 1989:22f.). Detta gäller också en sorts fasta uttryck med formellt subjekt. Den skarpsynte noterar att alla kombinationer av form och funktion i uppställningen här ovanför är unika utom adv-sättsadvl: *ligga skavfötters*, *ligga sist* och PP-sättsadvl: *ligga på mage*, *ligga i skilsmässa*. Detta visar att det satsdelsfunktionella kriteriet inte riktigt räcker till.

För att skilja de formellt och satsdelsfunktionellt överlappande frasernas betydelser åt, måste vi ta ytterligare en dimension i beaktande, nämligen grad av kroppslighet. Denna dimension har sitt ursprung i verbets grundläggande betydelse, 'befinna sig i vågrätt läge', och skulle kunna ses som en annan nivå av begreppet funktion. De tre första fraserna i uppställningen handlar alla om den liggande kroppen som sådan, den fjärde om var, i fysisk bemärkelse, kroppen befinner sig. De tre nästa beskriver en omständighet eller situation som kroppen befinner sig i och de fyra sista handlar om hur kroppen är, utan egentligt särskilt avseende på liggandet i sig (den lite avvikande *det*-konstruktionen diskuteras nedan). Därmed skiljer sig *ligga skavfötters* och *ligga på mage* från *ligga sist* och *ligga i skilsmässa* i det att *skavfötters* och *på mage* beskriver den fysiska kroppens position medan *sist* och *i skilsmässa* handlar om den omständighet eller situation kroppen befinner sig i.

Utifrån ovanstående kan vi gruppera uttrycken så att vi isolerar fyra betydelsemoment:

1) POSITIONELL BELÄGENHET. Den första gruppen utgörs av konstruktioner och fraser där verbet *ligga* används i sin grundläggande betydelse, d.v.s. 'befinna sig i vågrätt läge'. Att det är den mest grundläggande betydelsen det är fråga om stöds formellt av att det är den betydelse verbet har utan ackompanjering av några betydelsemodifierande argument. I den här gruppen hamnar dessutom de fraser som har argument vilkas funktion kan sägas "kommentera" verbet utan att modifiera betydelsen. Med avseende på verbets betydelse kan vi därför kalla dem optionella, även om de inte är det i sina respektive fraser (eftersom de då inte vore dessa fraser; en självklarhet som kanske ändå

måste lyftas fram i sammanhanget). Konstruktionssättet kan sammanfattas som V[+PP/adv] och funktionen hos de optionella argumenten är alltid sättsadverbial.

2) SPATIAL BELÄGENHET. Den andra gruppen består av fraser som handlar om var någonstans ngn/ngt är beläget med avseende på fysisk omgivning; de svarar alltid på frågan "var?". Argumentet är helt obligatoriskt, utan det förändras verbets betydelse. De består alltid av V+PP och funktionen hos det obligatoriska argumentet är alltid rumsadverbial.

3) SITUATIONELL BELÄGENHET. Den tredje gruppen samlar de fraser som beskriver en situation eller omständighet som ngn/ngt befinner sig i. De svarar i första hand på frågan "hur?", men många gånger är uttrycken förvillande lika de som handlar om spatial belägenhet, bl. a. genom att de ofta konstrueras med PP, men även på andra sätt t.ex. frasen *ligga sist*. Det som får avgöra sådana fall är om belägenheten kan knytas till en fysiskt bestämd plats i omgivningen, eller om det snarare handlar om en plats som bestäms i relation till någon eller något annat. Här finns också en konflikt i att t.ex. många metaforer i sin bokstavliga betydelse beskriver en spatial belägenhet, medan deras metaforiska användning handlar om situationell belägenhet. Den övergripande principen i arbetet med SKO är att metaforer placeras i det betydelsemoment som beskriver deras metaforiska användning (utom då även den bokstavliga betydelsen kan räknas som kollokation; i sådana fall har den bokstavliga betydelsen högre prioritet för frasens placering). De konstruktioner som förekommer är V+adv/[NP]PP och här finns en obligatorisk sättsadverbial funktion och ett optionellt objekt (obligatorisk respektive optionell i de bemärkelser som nämndes i beskrivningen av positionell belägenhet ovan).

4) ALLMÄNT VARANDE. Den fjärde gruppen skiljer sig från de övriga i det att den inte kan sägas beteckna belägenhet. Här återfinns uttryck där argumenten beskriver subjektet. Verbet i dessa uttryck är nästan tomt på lexikalt innehåll och kan snarast sägas

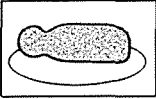
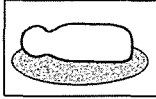


vara kopulativt; de uttrycker ett slags allmänt varande som i t.ex. *ligga öde*; *ligga* kan t.o.m. ersättas med verbet *vara* utan att betydelsen förändras eller utarmas nämnvärt. Ett annat fenomen som kanske stöder analysen om det allmänna varandet är att fraserna i denna grupp i flera fall förekommer i parallella uttryck mellan de olika belägenhetsverben, man kan t.ex. både *ligga lik* och *stå lik*, man kan både *stå modell* och *sitta modell* och det spelar för det mesta förmodligen ingen större roll vilket av verben man väljer (?). (Däremot kan man knappast **sitta lik* eller **ligga modell*, så visst finns ett uns av lexikalt innehåll hos verben även i dessa uttryck.)

Till den här fjärde gruppen har jag också räknat en sorts fast *ligga*-frastyp som kräver formellt subjekt (det +V +NP +PP), t.ex. *det ligger sanning i ngt*. Detta därför att NP:n i dessa fraser är egentligt subjekt, och alltså kan sägas vara ekvivalent med frasens subjekt på ett sätt som liknar de subjektiva predikativen. Man skulle kunna argumentera mot detta genom att hävda att man kan parafrasera *det ligger sanning i ngt* med *sanning ligger i ngt*, och därmed skulle det, åtminstone i formell mening enligt den modell som diskuteras här, vara fråga om spatial, eller möjligen situationell belägenhet. Jag tycker att resonemanget om kroppslighet ovan talar emot en sådan analys; här är det inte fråga om att t.ex. *sanning* är beläget någonstans i fysisk mening, inte heller är *sanning* beläget i en viss omständighet. Snarare har vi att göra med just ett allmänt varande, något *är* sant. (Man kan naturligtvis diskutera detta ytterligare och mer ingående, bl.a. om det kanske vore bättre att sortera denna frastyp i en egen grupp, men jag tänker inte uppehålla mig vid detta här.)

Konstruktionssätten i den fjärde gruppen är V+adj/+NP[+PP] och den subjektivt predikativa/egentligt subjektiva funktionen är naturligtvis en förutsättning här, d.v.s. helt obligatorisk, medan förekommande adverbial är optionella (i ovan nämnda bemärkelse).

Ovanstående beskriver alltså en grundbetydelse (1) och tre grader av utvidgad betydelse (2, 3, och 4). Ordningen hos dessa

fyra betydelser kan man betrakta som en skala av abstraktionsgrad som återspeglas på olika nivåer:

SEMANTISK NIVÅ:			
+	←	LEXIKALT INNEHÅLL	→ -
KROPPSLIGHET:			
+	←	→	-
			
FUNKTIONELL NIVÅ:			
POSITIONELL	>	SPATIAL	>
BELÄGENHET		BELÄGENHET	>
		BELÄGENHET	>
			ALLMÄNT
			VARANDE
STRUKTURELL NIVÅ:			
V[+adv/PP]		V +PP	
		V+adv	
		V[+NP]+PP	
			V+adj
			V+NP PP]

Ju längre till vänster i ordningen ovan, desto mer konkret: mer lexikalt innehåll, mer kroppsligt, mer specifik funktion och färre och lättare obligatoriska argument. Ju längre till höger, desto mer abstrakt: mindre lexikalt innehåll, bortom kroppslighet, mindre specifik funktion, fler möjliga och tyngre obligatoriska argument.

Man kan också se att längs den här skalan "rör sig" de lexikaliserade metaforerna (om de rör sig) åt höger; den bokstavliga betydelsen av ett metaforiskt uttryck finns alltid i samma grupp eller till vänster om den metaforiska användningen. Metaforer återfinns inom alla grupperna, men de allra flesta placerar sig i den grupp som beskriver situationell belägenhet.

3.2. Överförda betydelser

Det jag här kallar överförda betydelser omtalas ibland som metaforiska betydelser. Det är alltså fråga om att ord "adopterar" till en helt annan innebörd än den grundläggande via de bildliga betydelser de får ge språklig dräkt åt i lexikaliserade metaforiska

fraser. I SKO arbetar vi efter principen att varje lexikaliserad fras (d.v.s. alla fraser och uttryck som inte är rena argumentstrukturbeskrivningar eller öppna kollokationer) huvudsakligen behandlas i den artikel som rör frasens huvudord. Fraserna *ligga på latsidan* respektive *ligga ngn till last* behandlas alltså huvudsakligen i artiklarna *latsida* respektive *last*. De ska även nämnas någonstans i artikeln *ligga*, men bara för att hänvisas till respektive huvudordartikel.

Detta innebär inte att vi inte godkänner några som helst överförda betydelser som egna polysemer, men lite tillspetsat blir effekten att en överförd betydelse kräver en produktiv argumentstruktur för att räknas som en polysem¹. Med produktiv menar jag att de argument som omger ett verb inte kräver ett visst lexikalt innehåll för att verbet ska ha en viss innebörd. Detta krav kan ha en mjukare och en hårdare tolkning.

Med den mjukare tolkningen räknar man substantiv, adjektiv och adverb som de lexikalt bärande ordklasserna, och så kan man använda det för att skilja mellan polysemer och slutna kollokationer. Härigenom klassas exempelvis *ligga på halster*, vars argument närmast dignar av lexikalt innehåll, som en sluten kollokation som ska hänvisas till artikeln *halster*, och så sker förslagsvis inom ramen för den polysem vi hittills definierat som 'situationell belägenhet'. Mot detta kan ställas *ligga på ngt*, vars argument alltså inte kräver ett visst lexikalt innehåll för att verbet ska ha den vissa innebörd, 'inte ge ifrån sig, tjuvhålla', som ju avsevärt skiljer sig från grundbetydelsen hos verbet *ligga*. Därmed kan man betrakta *ligga på ngt* som en egen polysem av verbet *ligga*. Detsamma gäller fraserna *ligga med ngn* och *ligga på ngn*.

Med den hårdare tolkningen skulle man räkna även prepositioner som lexikalt bemängda. Kravet skulle då göra skillnad mellan konstruktioner vars formalisering förutsätter en specifik preposition, t.ex. *ligga på ngt*, och konstruktioner vars argument kan utgöras av exempelvis en PP med ospecificerad preposition,

¹ Detta är inte en uttalad hållning inom SKO-projektet som sådant, utan en del av det teoretiska resonemang som förs i denna framställning.

t.ex. *ligga* +(ospec.)sättsadvl i bemärkelsen 'för tillfället inte utnyttjas eller behandlas', som det formuleras i t.ex. SOB/NEO (4), kunna accepteras (om man anser att det är en adekvat tolkning av betydelsen). Detta därför att man kan hävda att *på* i *ligga på ngt* bär på så mycket lexikal information att det inte finns någon annan preposition som kan ackompanjera verbet med bibehållen betydelse, och i så fall är det inte verbet som ensamt står för den lexikala informationen. I stället skulle frasen kunna beskrivas som en bildlig, eller metaforisk användning, av t.ex. den polysem som uttrycker 'spatial belägenhet'. Detsamma skulle gälla fraserna *ligga med ngn* och *ligga på ngn*.

En uppmjukad variant av denna hårdare tolkning skulle kunna medge en PP som kan varieras, om än med ett (väldigt) begränsat antal alternativ. Därigenom skulle *ligga med/hos ngn* kunna passera som egen polysem, men lämna *ligga på ngn* och *ligga på ngt* utanför.

4. Systematiseringen och dess effekter för belägenhetsverben i SKO

I det föregående har jag resonerat kring ett sätt att avgränsa och systematisera användningen av betydelsemoment, eller polysemer, för beskrivningen av belägenhetsverben i SKO. Jag har utgått från en skillnad mellan det jag menar med utvidgade respektive överförda betydelser hos dessa verb. De utvidgade betydelserna har jag försökt isolera med hjälp av argumentens funktion. Syntaktisk funktion visade sig vara ett otillräckligt verktyg, så verbets grad av kroppslighet fick komma till som en ytterligare dimension för att komma fram till grader av betydelseutvidgning. Verbets grundläggande betydelse tillsammans med graderna av utvidgad användning av densamma bildar en sorts primär struktur som bl.a. återspeglar abstraktionsgrad. Dessa formerar de fyra första betydelsemomenten, polysemerna, hos belägenhetsverben. (Jag har, som sagt, koncentrerat mig på att beskriva verbet *ligga* i den här framställningen; jag hoppas återkomma framöver med en samlad beskrivning av de tre belägen-

hetsverben och den reguljära polysemi (jfr Malmberg:1988) som kan sägas råda dem emellan avseende de utvidgade betydelse(rna).)

UTVIDGAD BETYDELSE		ÖVERFÖRD BETYDELSE				
- + abstraktionsgrad - + kroppslighet	POSITIONELL BELÄGENHET	⇒	(ett femte moment')	(ett sjätte moment')	(ett sjunde moment')	(ett åttonde moment')
	SPATIAL BELÄGENHET	⇒	(ett femte moment'')			
	SITUATIONELL BELÄGENHET	⇒	(ett femte moment''')			
	ALLMÄNT VARANDE	⇒	(ett femte moment''''')			

Ur vart och ett av dessa betydelsemoment kan man tänka sig avknoppning av ytterligare polysemer med överförd betydelse av verbet. För att göra en formell skillnad mellan slutna kollokationer och "faktiska" polysemer av verbet, har jag för polysemer med överförd betydelse formulerat ett krav på produktiv argumentstruktur. Detta krav kan ges en mjuk eller en hård tolkning. Med den mjuka tolkningen accepterar man *ligga på ngt* som en egen polysem, men inte t.ex. den betydelse som följer av uttrycket *ligga på halster*. Med den hårda tolkningen accepterar man inte heller *ligga på ngt* som egen polysem, däremot skulle *ligga* [+sättsadvl] i bemärkelsen 'för tillfället inte utnyttjas eller behandlas' kunna accepteras. Jag formulerade också en uppmjukad variant av denna hårdare tolkning som släpper igenom *ligga med ngn* eftersom prepositionen kan varieras, om än i ytterligt begränsad omfattning (*med/hos*), vilket inte alls förekommer hos uttrycken *ligga på ngn* och *ligga på ngt*.

I mitt arbete med belägenhetsverben i SKO lutar jag i skrivande stund åt den uppmjukade hårda tolkningen av kravet på polysemer med överförda betydelse(r). Därmed ser verbet *ligga* för närvarande ut att få följande betydelsemoment:

- (1) vara belägen i vågrätt position
- (2) vara belägen på viss plats
- (3) vara belägen i viss situation el. omständighet
- (4) vara
- (5) ha sexuellt umgänge

Jag vill understryka att detta är en preliminär indelning och att jag säkert kommer att ändra mig flera gånger. Men man måste ju ändå resonera kring det hela. Kanske är det t.ex. inte rimligt att i en ordbok följa en teoretisk ram av struktur och funktion på det sätt som diskuterats här. Bl.a. kan man fråga sig hur intuitivt det är för en ordboksanvändare att *ligga* i betydelsen 'ha gått och lagt sig för natten' presenteras i betydelsemoment (1), *ligga i sängen* i moment (2) och *ligga sömnlös* i moment (4). Kanske vore det i det avseendet mer behändigt att försöka sig på en systematiskt konceptuell sortering av fraserna? Eller kan man tänka sig någon ytterligare möjlighet till lösning? Det får framtida funderande, av mig och andra, avgöra.

Litteratur

A. ORDBÖCKER

Bonniers svenska ordbok. 1994. Bonnier Alba.

LEXIN. 1986. Esselte Studium AB.

Nationalencyklopedins ordbok. 1996. Språkdata och Bokförlaget Bra Böcker AB.

Ord för ord. 1960. Norstedts.

Svensk handordbok. 1966. Norstedts.

Svensk ordbok. 1986. Språkdata och Esselte Studium AB.

B. ANNAN LITTERATUR

Clausén, Ulla. 2001. "Funktion och attityd i Svensk konstruktionsordbok" (denna volym s. 80–93).

- Ekberg, Lena. 1989. *Gå till anfall och falla i sömn. En strukturell och funktionell beskrivning av abstrakta övergångsfraser*. Studentlitteratur, Lund.
- Malmgren, Sven-Göran. 1988. "On Regular Polysemy in Swedish." *Studies in Computer-aided Lexicon*, 179–200. Stockholm.
- Malmgren, Sven-Göran. 1992. "Om behandlingen av polysema ord i enspråkiga ordböcker." I: *Nordiske Studier i Leksikografi*. Nordisk Forening for Leksikografi, Oslo.
- Malmgren, Sven-Göran. 1994. *Svensk Lexikologi. Ord, ordbildning, ordböcker och databaser*. Studentlitteratur, Lund.
- Svensén, Bo. 1987. *Handbok i svensk lexikografi. Principer och metoder i ordboksarbetet*. Esselte Studium/Tekniska Nomenklaturcentralen, Stockholm.

Appendix

Betydelsemoment hos verbet *ligga* enligt ...

Svensk handordbok

I om person el. djur (befinna sig i horisontalläge) **1)** sova, övernatta; vara sängliggande sjuk; vara lat, lata sig **2)** i uttr. ha samlag med **3)** om avliden **4)** oeg. **5)** om fågel **6)** med prep. i vara i verksamhet (med)

II om föremål i horisontalläge o.d. **1) – 2)** om vapen, redskap o.d. **3)** om fordon **4)** sjö. i uttr. för djupgående **5)** kortsp. om utspelat kort icke tagas upp för att ersättas med annat kort **6)** förbli oanvänd el. osåld o.d.; förvaras, vara placerad **7)** i uttr. för att ngt störtat samman o.d. **8)** om is täcka ett vattendrag; om sjö vara isbelagd, tillfrusen **9)** i uttr.

III bildl. **1)** i uttr. ha ngt till låns el. under behandling; dröja med att lämna ifrån sig ngt **2)** vara under behandling, vänta på avgörande **3)** i uttr. lämna ngt åt sitt öde, försumma **4)** tynga trycka på ngt; om andligt tryck, arbetsbörda o.d. **5)** i uttr. det är viktigt, betydelsefullt **6)** i uttr. utgöra grundvalen för

IV– **1)** utbreda sig, utsträcka sig, vara belägen; vara riktad el. vänd på visst sätt; vetta åt visst håll **2)** bildl. vara beskaffad, befinna

sig i visst läge **3**) om ljussken, mörker o.d. som är utbrett över el. inhöljer ngt; äv. om landskap o.d. som utbreder sig framför en **4**) i vissa uttr. är tydligt; vara uppenbar; är övergiven **5**) i uttr. utgöra en fortsättning av; äv. bildl. överensstämma med, passa ihop med **6**) sjö. ha visst läge, hålla viss kurs **7**) om vinden, solen o.d. träffa, falla **8**) bildl., i uttr. angivande anlag, lämplighet o.d.

V – **1**) uppehålla sig, vistas; vara förlagd, vara borta el. ute någonstans **2**) om fartyg vara förankrad el. förtöjd **3**) i uttr. **4**) i uttr.

VI med allmänt lokaliserande innebörd **1**) ha sin plats, befinna sig vara; är ett släktdrag **2**) i sht sport. inta en viss plats **3**) i uttr. **4**) i uttr. är föremål för ngn's tankar el. omsorger **5**) i sht hand. befinna sig högt, lågt osv. på prisskalan; stå i viss kurs

VII – **1**) vara sysselsatt med, hålla på med, vara invecklad i; anstränga sig, vara ivrigt sysselsatt; sport. träna; ägna sig åt försvarsspel **2**) fackspr., i uttr. hålla på att lastas, lossas

VIII i sht lant., i uttr. som anger att ett jordstykke icke odlas el. att det brukas på ett visst angivet sätt **IX** fackspr. **1**) tillhöra samma byalag som; brukas tillsammans med

Ord för ord

1) vara l. ligga utsträckt; vara sängliggande, (skämts.); (bildl.) vila l. ruva l. utbre(da) sig (över) **2**) vistas, uppehålla sig, befinna sig, hålla sig l. löpa l. springa l. åka, köra (bakom l. efter resp. före); vara förlagd, vara stationerad; (sjö.) vara förankrad, vara förtöjd; vara belägen, sträcka sig, utbreda sig, vetta **3**) ha sin plats, förvaras **4**) 'falla' l. 'lägga sig' (i veck l. bucklor); **5**) (om höna) ligga på ägg, ruva; (om säd) ligga l. vara nedslagen (av regn); (om sjö) vara isbelagd l. tillfrusen

LEXIN

vara i vågrätt läge, vila; vara placerad försvagat finnas

Bonniers svenska ordbok

vara, befinna sig, eg. i vågrätt läge; vara belägen; (friare el. bildl. i många fall)

Svensk ordbok / Nationalencyklopedins ordbok

1) vila med den största ytan mot underlaget om varelse eller föremål
2) vara belägen i visst område e.d. **3)** uppehålla sig på viss plats; ofta med antytt syfte **4)** för tillfället inte utnyttjas eller behandlas **5)** med prep. **på** avvakta med att behandla ngt ärende e.d. **6)** med prep. **i** vara i verksamhet (med)

*Godelieve Laureys
Gudrun Rawoens*

Bilingual leksikografi Nederlandsk–dansk ordbogsprojekt

In 1997 the Department of Nordic languages of Ghent University started up a Dutch–Danish dictionary project. The aim is to create a bilingual database of about 40.000 lemmas. The commissioner of the project is the CLVV ("Committee for lexicographical interlingual resources"), a Dutch–Belgian official agency that coordinates bilingual dictionary projects with Dutch as one of the involved languages. A Dutch–Danish bilingual lexical database is produced by means of the editor OMBI with a reverse function as one of its key features. RBN ("Referentiebestand Nederlands"), a corpus based predictionary lexical database, forms our material for the source language. In this article we intend to present this project by describing our way of working with the OMBI-program (network, advantages, problems) and with the RBN-material (macro, micro). A few editorial matters will be highlighted and illustrated with examples.

Indledning

Der eksisterer i øjeblikket ikke nogen udførlig nederlandsk–dansk/dansk–nederlandsk ordbog. Udbuddet begrænser sig til et par nederlandsk–danske og dansk–nederlandse lommeordbøger og nogle få ordlister.¹ En gedigen bilingual ordbog, der behandler nutidssproget og som omfatter en betydelig mikrostruktur, har længe været savnet. Selvfølgelig bør nævnes De Vries' nederlandsk–danske og dansk–nederlandse ordbog, der udkom første gang i 1971, og som blev genoptrykt i 1986. Denne ordbog rummer ganske vist ca. 25 000 opslagsord i hver del, men den adskiller sig ikke væsentlig fra en almindelig parlør hvad mikrostrukturen angår. Den er endvidere udgået.

¹ Van Goor's Mini-pockets Deens woordenboek, 1995/13 (12.000 lem-maer); Wolter's Miniwoordenboek Deens–Nederlands, Nederlands–Deens, 1999/5 (10.000 lem-maer); H. Alkema, I. Marsman, H. Westra-Lankamp, Nederlands Scandinavisch Basiswoordenschat Deens–Noors–Zweeds–IJslands, 1989.

Paludans hollandsk–dansk, dansk–hollandske ordbog i ét bind (12.000 opslagsord), der udkom i 1997, er det seneste skud på stammen. Det viste sig imidlertid at være et rent og skært genoptryk af Van Goors lommeordbog, inklusiv forordet, hvor man bl.a. understreger relevansen af ordbogen i forbindelse med Danmarks indtræden i EF (sic!).¹

Præsentation af projektet: Størrelse og profil

I 1997 påbegyndtes arbejdet med en ny nederlandsk–dansk ordbog ved Institutet for nordiske sprog ved universitetet i Gent i Belgien.

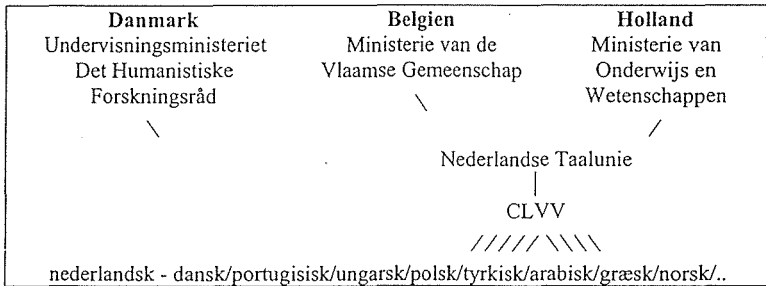
Målsætningen med projektet er at udarbejde en bilingval nederlandsk–dansk kontrastiv leksikalsk database på cirka 40 000 lemmata, af hvilken en tosprogsordbog kan fremstilles. Ordbogen og databasen er koncipieret som et bidirektionelt værktøj, hvilket indebærer, at den skal være lige så anvendelig for dansksprogede som for nederlandsksprogede brugere. Den nederlandsk–danske del fungerer fortrinsvis som en produktionsordbog for nederlandsktalende og som en receptionsordbog for dansk-talende. Det bidirektionelle udgangspunkt medfører en række redaktionelle konsekvenser. Databasen og den tiltænkte ordbog begrænses til 40 000 lemmata, men fokuserer på en rig mikrostruktur. Dette valg er grundet i det faktum, at nederlandsk og dansk er nært beslægtede sprog. For ordbogsbrugere, der arbejder med dette sprogpår, indebærer det, at de allerede på et tidligt stadium har mulighed for at søge hjælp i monolingvale ordbøger. Vi anser altså, at den største værdi for ordbogsbrugeren er et stort udvalg af grammatiske og leksikalske ordforbindelser, illustrative eksempelsætninger og faste udtryk.

Projektet udføres i opdrag af det nederlandsk–belgiske udvalg CLVV, "Committee for lexicographical interlingual resources". Søsterprojektet, dvs. den dansk–nederlandske del, bliver udar-

¹ Ordbogen fik en kritisk anmeldelse af Bergenholz og Van Hees. Se: Henning Bergenholz & Annelies van Hees, 1997.

bejdet ved Amsterdam Universitet. CLVV har nedsat en fælles styregruppe, der koordinerer begge projekter.

Projektet er finansieret med offentlige midler, der er stillet til rådighed af Danmark, Belgien og Holland. Midlerne forvaltes af CLVV, som sorterer under sprognævnet for det nederlandske sprogområde ("Nederlandse Taalunie"), og som koordinerer en række tosprogsordbogsprojekter, hvor nederlandsk udgør et af sprogene.



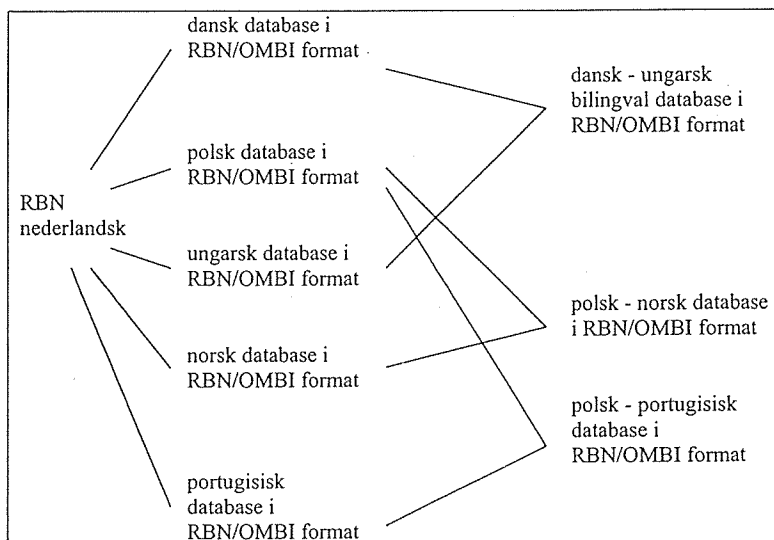
Det nederlandske kildemateriale

Det nederlandske kildemateriale for ordbogsprojektet er en relationel leksikalsk database ved navn RBN ("**R**eferentie**b**estand **N**ederlands").

RBN er dels baseret på et korpus på ca. 5 millioner ord, udarbejdet ved INL ("Instituut voor de Nederlandse Lexicografie" ved universitetet i Leiden), der indeholder materiale fra både skriftlige og mundtlige kilder, og dels på en 21 000 ords leksikalsk database udarbejdet ved Instituttet for Leksikologi ved Vrije universiteit Amsterdam. RBN har et format, der gør det muligt at importere det direkte i editor-programmet. RBN er udarbejdet på foranledning af CLVV som et samarbejdsprojekt mellem universiteterne i Amsterdam, Utrecht, Leiden og Leuven. CLVV har to målsætninger for RBN, en praktisk og en forskningsmæssig. Den praktiske hensigt er at stille en nederlandsk

leksikalsk database til rådighed for de forskellige projekter, der udføres under koordination af CLVV, således at de på forhånd ikke er afhængige af et forlag. De enkelte projekter kan selv beslutte, om de bruger RBN som kildesprogsmateriale, målsprogsmateriale, eller kun som referencedatabase. Desuden kan hvert projekt selektere relevante dele af makro- og mikrostrukturen i overensstemmelse med den type af ordbog, der fremstilles. For fremstillingen af en ordliste på 20.000 lemmaer sletter man f.eks. halvdelen af RBN-lemmaerne, og fjerner eksempelsætningerne.

Den forskningsmæssige målsætning er at teste den såkaldte "hub and spoke" model¹ ved at matche (varianter af) et og samme kildesprog med forskellige målsprog. I og med at de respektive fremmedsproglige databaser er ens strukturerede, vil det være muligt på et senere tidspunkt at kunne sammenføre to "målsprogsdatabaser" med hinanden. Tanken er at en sådan ny bilingval database danner udgangspunkt for udarbejdelsen af en bilingval ordbog for et nyt sprogpar.



¹ Willy Martin, Anne Tamm, Karoli Gaspar, OMBI: An Editor for Constructing Reversible Lexical Databases, i: Euralex '96 Proceedings, s. 675-687.

Editor

Den nederlandsk-danske ordbog fremstilles ved hjælp af ordbogsredigeringsprogrammet OMBI. OMBI er et akronym for "**O**mkeerbare **B**ilinguale **L**exicale **D**atabanken" dvs. et program til udarbejdelse af tosprogsordbøger, som automatisk tager højde for at bytte om på kilde- og målsprog. OMBI er desuden skræddersyet til RBN.

Programmet OMBI er blevet installeret på stedet i Gent i et flerbrugermiljø på Windows NT-computere, hvor OMBI kører på workstations, mens ordbogsdatabaseen ligger centralt på en databaseserver. Programmet OMBI har en vis såkaldt "scalability", hvilket indebærer, at det er muligt at konfigurere den som "stand-alone" på en Windows 95/98/NT computer. Desuden kan OMBI installeres i en "multi-user" miljø, hvori OMBI ("client program") kører på en "workstation" mens ordbogsdatabaseen administreres af en database server (Interbase).

I øjeblikket bruger vi Windows NT både til disse "workstations" og til serveren, på grund af den større stabilitet som Windows NT tilbyder i sammenligning med Windows 95/98. Vi har udført nogle prøver, hvori vi lod OMBI køre på en Linux Interbase server. Det fremgik heraf, at Linux-omgivelsen er meget hurtigere end Windows NT opstillingen. For vort projekt tilbyder Windows NT Interbase serveren imidlertid den bedste løsning, eftersom serveren da også kan bruges som ekstra data-entry station.

Selv om OMBI er et avanceret program med mange interessante træk og funktioner, er der også nogle ulemper. Et eksempel er mangelen på "managing capabilities" og på en anordning for replication/version-control.

- Ved hjælp af "managing capabilities" ville det være muligt at foretage en slags "user- and database-management". På den måde ville man kunne specificere, hvilken form for adgang en bruger skal have til en specifik del af databasen, hvilket f.eks. ville indebære, at man kunne begrænse denne persons adgang til nogle specifikke handlinger på databasen.

- Hvis der var blevet indbygget en "replication-functionality" i programmet, havde det været muligt at arbejde samtidig ("concurrent") på forskellige steder (f.eks. medarbejderne i Danmark). Takket være "replication" ville det være muligt at sammenføre databaserne uden at man skulle være bange for at "overskrive" data. Dette ville kunne lade sig gøre blandt andet ved at installere Windows NT Terminal server edition, der gør det muligt at køre "concurrent, multiple user-sessions" på en NT-server. Dette ville medføre, at en medarbejder, der ikke befinder sig på stedet, via Internet kunne arbejde direkte i en "graphical terminal session" på serveren i Gent.

Arbejdsgang og hjælpemidler

Redaktionsgruppen består af dansksprogede og nederlandsksprogede redaktører, der har gode kundskaber i det respektive andre sprog, samt et par tosprogede redaktører, der har en seminatale kompetence i begge sprog. Alle redaktører har en akademisk uddannelse med dansk og/eller nederlandsk som fag. Fire redaktører arbejder permanent på stedet, fire korrekturlæsere har distancearbejde på freelance-basis, og endelig er der 2 deltidsredaktører, der pendler mellem Danmark og Belgien.

Udover diverse elektroniske ordbøger har redaktørerne adgang til Bergenholtz danske 4 millioner ords korpus, 6 millioner ords korpusset fra den danske ordbog, tilsvarende nederlandske korpora, samt alle de informationskilder, som internettet giver adgang til.

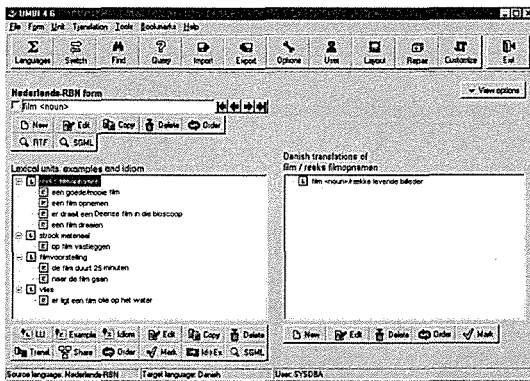
Redaktionsarbejdet er opdelt i forskellige faser. Før selve redaktionsarbejdet foretages en såkaldt nulfase, hvor en nederlandsktalende redaktør bearbejder den nederlandske leksikalske database og importerer det selekterede materiale i ordbogsredigeringsprogrammet. I den første fase oversætter redaktører med dansk som modersmål fra nederlandsk til dansk i redigeringsprogrammet. Fase 2 er den første korrektionsfase, hvor korrektur-

læsningen sker på papir af en korrekturlæser med dansk som modersmål. I fase 3 tages der stilling til bemærkninger, rettelser og ændringsforslag fra den forrige fase. Dette arbejde udføres af hovedredaktøren i samråd med en af de dansksprogede redaktører (dvs. en som ikke selv har oversat dokumentet). De endeligt fastslåede rettelser indføres nu også i den elektroniske database. Den sidste fase omfatter gennemlæsning og endelig redigering.

Oversættelse ved hjælp af OMBI

Programmet har en vendefunktion, hvilket indebærer, at der under oversættelsesprocessen simultant fremstilles en leksikalsk database over målsproget.

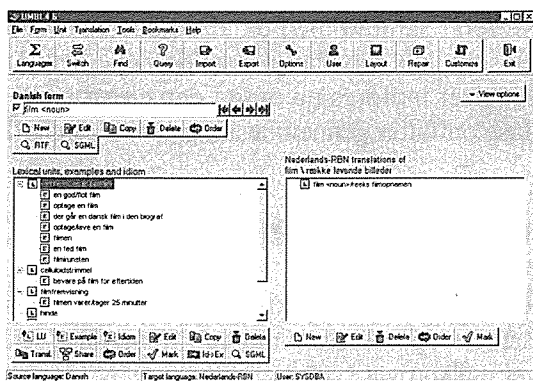
Vi vil prøve at demonstrere databasens struktur og oversættelsesprocessen ved et konkret eksempel.



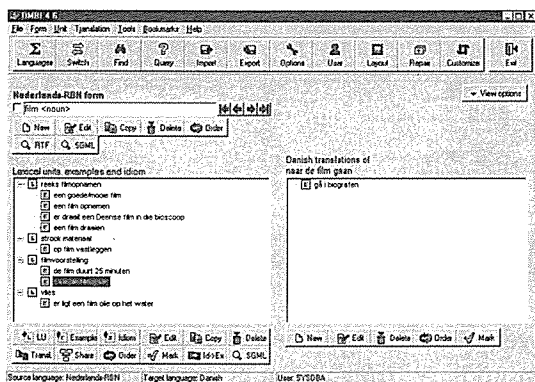
I feltet til venstre står det nederlandske opslagsord – i OMBI kaldt ”form unit” – ”*film*”, som er opdelt i forskellige betydningsvarianter – såkaldte ”lexical units”.

Eksempelsætningerne er hængt op under en bestemt lexical unit. Oversætteren starter med den første nederlandske lexical unit og gengiver den med en dansk lexical unit. I feltet til højre ser vi resultatet: ”*film*”. Ved hjælp af vendefunktionen bliver den

danske oversættelsesækvivalent nu indført i den danske database.

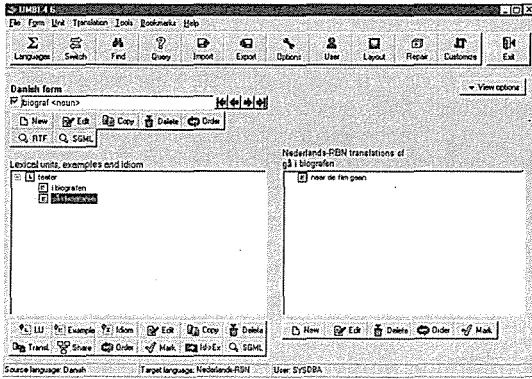


På dette skærmbillede ser vi resultatet, når hele opslagsordet ”*film*” er blevet oversat til dansk. Vi kan nu se at det danske lemma ”*film*” ligeledes er opdelt i betydningsvarianter (LU), og at eksemplerne er indført under den rette betydningsvariant.

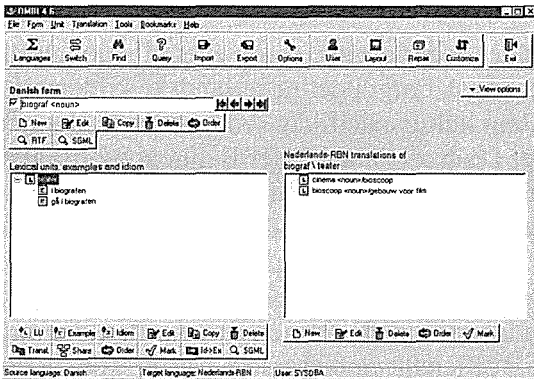


Dette screenshot viser, at eksempelsætningen ”*naar de film gaan*” er blevet oversat med ”*gå i biografen*”. Den danske eksempelsætning vil derfor blive hængt op under det danske lemma ”*biograf*”. Ved at aktivere vendefunktionen vil man så

kunne se at oversættelsen af eksempelsætningen ”gå i biografen” allerede figurerer i den nederlandske database:



Dette proces medfører, at det danske lemma ”*biograf*” i systemet selvfølgelig også lænkes til andre nederlandske oversættelsesækvivalenter: ”*bioscoop*”, ”*cinema*” som det fremgår af næste skærmbillede:



I princippet foretrækker redaktionsgrupper ”en-til-en” oversættelser, men i en række tilfælde har vi valgt at angive flere mulige oversættelsesækvivalenter. Det forekommer også, at et opslagsord kan gengives på dansk med flere forskellige ækvivalenter,

der dog hver har specifikke selektionsbegrænsninger. Disse begrænsninger angives efter oversættelsen med < >.

emotioneel ¹ <adj> (adv; ncompar) **1. m.b.t. emoties følelsesmæssig, emotionel, følelsesbetonet** **a.** emotioneel stabiel *følelsesmæssigt stabil* **b.** een emotionele band met iets of iemand hebben *føle sig følelsesmæssigt forbundet med noget eller nogen* **c.** emotionele problemen *følelsesmæssige problemer, emotionelle problemer.*

emotioneel ² <adj> (adv; -er/-st) **1. gevoelig følelsesladet** <situation, handling>, *letbevægelig* <persoon>, *emotionel* <persoon> **a.** een emotioneel afscheid *en følelsesladet afsked* **b.** een emotionele reactie *en følelsesladet reaktion* **c.** emotioneel reageren *reagere med stærke følelser.*

Det er også muligt at angive, at en oversættelse har et begrænset brugsområde: f.eks. ved *Kamerdebat*, hvor <i DK> angiver, at oversættelsen *folketingsdebat* kun gælder den danske parlamentariske forsamling.

Kamerdebat <noun> (-ten; het; n) folketingsdebat <i DK>, parlamentarisk debat

I og med at det nederlandske sprog er et binationalt sprog bliver dette princip også brugt i det nederlandske kildemateriale for at skelne mellem kulturspecifikke ord og termer, der udelukkende forekommer i enten Holland eller Belgien, og for at angive den intralingvale variation.¹ Der bruges forskellig "labelling" for de såkaldte flamske ord, dvs. ord, der kun bruges i den nederlansktalende del af Belgien men ikke i Holland. f.eks.: '*panikeren*' (= '*in paniek raken*') eller '*microgolfoven*' (= '*magnetron*'), og for de såkaldte institutionelle termer, dvs. termer, der angiver institutioner eller officielle embeder som har forskellige navne i resp. Holland og Belgien.

byrådsmedlem <noun> (-met; -mer) wethouder <i NL>, schepen <i B>

¹ For en uddybning af begrebet "natiolekt" se: Godelieve Laureys, 1997.

I visse tilfælde har vi valgt ikke at oversætte kulturspecifikke ord, men at gengive dem med en encyklopædisk beskrivelse, som f.eks. ved *flevopolder*:

flevopolder <noun> (-; de; mf) *inddæmmet område som udgør provinsen Flevoland i Holland.*

Makrostruktur

Selv om RBN-materialet er ”skræddersyet” til bilingval leksikografi, besluttede redaktionsteamet dog at gennemføre yderligere et udvalg af makrostrukturen. RBN har en størrelse på 45 000 ord eller lemmaer, mens vort projekt er beregnet på en ordbog på ca. 40 000 lemmaer.

Udvælgelseskriterierne for lemmaerne er frekvens, kulturspecificitet og kontrastiv relevans. Desuden fjernede vi et antal lemmaer, som var for ”korpuspåvirkede”. Her tænker vi bl.a. på en række termer fra skibsindustrien.

Derfor opstillede vi i en forberedelsesfase udvælgelseskriterier for at slette visse lemmaer og eventuelt tilføje andre. Alle forandringer gemmes i separate filer. Dette arbejde bliver udført af en nederlandsktalende redaktør i samråd med medarbejdere fra instituttet for det nederlandske sprog.

Opdelingen af lemmaerne er baseret på rent formelle, dvs. kategoriale eller subkategoriale, kriterier. Den mindste forskel i de morfologiske paradigme giver anledning til et nyt lemma. f.eks.

chagrijn¹ <noun> (-; het; n) **1.** *knorrigheid* [informal] *græmmelse, gnaveri, surmuleri a. iemand met chagrijn en surmuler.*

chagrijn² <noun> (-en; de; m/f) **1.** *persoon* [informal] *stivstikker, tørvetriller, dødbider, surmuler a. een stuk chagrijn en sur skid, dødbider.*

Mikrostruktur

Mikrostrukturen i den elektroniske database indeholder information om forskellige kategorier. Denne information er til rådighed for både redaktører og korrekturlæsere. Redaktionsgruppen har udarbejdet en manual, hvoraf det fremgår hvilken type informationer, der skal synliggøres i ordbogsformatet. De øvrige oplysninger gemmes i systemet, hvilket på et senere tidspunkt vil gøre det muligt for forlæggeren eller fremtidige brugere af databasen selv at selektere den type information, som anses for værende relevant. I RBN bruges engelsk som metasprog.

I systemet ligger der for hver LU et résumé eller en kortfattet definition. Ved polysemiske opslagsord fungerer résuméet som betydningsdiskriminator for de forskellige LU. Ved polysemiske opslagsord er det meningen, at de kommer med i ordbogen som hjælp for ordbogsbrugeren. På et dybere niveau ligger der også en mere udførlig definition, der svarer til behandlingen af opslagsordet i en monolingval ordbog. Dette gør ikke kun arbejdet for oversættere og korrekturlæsere lettere, men garanterer også en entydig tolkning af hver enkel LU.

Hvert lemma er forsynet med oplysninger om stavemåde, morfologi, syntaks, semantik, pragmatik og kombinatorik.

Med henblik på stavemåde følges *Woordenlijst Nederlandse taal* (1997) for nederlandsk og *Retskrivningsordbog* (1998) for dansk. Dobbeltformer angives i overensstemmelse med samme standarder.

For hvert lemma angives ordklasse og morfologisk type, f.eks. ved substantiver angives, om der er tale om et simplex, en sammensætning eller en afledning, og der angives pluralisendelse; ved verber angives bøjningsklasse og typologi.

Den syntaktiske beskrivelse af et opslagsord omfatter oplysninger om ordets kombinatorik, f.eks. ved substantiver angives genus og artikelbrug; ved verber, visse adjektiver og substantiver angives valensstruktur og faste præpositioner; ved adjektiver angives, om de kan bruges prædikativt og/eller adverbialt, og om de kan kompareres.

I det semantiske felt får opslagsordet en kode, der angiver semantiske tilhørsforhold, f.eks. ved substantiver angives referencestrukturen (f.eks. proprium, artnavn/dividua, individua), semantisk type (f.eks. abstrakt, konkret), leksikaliseret metonymi (f.eks. dynamisk, ikke-dynamisk).

Det pragmatiske felt omfatter syv forskellige etiketter, der indkredser ordets brugsområde: f.eks. angives, om et ord er et fremmedord, et regionalt ord, om det har en vis konnotation, om det tilhører et socialt register, og om det er tidsdateret. Systemets dybdestruktur kan f.eks. illustreres ved at etiketten ”konnotation” er repræsenteret af fem forskellige markører: *neutral* (default), *offensive*, *pejorative*, *euphemistic*, *jocular*.

Den rige mikrostruktur af RBN kendetegnes desuden af et stort udvalg af eksempelsætninger og idiomatiske udtryk. RBN omfatter to typer eksempler: kanoniske eksempler og kontekstuelle eksempler, der er taget direkte fra korpus. Som regel vælges den kanoniske form af pladshensyn, men ofte foretrækkes et (”plukket”) korpuseksempel eller et redaktionseksempel for at illustrere brugen af et opslagsord i en bestemt kontekst. Det gælder især faste vendinger og pragmatiske udtryk.

Den leksikalske database rummer altså en stor mængde information på mange forskellige niveauer. Den forfinede beskrivelse af kildeproget letter oversættelsesprocessen. Vi bestræber os på ved hjælp af alle disse oplysninger om de nederlandske opslagsord at ramme den mest hensigtsmæssige danske oversættelsesækvivalent. Samtidig udarbejder vi simultant en dansk database, der kan betegnes som en rå predictionary version. De danske oversættelsesækvivalenter bliver forsynet med et dansk résumé og linket til en nederlandsk LU. De øvrige oplysninger om morfologi, syntax, semantik, m.v. bliver ikke systematisk indført i den danske database, i og med at det ikke indgår i vores opgave at udarbejde den dansk-nederlandske del. Selvfølgelig er det fuldt muligt at supplere den danske database på et senere tidspunkt eller at lave uddrag af den til udforskning af forskellige sproglige fænomener, f.eks. valens. Det er vigtigt at påpege, at den nederlandske og den danske database er linket sammen på LU-niveau, hvilket indebærer, at ”en bestemt nederlandsk op-

slagsform med en bestemt betydning” læknes til ”en bestemt dansk opslagsform med tilsvarende betydning”. En vending af bestanden vil give anledning til en brugbar udgangspunkt for en mere dybtgående beskrivelse af den danske del af databasen i samme format som RBN. Dette kan ses som en implementering af det ovennævnte ”hub and spoke” model.¹

Til sidst vil vi illustrere arbejdet ved et uddrag af databasen.²

¹ Se grafisk forestilling på side 3.

² Disse eksempler er taget fra materiale der ligger i fase 3.

erg¹ <adj> (adv; -er/-st; attr) I. *ontzettend* *forfærdelig*, *frygtelig*
 a. iemands ergste vijand *nogens værste fjende* b. erge
 honger/haast hebben *være frygtelig sulten/have frygteligt*
travlt c. ik doe nooit inkopen tijdens de ergste drukte *jeg kø-*
ber aldrig ind midt i myldretiden d. erge buikpijn hebben
have (noget) så forfærdeligt ondt i maven, have en forfærde-
lig mavepine e. tijdens de ergste drukte rinkelde de telefoon
midt i det hele ringede telefonen f. zij heeft het erg te pakken
hun går meget op i det, hun er smaskforelsket <verliefdheid>

erg² <adj> (nonadv; -er/-st) 1. *naar* *dårlig*, *slem* a. iets erg
 vinden *synes at noget er dårligt* b. een erge ziekte *en slem*
sygdom, en alvorlig sygdom c. en, wat erger is, *og/men hvad*
værre er, d. (het wordt) hoe langer hoe erger (*det bliver*)
værre og værre (jo længere tid der går) e. wat erg! *det var*
da ærgeligt! f. in het ergste geval *i værste tilfældelfald* g. zo
 erg is het niet så *slemt er det ikke, det er ikke så slemt, det er*
ikke så farligt h. dat is niet erg, hoor! *det gør ikke noget!* i.
 dat is toch niet zo erg *det er vel ikke så slemt, det går nok* j. is
 het erg als ik na tien en bel? *gør det noget hvis jeg ringer ef-*
ter ti? k. het middel is erger dan de kwaal l. iets erger maken
 dan het is *gøre noget værre end det er* m. ik vind het echt
 heel erg *det gør mig meget ondt* n. ik vind het erg dat hij die
 baan niet kreeg *jeg er ked af at han ikke fik det job, det var*
da ærgeligt at han ikke fik det job o. als je het niet erg vind,
 ga ik nu weg *hvis du ikke har noget imod det, tager jeg af*
sted nu p. er is iets ergs gebeurd *der er sket noget forfærde-*
ligt q. zoiets ergs heb ik nooit meegemaakt *det er det værste*
jeg har været ude for r. Gudrun is op het ergste voorbereid
Gudrun er forberedt på det værste s. het is meer dan erg *det*
er helt igennem forfærdeligt t. de toestand van de patient is
 erger geworden *patientens tilstand er blevet værre* u. ach,
 het had erger kunnen zijn! *det kunne have været værre* v. het
 is zo al erg genoeg *det er slemt nok som det er, det er galt*
nok som det er w. des te erger! *det gør ikke (ligefrem) sagen*
bedre! 2. *verkeerd slem, forfærdelig* a. iets ergs gedaan heb-
 ben *have gjort noget forfærdeligt* b. van kwaad tot erger
 vervallen *gå den gale vej, gå ned ad bakke* c. nu maakt hij
 het te erg! *nu er han gået for vidt!*

erg³ <noun> ¶ I. geen erg hebben in iets *ikke være opmærksom*
på noget, komme til at overse noget II. iets zonder erg doen
komme til at gøre noget, gøre noget uden at tænke sig om III.
 daar had ik geen erg in *det tænkte jeg ikke på.*

ernst <noun> (-; de; m) **1. ernstig zijn alvor** a. er klonk ernst in zijn stem *hans stemme lød alvorlig* b. ernst en luim c. maar nu in ernst *spøg til side, men alvorligt talt* d. je ernst bewaren *bevare Alvoren* **2. vastbeslotenheid, oprechtheid alvor** a. in ernst *alvorligt* b. het is mij (hoge/heilige) ernst *det er mit ramme alvor* c. in alle ernst *i ramme alvor* **3. oorzaak dan wel gevolg van ernst alvor** a. de ernst van de situatie *situationens alvor* b. (het is) bittere ernst (*det er*) *ramme alvor*, (*det er*) *bitter alvor* c. het wordt ernst met zijn plannen *han gør nu alvor af sine planer*.

ernstig <adj> (adv; -er/-st) **1. serieus alvorlig** a. iemand ernstig toespreken *sige nogen et alvorstord* b. iets ernstig overwegen *alvorligt overveje noget* **2. ingrijpend alvorlig** a. ernstig gewond *alvorligt såret* b. een ernstige beschuldiging *en alvorlig beskyldning* c. een ernstige ziekte *en alvorlig sygdom* d. een ernstig getroffen gebied *et område som er alvorligt ramt* e. ernstige gevolgen *alvorlige følger* f. iemand iets ernstig aanrekenen g. zich ernstig(e) zorgen maken *være alvorligt bekymret*.

denken <verb> (intr) (h ; dacht; gedacht) **1. het verstand laten werken tænke** a. denken dat ... *tro at ...* b. daar moet ik nog eens over denken *det skal jeg lige tænke over igen* c. ik denk van niet [*informal*] *det tror jeg ikke* d. wat denk je ervan? *hvad synes du?* e. denk om (de buren) *tænk på (naboerne)* f. dat geeft te denken *det giver stof til eftertanke* g. ik moet er niet aan denken *bare tanken* **2. plannen tænke** a. ik denk erover te stoppen met roken *jeg tænker på at holde op med at ryge, jeg overvejer at holde op med at ryge* b. hoe denk je erover? *hvad synes du om de?, hvad mener du om det?* **3. niet vergeten** a. aan iemand denken *tænke på nogen* b. denk aan je tandenborstel *husk lige tandbørsten*.

Litteratur

- Jens Axelsen. 1998. *Retskrivningsordbog*.
 Henning Bergenholz, Ilse Cantell, Ruth Vatvedt Fjeld, Dag Gundersen, Jón Hilmar Jónsson, Bo Svensén. 1997. *Nordisk leksikografisk ordbok*.
 Henning Bergenholz & Annelies van Hees. 1997. "Hollandsk-Dansk Dansk-Hollandsk Ordbog." København Paludan. Anmeldelse i: *Hermes* 19, Aarhus 1997:246-250.

- De Vries nederlandsk–dansk/dansk–nederlandsk*, 1971/76 (Van Goor), 1986 (Kramers).
- Godelieve Laureys. 1999. "Den svensk–nederländska/nederländsk–svenska ordboken. En bidirektionell ordbok i ett dubbelt binationellt perspektiv." *Nordiska studier i lexikografi*. 4. (Skrifter utgivna av Nordiska föreningen för lexikografi. 5.) Helsingfors. S. 249–254.
- Godelieve Laureys & Hans de Groot. 1996. *Van Dale Handwoordenboek Zweeds–Nederlands/Nederlands–Zweeds – Norstedts svensk–nederländska/nederländsk–svenska ordbok*.
- Willy Martin, Anne Tamm, Karoli Gaspar. "OMBI: An Editor for Constructing Reversible Lexical Databases." *Euralex '96 Proceedings*. S. 675–687.
- Willy Martin, Hanny Demeersseman, Isa Maks. "Referentiebestand Nederlands. Documentatie." (Interne publikation.)
- Bo Svensén. 1991. *Handbok i lexikografi*. Stockholm.
- Mark van Elswijk, Cecile Wijne, Isa Maks, Willy Martin. "OMBI 4 User Manual." (Interne publikation.)

Henrik Lorentzen

Jagten på det gode citat

Om vanskelighederne ved at finde egnede ordbogseksempler i et korpus

The paper deals with the difficulties involved in selecting good dictionary examples from a large text corpus, especially if all examples are to be taken directly from the corpus without modifications. The author gives examples of the problems that the lexicographer has to cope with when he is put face to face with authentic language, such as length, disturbing elements, difficult words, errors, and lack of naturalness. Thus, the emergence of text corpora has not made the lexicographer's work easier, but the result will hopefully be a better dictionary with better examples.

Indledning

Denne artikel handler om sprogbrugseksempler i en stor monolingval ordbog og i særdeleshed om de vanskeligheder der er forbundet med at udvælge autentiske eksempler i form af citater fra et større, elektronisk søgbart tekstkorpus. Umiddelbart virker det jo tillokkende at have en stor tekstmængde til rådighed bestående af "rigtigt" sprog, sætninger som folk faktisk har skrevet eller sagt. Nu skulle man endelig få mulighed for let at finde eksempler på sproget som det bruges ude i verden, modsat sådan som ordbogsredaktøren forestiller sig sproget. Virkeligheden er imidlertid lidt anderledes. Ganske vist er alle de forekomster man møder, autentisk, faktisk forekommet sprog, men det er ikke ensbetydende med at de egner sig til at være citater i en ordbog. Ud af flere hundrede forekomster kan det ofte være svært at finde bare et citategnet belæg, bl.a. fordi de autentiske belæg netop forekommer i en sammenhæng og derfor har svært ved at fungere som løsrevne ordbogseksempler. I artiklen vil jeg give eksempler på de forhold der kan diskvalificere et citat, og i den sammenhæng se på citatets funktion i ordbogsartiklen.

Mine erfaringer og mit eksempelmateriale stammer fra mit arbejde som redaktør på Den Danske Ordbog, et projekt under Det Danske Sprog- og Litteraturselskab. Arbejdet med denne nye danske modersmålsordbog har nu været i gang i knap 8 år og forventes afsluttet i 2002, hvor ordbogen begynder at udkomme i 6 bind. Den Danske Ordbog skal efterfølge *Ordbog over det Danske Sprog* (ODS), som beskriver dansk indtil ca. 1955. Ordbogen bygger primært på et elektronisk tekstkorpus der dækker perioden 1983–92. Ud over at ligge til grund for beskrivelsen af ordenes betydninger skal korpus også levere autentiske eksempler, citater, til illustration af ordenes betydning og brug (se evt. mere om Den Danske Ordbog på Internettet: www.dsl.dk/ddo.html)

Når man ser bort fra ODS, er det noget nyt at en dansk almenordbog benytter autentiske citater som eksempler. Den hidtidige praksis synes at have været konstruerede, redaktionelle eksempler, evt. i form af bearbejdede autentiske eksempler. I litteraturen diskuteres det indgående om man skal foretrække redaktionelle eller autentiske eksempler. Tilhængerne af redaktionelle eksempler holder på at ordbogsredaktøren med sin store viden om sproget er den bedste til kort og klart at eksemplificere det pågældende opslagsords brug, uden at blande alt for mange forstyrrende elementer ind i billedet. Fortalerne for autentiske eksempler mener derimod at de konstruerede eksempler netop virker konstruerede, abstrakte og kunstige i forhold til sætninger der påviseligt er forekommet i faktisk sprogbrug og kan bringe en ny friskhed som de konstruerede og måske stereotype eksempler ikke har (Bergenholtz 1994). Ikke mindst britiske leksikografer (især folkene bag Cobuild) har gjort sig til talsmænd for kun at bruge autentiske eksempler, nu hvor disse er så let tilgængelige via elektroniske korpora.

Den Danske Ordbog, som altså også bygger på et korpus, indeholder også (stort set) kun autentiske eksempler i form af citater. Vi lægger os hermed i forlængelse af de eksisterende store modersmålsordbøger i Skandinavien (ODS, *Svenska Akademiens ordbok*, *Norsk Riksmålsordbok*), som favoriserer de autentiske eksempler.

Hvad er et godt citat?

I det følgende vil jeg beskæftige mig med hvordan et godt citat ser ud, og hvad der gør et citat mindre egnet til at stå i en ordbog.

Før man kan afgøre hvad et godt citat er, må man først overveje hvad citatets funktion er. Hvis citatet blot skal dokumentere at ordet findes i sproget på et bestemt tidspunkt, behøver man ikke stille særlig store krav til det. En forekomst som opslagsord i en ordbog, et leksikon eller en ordliste kan fx være tilstrækkeligt. Skal citatet derimod supplere eller illustrere de allerede givne oplysninger om ordet, dvs. bidrage yderligere til forståelsen af ordet og vise det i naturlig sammenhæng med andre ord, viser det sig hurtigt at langt de fleste autentiske ytringer ikke er velegnede.

Sven-Göran Malmgren har i en artikel i *LexicoNordica* (Malmgren 1994) opstillet fire kategorier af ordbogseksempler, inddelt efter hvilken type information der gives i dem. Disse fire kategorier er:

- (1) Semantiske eksempler.
- (2) Syntaktiske eksempler.
- (3) Konnotationelle eksempler.
- (4) Encyklopædiske eksempler.

Semantiske eksempler understøtter og supplerer betydningsbeskrivelsen ved at vise ordet sammen med betydningsmæssigt beslægtede ord, fx synonyme eller antonymer, eller sammen med ord som det typisk optræder sammen med (kollokationer), fx for et substantivs vedkommende de adjektiver der typisk beskriver det, og for et verbs vedkommende typiske subjekter og typiske objekter.

Syntaktiske eksempler viser de almindelige syntaktiske forbindelser som ordet indgår i, fx et verbum konstrueret med direkte objekt, eller forbundet med præpositioner og i givet fald hvilke. Ofte vil et givet eksempel både illustrere semantiske og syntaktiske forhold, fx hvis et transitivt verbum optræder med et typisk objekt: *Hun spiller klaver.*

Konnotationelle eksempler antyder typiske associationer, følelser og situationer som opslagsordet er knyttet til eller giver anledning til, fx ved ordet *fødselsdag*, hvor et godt eksempel vil kunne handle om fest og gaver.

Encyklopædiske eksempler giver faktuel, ikke-sproglig viden om det begreb som ordet refererer til, fx statistiske oplysninger om forekomsten af en sygdom.

Jeg ved ikke om denne rækkefølge er tilfældig fra Malmgrens side, men det stemmer med min egen opfattelse at et godt citat primært bør illustrere opslagsordets naturlige semantik og syntaks, mens konnotationer og encyklopædiske oplysninger kommer i anden række.

Når korpus kun består af autentisk sprog, burde det jo være let at finde eksempler på naturlig sprogbrug. I praksis viser det sig dog at være ret vanskeligt at udvælge gode citater. Jeg har forsøgt at inddele problemerne i en række kategorier som i høj grad modsvarer de krav jeg selv stiller til et godt citat, nemlig:

- (1) Længde.
- (2) Forstyrrende elementer.
- (3) Svære ord.
- (4) Manglende naturlighed med hensyn til sprog eller situation.
- (5) Fejl.
- (6) Tendentløse eller anstødelige udsagn.
- (7) Manglende underholdningsværdi.

Jeg vil i det følgende give konkrete eksempler på nogle af disse problemer, idet der dog ofte er tale om en kombination af flere forhold.

Længde

Det er en god gammel leksikografisk grundregel at man skal spare på pladsen. Eksempler er ingen undtagelse til denne regel. De fleste redaktionelle ordbogseksempler er korte, mens auten-

tiske eksempler har en tendens til at blive ganske lange, især hvis man (som vi på Den Danske Ordbog) tilstræber at alle citater indeholder et finit verbum. Man kan dog ofte forkorte ved at udelade passager der ikke er direkte relevante for opslagsordet, og erstatte disse med prikker. Vi tillader os også at resumere passager og erstatte dem med en kort omskrivning i firkantet parentes, men man kan hurtigt komme i konflikt med kravet om autenticitet. Hvor meget må man gøre ved citatet og stadigvæk påberåbe sig at det er autentisk? Og endelig hvad betyder det for læsbarheden at citaterne indeholder underlige tegn og gerninger i form af prikker og parenteser? For at dæmme op for dette har vi på Den Danske Ordbog ikke tilladt at kombinere disse forkortelsesmetoder, men alligevel ser vi ofte at ønsket om at holde længden nede giver en høj frekvens af prikker og parenteser. Eks. 1 viser hvordan man kan gøre vold på et belæg for at forkorte det, men uden at resultatet er specielt vellykket. Opslagsordet er *klarsyn*:

- (1) Problemet er blot at finde konkrete psykiske fænomener og dermed personer, som kan tillægges paranormale evner. Det vil sige evner, der ikke kan forklares ud fra kendte naturvidenskabelige principper. For eksempel tankelæsning (telepati), forudviden (prækognition) og klarsyn (clairvoyance) (Michael Leslie Clemmensen: *Er der mere mellem himmel og jord?*, 1992)

Ved hjælp af parenteser og prikker har man forsøgt at redigere denne lange smøre ned til:

- (1a) [paranormale evner] vil sige evner, der ikke kan forklares ud fra kendte naturvidenskabelige principper. For eksempel .. klarsyn (clairvoyance)

Citatet er ganske vist blevet meget kortere – til gengæld er der både brugt firkantede parenteser og prikker, foruden de runde parenteser som citatet i forvejen indeholder. Et sådant belæg er ikke muligt at redigere på en fair måde og egner sig ikke særlig til ordbogsbrug, også fordi det ikke primært handler om opslagsordet *klarsyn*, men blot nævner det i en opremsning sammen med andre paranormale evner.

Eks. 2 er både kortere, lettere tilgængeligt og frit for redaktionelle indgreb:

- (2) Guds profet Samuel var en sådan "seer", der i sit klarsyn kunne se, hvor bortløbne dyr befandt sig (Den Nye Dialog, 5/1992)

Forstyrrende elementer

Autentiske eksempler indeholder ofte mange oplysninger og elementer som er relevante i den konkrete kommunikationssituation hvor sætningen er blevet brugt. Det kan fx være navne på personer og steder, årstal, andre tal, og det kan være deiktiske udtryk og tekstmarkører af typen *her, så, desuden, derfor, således*. Lige så naturlige og nødvendige som den type oplysninger er i faktisk sprogbrug, lige så unaturlige og uvedkommende kan de forekomme når citaterne bruges, helt løsrevet fra deres kontekst, i en ordbog hvor der skal illustreres almene forhold vedrørende ordenes betydning og brug. Her et eksempel vedr. ordet *ordensregel*:

- (3) Ø C K og K O K er imidlertid indforstået med at udarbejde ordensregler, hvorefter medlemmerne skal parkere på parkeringspladsen nord for Svanholmskolen (Frederikssund Avis, 1983)

I princippet er citatet godt fordi det illustrerer hvad ordensregler kan omfatte (her regler for parkering), men det fungerer ikke så godt uden for kontekst fordi man ikke kan regne ud hvad Ø C K og K O K står for, og måske endda bliver lidt forvirret af det (der er tale om to sportsklubber).

Et bedre citat ville være det følgende, som stammer fra en lejekontrakt:

- (4) Lejeren skal have et eksemplar af ordensreglerne og skal følge disse regler (Labor Lejekontrakt, 1983)

Vi får – også her – sneget en væsentlig encyklopædisk oplysning ind, nemlig at der findes ordensregler i boligejendomme, til gengæld indeholder citatet ingen forstyrrende, uigennemskuelige

forkortelser, og endelig er det kun halvt så langt. I eks. 5 ser vi et andet tilfælde hvor elementer peger ud af teksten på en forvirrende måde (belæg på ordet *ulykkelig*):

- (5) Derfor var man langt ringere stillet end i dag, hvis man var så ulykkelig at blive ramt af sygdom (Sven Frøkjær-Jensen: *Segl og scepter – danskernes virkelighed*, 1991)

Uden for kontekst er det selvfølgelig umuligt at vide hvad *derfor* henviser til, og citatet er af den grund ikke velegnet i ordbogen.

Svære ord

Vi tilstræber på Den Danske Ordbog at det ordforråd og den syntaks der bruges i definitionerne, er så ligetil som muligt. Lidt anderledes kan det forholde sig med citaterne. Her er det jo sproget i "fri dressur" som vi afspejler, og derfor er det usandsynligt at alle ord og udtryk fra citaterne vil figurere som opslagsord i ordbogen. Ikke desto mindre bør citaterne ikke være for svære at tilegne sig da de ellers ikke vil fremme forståelsen af definitionerne, og i værste fald måske endda blokere for forståelsen. Ordet pesticid er – af gode grunde – blevet meget udbredt i samfundsdebatten, men det følgende belæg på ordet er for indviklet (og for langt) til virkelig at gøre gavn:

- (6) ved at overføre det toxinkodende gen, håber det store amerikanske frø- og kemikaliefirma Monsanto at have et mikrobielt pesticid til bekæmpelse af jordboende sommerfuglelarver (Morten Kvistgaard et al.: *Bioteknologi og miljø*, 1987)

Mange brugere vil nok have problemer med ordene *toxinkodende*, *mikrobiel* og *jordboende* sådan at opmærksomheden fjernes fra opslagsordet selv. Formentlig har man større glæde af følgende citat.

- (7) Kommissionen anbefaler, at forbruget af pesticider i landbruget formindskes (Statens Information: *Status*, 1986)

som ud over at være enklere og kortere også nævner landbruget, der jo er kendt som storforbruger af pesticider.

Manglende naturlighed med hensyn til sprog eller situation

Som redaktør vil man ofte være optaget af at finde et citat hvor flest mulige af Malmgrens fire kategorier kommer med (semantik, syntaks, konnotation og encyklopædi). Især kan man være fristet af citater hvor der gives encyklopædiske oplysninger, men her risikerer man at fjerne sig for langt fra den naturlige sprogbrug. Det er nemlig ret sjældent at man i almindelige samtale-situationer beskæftiger sig med hvordan diverse begreber helt præcist defineres, eller hvilken familie en bestemt blomst henregnes til. I følgende eksempel beskæftiger man sig fx med den præcise meteorologiske definition af fænomenet "hvid jul":

- (8) En hvid jul betyder, at mere end 90% af Danmark skal være dækket af mindst 1/2 cm sne den 24. december om eftermiddagen (AKB-beboeren, 4/1998)

Der er ingen tvivl om man her får megen encyklopædisk information, men til gengæld er det ikke et eksempel på naturlig brug af udtrykket *hvid jul*. I vores korpus beskæftiger man sig ganske vist i nogle få tilfælde med statistikken for hvid jul i Danmark i 1900-tallet (ca. hvert tiende år i øvrigt!), men mere udbredt er det at drømme om eller håbe på en *hvid jul*, så et mere karakteristisk citat ville få fat i udtrykkets konnotationer, nemlig den romantiske drøm om hvid jul:

- (9) tunge skyer hang lavt og varslede sne. Måske ville det endelig blive en hvid jul? (Hjemmet, 1988)

Et andet eksempel på en atypisk kontekst mødte jeg ved forkortelsen for *kilometer*: *km*. Hvis man bliver forblændet af trangen til at oplyse, fristes man måske til at bringe følgende citat som belæg på forkortelsen *km*:

- (10) I stedet for at sige 1000 m siger vi 1 kilometer og skriver 1 km (Mogens Nørgaard Olesen: *Alt er tal*, 1990)

Som citat er dette helt uegnet, dels fordi det slet ikke giver nogen supplerende oplysninger (artiklen oplyser jo naturligvis at *km* er forkortelse for *kilometer*), dels fordi den lærebogsagtige formulering (som netop også er fra en skolebog) er helt atypisk sprogbrug i forbindelse med denne forkortelse. Meget mere naturlige og typiske citater vil være som de følgende:

- (11) Jeg løb ca. 7 km på 45 min, hvilket jeg tror er nær en rekord (dagbog, 1992)

- (12) den tilladte hastighed er 50 km i timen (Jyllands-Posten, 1992)

Fejl

Det er et meget kontroversielt spørgsmål om ordbogseksempler må indeholde fejl. Hvis der kun bringes redaktionelle eksempler, bør problemet ikke melde sig da redaktøren jo bare kan lade være at lave fejl! Anderledes forholder det sig med autentiske eksempler. Det var fristende at sige at når der nu er tale om autentisk sprog, så afspejler man bare det ubearbejdede sprog som det foreligger, og det er som sådan ofte fejlbehæftet, og redaktionen står ikke inde for det. Denne holdning er imidlertid ikke gangbar, specielt når man betænker den normerende virkning og de mange læsere som en ordbog som Den Danske Ordbog formentlig vil få. Det store flertal af brugerne vil uden tvivl forvente at finde korrekt sprogbrug i eksemplerne, sprogbrug som de vil kunne efterligne uden at få problemer med dansklærere, arbejdsgivere eller andre. Det giver os store problemer som ordbogsredaktører fordi en væsentlig del af alle eksempler på et givet ord indeholder større eller mindre sproglige eller faktuelle fejl.

Disse fejl kan så være mere eller mindre graverende. For at overholde vores målsætning om at være deskriptive (men altså

også for at få citategnede eksempler nok at vælge imellem!) tillader vi bl.a. følgende typer fejl i citaterne:

- kommafejl,
- forkert sær- og sammenskrivning (*idag* over for det normerede *i dag*),
- forkerte forkortelser (*kr* uden punktum; *TV*, som normeres *tv*),
- diskutabel brug af bindestreg (*gen-transplantation*; *latin-prøve*),
- forkert brug af apostroffer (*Kirsten's computer*).

Disse typer fejl anser vi altså for at være under bagatelgrænsen og er ikke bange for at lade dem slippe igennem i citater.

Hvis der ellers er stavfejl (både i og uden for opslagsordet), må vi gå uden om det pågældende citat. Der er således utallige tilfælde hvor – ellers gode – citater må falde fordi de indeholder en stavfejl. Jeg skal nøjes med at give et enkelt eksempel, nemlig til ordet *indgroet* i overført betydning:

- (13) Alt for ofte lader vi os styre mere af indgroede vanemønstre end af egentlige refleksioner over vort liv og forhold til andre mennesker (Massage Meditation Yoga, Center for bevidsthed og vækst, 1991)

Refleksion staves på dansk med -s-, så citatet, der ellers på glimrende vis modstiller vanemønstre og refleksion, må droppes, måske til fordel for et ringere citat – uden fejl!

Med hensyn til fejltyper der ligger uden for *Retskrivningsordbogens* område, altså den ortografiske norm, er situationen lidt vanskeligere. Selvom man som redaktør ønsker at være deskriptiv, går man i sit valg af citater mere eller mindre bevidst efter citater helt uden fejl eller kritisabel sprogbrug. Ej heller er man særlig tilfreds med at bringe et citat der indeholder en oplysning som man positivt ved er forkert (fx et forkert årstal). Eks. 14 indeholder en almindelig, men uheldig konstruktion (belæg på ordet *vandføring*):

- (14) Med en vandføring på ti kubikmeter pr. sekund stejler den lille gummi-båd i vandfaldet (B.T., 1991)

Verbalsubstantivet *vandføring* knyttes her til et forkert logisk subjekt, idet det selvfølgelig er vandfaldet der har en vandføring, og ikke gummibåden. Til trods for citatets åbenlyse kvaliteter (korthed, angivelse af vandmængde, sammenhæng med vandfald) vil vi føle os nødsaget til at droppe det for ikke at gengive noget der regnes for dårligt sprog.

Tilsvarende forholder det sig med talesprogscitater. Med alle de fænomener der er karakteristiske for spontan tale: tøvefænomener, selvrettelser, gentagelser m.m., er udskrevet talesprog meget svært at håndtere og bringe som citat uden at det kommer til at se meget mærkeligt ud for nogen der ikke er vant til at se den slags. Det følgende eksempel stammer fra en udskrevet samtale med en ung pige (belæg på ordet *rideskole*):

- (15) ude på rideskolen hvor jeg rider der er både {pause} altså {pause} privat {pause} altså {tøven} privatheste, ikke? og så rideskolens (Projekt Bysociolingvistik, 1987)

Der er her klart tale om spontant talesprog som vi alle sammen kender det, men det er kun sprogfolk der er klar over at det faktisk er sådan vi taler. Et eksempel af denne type kræver skrap redigering for bare at nærme sig de "normale" skriftsprogscitater, og selv da vil de fleste brugere nok stejle over det og ikke få det tilsigtede udbytte.

Tendentiøse eller anstødelige udsagn

Som omtalt kan vi som ordbogsredaktører til en vis grad fralægge os ansvaret for mindre sproglige bommerter i citaterne (kommafejl, bindestregsfejl e.l.). Men kan vi tillade os at viderebringe tendentiøse eller direkte anstødelige eller fornærmende udsagn under henvisning til at vi jo bare citerer autentisk sprog der opfylder vores formål, nemlig at illustrere brugen af et givet opslagsord? Lad mig give et par eksempler. Det første, som om-

handler offentlige skoler over for privatskoler, er vel egentlig ganske ufarligt selvom nogle nok vil føle sig stødt (belæg på *kommuneskole*):

- (16) da de københavnske kommuneskoler ikke har ry for at være særlig gode . . . har vi foretrukket at lade dig gå et sted hvor der er mere system i tingene [: en privatskole] (dagbog, 1986)

men langt værre bliver det når man kommer over i de racistiske og på anden måde groft diskriminerende udtalelser (belæg på ordene *jødesvin* og *perker*¹):

- (17) Nu har vi dig, dit kommunistiske jødesvin! (Bo Torstensen: *Sammen vi står!*, 1988)

og

- (18) Jeg hader at se perkerne skumme fløden imedens danskerne dør på grund af ventetid [i] hospitalssystemet (Indlæg til Fælleslisten mod indvandringen, Internettet 1997)

Ifølge Malmgrens artikel vil han tilsyneladende gerne være med til at optage enkelte af den slags opslagsord, men mener at det er etisk uforvarsligt at bringe racistisk betonedede eksempler som de ovenstående. Jeg mener at man principielt ikke bør afstå fra at illustrere et ords brug selvom man personligt tager afstand fra ordet og måske også fra de mennesker der bruger det. Det der er vigtigt, er selvfølgelig at give en kraftig advarsel om ordets stærkt nedsættende valør og netop kun bringe et sådant citat i den artikel der har det pågældende ord som opslagsord.

Manglende underholdningsværdi

Ud over at være eksempel på naturlig sprogbrug må et citat godt bringe billeder af konkrete situationer frem i læserens bevidsthed. Citaterne kan være individualiseringer af de mere formaliserede oplysninger der er bragt tidligere i artiklen i form af definition,

¹ Ordet *perker* er en stærkt nedsættende betegnelse for en mørk indvandrer.

eventuelle synonymer, antonymer, valensoplysninger m.m. Derfor vil det kunne diskvalificere et citat hvis det slet og ret er for kedeligt. Et enkelt eksempel hvor et tørt og faktisk citat er skiftet ud med et mere levende (belæg på ordet *tørretumbler*):

- (19) Der kom forslag om opsætning af møntautomat til betaling af det øgede strømforbrug i forbindelse med tørretumbler (Referat fra beboermøde, over for 1986)
- (20) jeg ordner lige lidt vasketøj. Det gør jeg næsten hver dag. Skønt med tørretumbler, når det øser sådan ned (dagbog, 1992)

Karakteristisk nok stammer det sidste citat fra en meget vigtig kilde til ord fra dagligdagen, nemlig en lang række dagbøger der blev skrevet på Nationalmuseets foranledning den 2. sept. 1992. Herfra får vi et stort udvalg af ord, konstruktioner og situationer som stammer fra almindelige menneskers almindelige hverdag.

Intet citat

I mange tilfælde er vi ude for at der selv blandt mange belæg slet ingen er der egner sig til at blive bragt som citat. Det gælder fx for mange ord med meget konkret betydning at belæggene ofte er temmelig intetsigende og kun kan tjene til dokumentation. Måske indgår ordet mest i opremsninger som ikke siger meget om ordet selv, eller måske står det som opslagsord i et leksikon der er blandt kilderne i korpus. I så tilfælde kan man lige så godt undvære et citat, medmindre formålet er for enhver pris at dokumentere at ordet findes i ens korpus.

En gruppe ord hvor det især er vanskeligt at finde citater, er nationalitetsord (*nordmand, norsk*). Ofte, især ved små og fjerne nationer, er ordene knap nok repræsenteret i vores korpus; i andre tilfælde er der masser af eksempler at tage af, men i så fald er de fleste belæg enten fulde af forældede fordomme, eller også stammer de ensidigt fra krigs- eller sportsreportager. Her er man ofte nødt til at lade være med at bringe noget citat selvom ordets

forekomsttal absolut berettiger til det, eller bringe et farveløst og politisk korrekt citat som fx:

- (21) Alt tyder på, at der ikke bliver rokket ved regeringens aftale med svenskerne om Øresundsforbindelsen (B.T., 1991)

En maliciøs sjæl havde måske lyst til at drille broderfolket med citater af typen:

- (22) Fuld svensk vægtløfter mørbankede to politifolk (Berlingske Tidende, 1991)

men der er jo ingen grund til at sætte det gode naboskab over styr!

Afslutning

Jeg har i det foregående givet eksempler på hvor vanskeligt det er at finde citategnede belæg i vores store tekstkorpus. Det ideelle citat skal, som det er fremgået, imødekomme en række krav som jeg kort skal resumere

- det skal være tilpas kort;
- det må ikke indeholde forstyrrende elementer der tager opmærksomheden bort fra opslagsordet;
- det må ikke indeholde for svære ord;
- det skal være naturligt med hensyn til sprog og situation;
- det må ikke indeholde fejl;
- det bør kun indeholde anstødelige udsagn hvis man gør udtrykkeligt opmærksom på stilværdien;
- det må gerne være underholdende og levende.

Disse krav er svære at opfylde på en gang, og man må i mange tilfælde stille sig tilfreds med et mindre godt citat eller acceptere forstyrrende afkortninger, indskudte forklaringer e.l.

Hvorfor kan I dog ikke bare rette fejlene og skære citaterne til så de passer, uden at gøre rede for det? vil nogen måske spørge. Dertil er svaret: fordi vi lægger os i ODS-traditionen med hensyn til videnskabelighed. Vi citerer korrekt, og hvis vi ændrer noget i et citat, skal det fremgå. Desuden skal citaterne være forsynet med kildehenvisning så brugeren kan se hvor og hvornår den pågældende sætning blev skrevet eller sagt. Til gengæld må vi så leve med at der bliver lidt ”støj” i form af prikker og parenteser, men disse gener bliver i højeste grad opvejet af fordelene ved de autentiske eksempler, nemlig det nye og spændende som de kan tilføre ordbogen, modsat de redaktionelle eksempler.

Litteratur

- Bergenholtz, Henning. 1994. ”Beispiele in Fachwörterbüchern.”
Burkhard Schaedler/Henning Bergenholtz (Hrsg.): *Fachlexikographie. Fachwissen und seine Repräsentation in Wörterbüchern*. Tübingen: Gunter Narr Verlag. S. 321–439.
- Malmgren, Sven-Göran. 1994. ”Språkprovens form og funktion i svenska betydelseordböcker från Östergrens Nusvensk ordbok till Svensk ordbok.” *LexicoNordica* 1:107–117.

Nina Martola

Översättningsmodeller i en ordbok? En liten användarundersökning

In the new Finnish-Swedish bilingual dictionary published in 1997 there are models for how to translate certain types of Finnish words. The models are part of the front matter and from the entries there are references to the models. The word types in question are derivatives that have no equivalents at all in Swedish, or the equivalents of which cannot be constituents of the same type of syntactical constructions as the Finnish words. In order to find out whether this concept seems to work I have carried out user research where students were asked to translate sentences containing these types of constructions with the help of our models. A control group got another text about translating instead. The test persons were also asked to evaluate the models. Even if the differences between the two groups were not very big, the result was not entirely discouraging.

1. Bakgrund

I denna artikel kommer jag att redogöra för en användarundersökning som jag har utfört med ett antal svenskstuderande vid olika finländska universitet som testpersoner. Själva undersökningen genomfördes våren 1997 och analysen av resultaten våren 1999.

Det jag ville undersöka var hur väl de översättningsmodeller som finns med i *Stora finsk-svenska ordboken* (1997) (i fortsättningen SFSO) verkar fungera. Modellerna finns presenterade i artiklar i *LexicoNordica 2* och i *Nordiska studier i lexikografi 2*, så jag skall inte redogöra så ingående för dem här, men en kort presentation behövs säkert som bakgrund.

Modellerna är fyra till antalet, eller fem beroende på definition. I tre stycken behandlas olika typer av nominalfraser, i den fjärde kausativa verb och i den femte olika typer av tidsuttryck. Det är de tre modellerna för översättning av nominalfraser som jag har testat.

Det rör sig om finska nominalfraser som inte kan översättas ord för ord, eftersom den avledning som utgör huvudord aningen helt saknar motsvarigheter på lexikal nivå på svenska eller har ekvivalenter som är underkastade andra syntaktiska restriktioner än den finska avledningen. Som uppslagsord i en ordbok är dessa avledningar därför ytterst problematiska. I regel har detta hittills ignorerats i finsk-svenska ordböcker, men i SFSO finns alltså ett försök till lösning av problematiken i form av modellöversättningar för satser som innehåller de besvärliga ordtyperna. Typerna är egenskapssubstantiv, nomen actionis och nomen agentis. Modellerna finns både på svenska och finska och består i huvudsak av exempel på översättningar med kortfattade förklarande texter. (Se appendix.) Från alla lemman av denna typ finns det hänvisningar till modellerna. Just det aktuella ordet behöver inte finnas med bland modellöversättningarna, utan dessa är tänkta att fungera som typexempel som skall kunna tillämpas också vid översättningar av andra ord av samma typ. Hänvisningen kan antingen stå ensam, vid total avsaknad av ekvivalenter, eller ges utöver de ekvivalenter och/eller exempel som finns i själva artikeln. Sammanlagt finns det över 7 000 modellhänvisningar i ordboken

Vi var naturligtvis intresserade av att försöka få reda på hur det här försöket verkade ha utfallit, och det var därför jag beslöt genomföra en användarundersökning.

2. Testet

2.1. Testmaterialet

2.1.1. Översättningsuppgiften

Studenterna fick sammanlagt 30 meningar/meningsgrupper att översätta. Dessa innehöll problematiska konstruktioner som inte gick att översätta ordagrant; 10 var exempel på egenskaps-substantiv, 10 på nomen actionis (i testmaterialet och i ordboken rubricerat som *verbalsubstantiv som anger en handling*) och 10

på nomen agentis (*verbalsubstantiv som anger den handlande*). Meningarna var hämtade ur en finsk korpus och således autentiska. De representerade alla tidningstext. Vissa längre namn och personbetecknande nominalsubstantiv var utbytta ut mot *X* för korthetens skull. För varje sats att översätta fanns det med så mycket kontext som behövdes för att det skulle gå att åstadkomma en vettig översättning, men testpersonerna fick koncentrera sig på de problematiska partierna. Text som var med som stödjande kontext var återgiven i petit. Vissa ord som hade vållat pilotgruppen bekymmer och som inte hade relevans för själva testet gavs längst ner på sidan. De ord som hänförde sig till de ordtyper som testet gällde fick försökspersonerna leta efter i en stencil med uppkopierade artiklar ur den existerande stora finsk-svenska ordboken (FSS). Testet genomfördes innan SFSO kommit ut, vilket var viktigt så att försökspersonerna inte hade haft tillfälle att läsa modellerna.

Seija Tiisala, som är lektor vid Helsingfors universitet och undervisar i översättning, hjälpte till att göra det slutliga urvalet genom att ta ställning till de satser jag hade gallrat fram. Hon kommenterade också testblanketterna inför pilotundersökningen och har också i egenskap av ordförande för redaktionsrådet för SFSO kommenterat översättningsmodellerna vid åtskilliga tillfällen medan vi höll på med utarbetningen.

2.1.2. Stödmaterial

I varje testgrupp fick hälften av studenterna översättningsmodellerna ur SFSO som stödmaterial medan hälften fick ett annat material. Vid tiden för testet skulle SFSO just gå i tryck, så det var en så gott som färdig version av översättningsmodellerna som användes. Endast några smärre layoutjusteringar har gjorts efter det.

Jämförelsegruppen fick ett stödmaterial bestående av ett utdrag ur Rune Ingos bok *Från källspråk till målspråk*, som finns både på finska och svenska (Ingo 1990 resp. 1991). Avsnittet är en allmänt hållen text om översättning, som i och för sig kunde vara

nyttig men som inte var avsedd att ge lika handfast och detaljerad hjälp.

Åbostudenterna, som läste svenska som modersmål fick stödmaterial på svenska, de övriga på finska.

2.1.3. Testpersonerna

För det egentliga testet kontaktade jag svensklärare vid olika universitet och av de sju jag var i kontakt med hade fem grupper att erbjuda och lämpliga tillfällen att genomföra testet: Joensuu, Tammerfors, Uleåborgs och Vasa universitet samt Åbo Akademi. Antalet fördelade sig på följande sätt:

Joensuu	15
Tammerfors	32
Uleåborg	33
Vasa	21
Åbo	14
Sammanlagt	115

Två testpapper fick uteslutas, det ena för att endast deluppgift A var gjord, det andra för att det inte framgick vilket av stödmaterialen försökspersonen hade haft.

Alla utom Åbostudenterna läste svenska som främmande språk. Hur långt studenterna hade hunnit i sina svenskstudier varierade, men största delen hade ett eller ett par studieår bakom sig.

2.1.4. Enkät

Försökspersonerna fick också i uppgift att bedöma hur bra stödmaterial var enligt deras uppfattning. De fick poängsätta det för var och en av de tre testdelarna på en poängskala från 0 till 4. Avsikten var att se om de som bedömt stödmaterial som nyttigt

också verkade ha haft större nytta av det – m.a.o. om bra poäng för översättningen korrelerade med högre bedömningspoäng för stödmaterialet.

Men samtidigt skulle enkäten tjäna som en allmän finger-visning om hur modellerna har tagits emot med tanke på framtida reviderade upplagor av SFSO. Därför ombads försökspersonerna också kommentera stödmaterialet i ord.

2.2. Testproceduren

2.2.1. Pilotundersökning

Som förberedelse för det egentliga testet genomförde jag en pilotundersökning på svenskstuderande vid Helsingfors universitet. De testen, med tre olika grupper, övervakade jag själv, för att se vilka instruktioner som behövdes och om något verkade gå fel, detta för att kunna ge så utförliga instruktioner som möjligt till de lärare som skulle genomföra testet vid de andra universiteten. Pilotundersökningen gav vid handen att tiden, 90 minuter, blev alltför knapp för att testa också modellen för de kausativa verben, vilket hade varit den ursprungliga tanken, så i det egentliga testet utelämnades den. Översättningarna till vissa ord som vållat pilotgrupperna problem lade jag också in under respektive uppgift.

2.2.2. Det egentliga testet

Vid de egentliga testtillfällena var det lärare vid institutionerna som övervakade studenterna. Utifrån pilotundersökningen hade jag utarbetat detaljerade anvisningar till lärarna; vad de skulle säga och hur mycket tid de skulle ge för varje moment. I vänsterkanten på lärarpapperet angavs löpande minuter som ytterligare stöd.

I anvisningarna till studenterna, både i de anvisningar som fanns på testpappren och i de instruktioner som övervakarna

ombads ge, betonades det att det *inte* var viktigt att hinna med alla satser utan att det viktiga var att försökspersonerna faktiskt utnyttjade stödmaterialet.

3. Resultatanalys

3.1. Bedömningen av översättningarna

Poängsättningen av översättningarna genomförde jag utan att veta vilket stödmaterial svarspersonen haft till förfogande eller från vilket universitet han eller hon kom, detta för att inte omedvetet favorisera dem som haft SFSO-modellen som hjälp och också för att undvika att få förutfattade meningar om något visst universitet. Jag poängsatte endast översättningarna av de problematiska konstruktionerna. Riktig konstruktion gav 2 poäng även om ordvalet inte var helt korrekt. Hur väl eller illa resten av satsen var översatt beaktade jag inte alls. Var konstruktionen någonting ditåt gav det 1 poäng eller i vissa fall 1/2. Också påbörjade konstruktioner där man kunde se att tanken varit riktig gav 1 eller 1/2 poäng.

Från helhetssynpunkt kan bedömningen alltså vara tämligen orättvis; en försöksperson som skrev knagglig svenska men lyckades åstadkomma någon form av "modellenliga" översättningar kunde få bättre poäng än en som skrev relativt flytande men hade missat på de aktuella fraserna.

För varje deluppgift (A = egenskapssubstantiv, B = verbal-substantiv som anger en handling och C = verbalsubstantiv som anger den handlande) räknades de erhållna poängen samman och ställdes i relation till de möjliga poängen. Möjliga poäng var det poängtal försökspersonen hade kommit upp till om alla de nominalfrasöversättningar han eller hon hade försökt sig på hade blivit rätt (och alltså *inte* det maximala poängtal det gick att få totalt för en uppgift). Den slutliga poängen var kvoten av dessa två poängtal gånger tio:

$$\frac{X}{Y} \times 10$$

X = erhållna poäng
Y = möjliga poäng

Maximipoängen per deluppgift var alltså 10 oavsett hur många eller hur få satser försökspersonen hade försökt sig på. Det här poängsättningssystemet använde jag dels eftersom de försökspersoner som faktiskt använde sig av modellerna på allvar kan antas ha hunnit med färre satser än de som satt och skrev rakt av, dels för att det ju inte var försökspersonernas allmänna svensk-kunskaper jag ville testa utan uttryckligen vad modellen kunde tänkas ge.

Först när denna poängsättning var gjord för samtliga svar kontrollerade jag vilket av stödmaterialet försökspersonerna hade haft och hur de hade bedömt det.

Av de 115 inlämnade testpapperen var det två som fick gallras bort direkt. En försöksperson hade inte lämnat in stödmaterialet tillsammans med testsvaren, så jag kunde alltså inte veta vilket-dera hon/han haft. En hade gjort bara deluppgift A. Alltså återstod 113.

När jag undersökte om försökspersonernas syn på modellerna korrelerade med deras prestationer fick jag utelämnat betydligt fler (27) testpapper. Det visade sig nämligen att flera av dem inte förstått att det var bara "översättningshjälpen", d.v.s. SFSO:s modellöversättningar respektive Ingos text de skulle bedöma. En del hade bedömt hur nyttig övning översättningsuppgiften i stort hade varit, många uttalade sig om stencilen med de uppkopierade ordboksartiklarna ur FSS och en del uttalade sig i huvudsak om de meningar de hade i uppdrag att översätta. Det här framgick av de omdömen i ord som försökspersonerna ombetts skriva. Vissa hade inte alls kommenterat i ord och för säkerhets skull uteslöt jag då också de svaren, eftersom jag inte kunde veta vad de egentligen poängsatt.

Stödmaterialet borde ha rubricerats tydligt med samma rubrik för översättningsmodellerna och den andra texten, så att rubriken kunde ha använts när man bad försökspersonerna bedöma stöd-

materialet i stället för att tala om *stencilen om översättning* eller *översättningsstencilen*.

3.2. Testresultat

Hur utföll då testet? Hade försökspersonerna haft nytta av översättningsmodellerna. Och vad hade de tyckt om dem?

3.2.1. Resultatjämförelse

På den poängskala från 0 till 10 som jag använde hade de som hade haft översättningsmodellerna ur SFSO ungefär en poäng mer i genomsnitt än de som haft Ingos text till hjälp för deluppgifterna A och B. För C var skillnaden mindre:

TABELL 1. *Genomsnittspoäng för testgrupp och kontrollgrupp (113 testpersoner).*

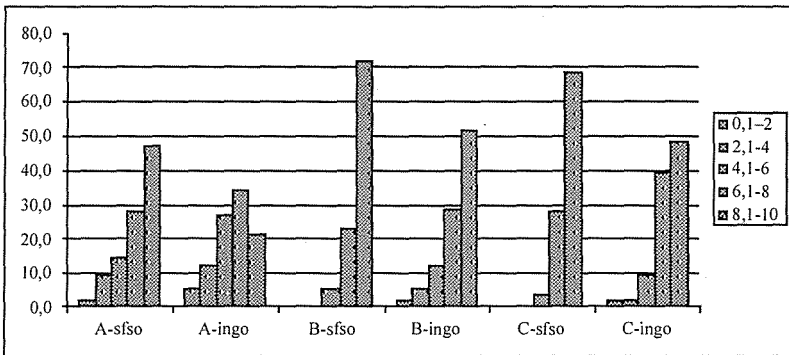
Mod.	SFSO	Ingo	skilln.
A	7,3	6,1	1,2
B	8,7	7,5	1,1
C	8,3	7,7	0,6
Antal	57	56	113

Skillnaden i genomsnittspoäng är alltså inte särskilt stor, men ser man på poängfördelningen verkar skillnaderna något större.

Skillnaden blir tydligare i grafisk framställning. Medan andelen högsta poäng (grå staplar med vita prickar) är stor eller mycket stor för dem som haft SFSO:s översättningsmodeller som hjälp är poängfördelningen jämnare för dem som haft Ingos text.

TABELL 2. *Poängfördelningen.*

Poäng	Modell A				Modell B				Modell C			
	SFSO		Ingo		SFSO		Ingo		SFSO		Ingo	
	Antal	%	Antal	%	Antal	%	Antal	%	Antal	%	Antal	%
0,0-2	1	1,8	3	5,4	0	0,0	1	1,8	0	0,0	1	1,8
2,1-4	5	8,8	7	12,5	0	0,0	3	5,4	0	0,0	1	1,8
4,1-6	8	14,0	15	26,8	3	5,3	7	12,5	2	3,5	5	8,9
6,1-8	17	28,1	19	33,9	13	22,8	16	28,6	16	28,1	22	39,3
8,1-10	26	47,4	12	21,4	41	71,9	29	51,8	39	68,4	27	48,2
	57	100	56	100	57	100	56	100	57	100	56	100

FIGUR 1. *Poängfördelningen i procent.*

Professor Timo Mäkeläinen vid Helsingfors universitets matematiska institution har haft vänligheten att göra en statistisk analys av resultatet och konstaterar att SFSO utifrån den verkar vara det bästa hjälpmedlet. Det bästa utslaget ger modell A.

	A	B	C
Wilcoxon	0.0051794	0.017614	0.019562
Permutationstest	0.0074791	0.0078863	0.012129

Ett tredje tal skulle dock behövas för att säkerställa signifikansen.

3.2.2. Studenternas poängsättning av modellerna

Den andra avsikten med undersökningen var alltså att ta reda på vad försökspersonerna hade ansett om översättningsmodellerna som hjälpmedel och om det fanns något samband mellan om-döme och resultat.

Studenterna fick i uppdrag att bedöma hur pass bra stödmaterialet var enligt deras uppfattning. De fick bedöma det på en poängskala mellan 0 och 4. SFSO-modellerna fick i genomsnitt drygt 2 poäng, vilket var 0,7–1,1 poäng mer än Ingos text.

I denna jämförelse kunde, såsom ovan konstaterats, endast 86 testpapper användas, eftersom många av försökspersonerna bedömt också annat än SFSO-modellerna resp. Ingos text. Konstateras kan i alla fall att två tredjedelar av dem som inte insett vad de skulle bedöma hörde till jämförelsegruppen, 18 stycken mot 9 för dem som hade haft översättningsmodellerna ur SFSO.

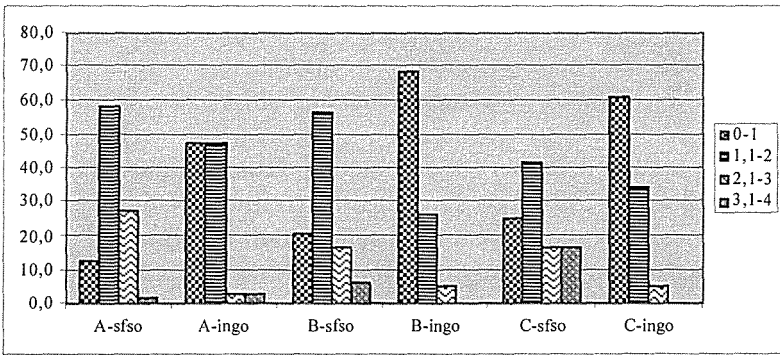
TABELL 3. *Genomsnittspoäng för stödmaterialet (skala 0–4, 86 testpersoner).*

Mod.	SFSO	Ingo	skilln.
A	2,2	1,5	0,7
B	2,1	1,1	1,0
C	2,2	1,1	1,1
Antal	48	38	86

Även vad gäller försökspersonernas bedömning av stödmaterialet blir skillnaderna tydligare om man ser på hur poängen fördelar sig (tabell 4). I figur 2 kan man se att det för SFSO är staplarna för 1,1–2 poäng (de tvärstrimmiga) som dominerar. För Ingos del är staplarna för 0–1 poäng högst (de schackrutiga) utom för modell A där båda är lika höga.

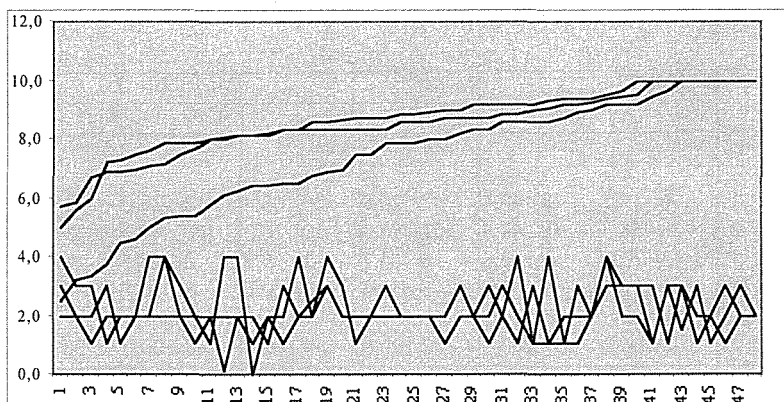
TABELL 4. *Poängfördelningen.*

Poäng	Modell A				Modell B				Modell C			
	SFSO		Ingo		SFSO		Ingo		SFSO		Ingo	
	Antal	%	Antal	%	Antal	%	Antal	%	Antal	%	Antal	%
0,0-1	6	12,5	18	47,4	10	20,8	26	68,4	12	25,0	23	60,5
1,1-2	28	58,3	18	47,4	27	56,3	10	26,3	20	41,7	13	34,2
2,1-3	13	27,1	1	2,6	8	16,7	2	5,3	8	16,7	2	5,3
3,1-4	1	2,1	1	2,6	3	6,3	0	0,0	8	16,7	0	0,0
	48	100,0	38	100,0	48	100,0	38	100,0	48	100,0	38	100,0

FIGUR 2. *Poängfördelningen i procent.*

3.2.3. Korrelation resultat – omdöme

Finns det då någon korrelation mellan översättningsresultat och omdöme? Det vill säga, har de som tyckt att översättningsmodellerna ur SFSO är bra faktiskt också haft mer nytta av dem i praktiken och alltså klarat översättningen bättre än de som inte tyckte modellerna var särskilt nyttiga? Svaret är nej.



FIGUR 3. *Korrelation omdöme – resultat.*
(SFSO – 48 testpersoner).

Medan resultatpoängen för de olika modellerna långsamt stiger från 2,5 (A) – 5,0 (B) – 5,7 (C) till maximala 10 poäng så går omdömeskurvorna mellan minimum 0 och maximum 4 poäng utan någon som helst stigande tendens. Här har alltså resultatpoängen sorterats från lägsta till högsta för var och en av modellerna medan omdömespoängen följt med den resultatpoäng den hörde till.

3.3. Studenternas kommentarer

Med tanke på framtida revideringar av översättningsmodellerna ombads försökspersonerna också motivera sin poängsättning. Kommentarererna kommer nog att vara till nytta, även om de till många delar är motsägelsefulla, vilket i och för sig inte var oväntat. Så brukar det ju bli när man ber en större mängd personer komma med synpunkter på någonting.

I tabellen nedan har kommentarererna fördelats på olika temagrupper och indelats i positiva, negativa och neutrala. Det är självfallet avsevärda gradskillnader mellan de olika kommentarererna av samma slag, alltifrån ”ingen nytta” till ”gav inte hjälp

med alla satserna”, och gränsen mellan t.ex. neutral och positiv resp. negativ är inte heller entydig. Få att få något slag av allmän översikt har jag i alla fall försökt mig på att göra en sammanfattning i tabellform, där jag räknat samman antalet positiva, negativa och neutrala kommentarer per modell.

TABELL 5. *Studenternas kommentarer till översättningsmodellerna (48 testpersoner).*

		A	B	C	
Använde den (just) inte		3	6	3	
Kunde det (mesta)		3	9	5	
För litet tid		6	4	5	
	+/-				
Exemplen	+	17	12	11	
	∅	2	2	2	
	-	7	6	3	
Allmännytt	+	19	15	16	
	∅	11	4	10	
	-	3	8	7	
Omfattning	+			1	
	∅		1		
	-	3	1	2	
Åskådlighet	+	8	7	8	
	∅	1	2	1	
	-	14	10	6	
Övrigt	+		1	2	
	∅	1	1	1	
	-	3		2	
Totalt:					
Positiva kommentarer		+	44	35	38
Neutrala kommentarer		∅	15	10	14
Negativa kommentarer		-	30	25	20
		S:a	89	70	72
%		+	49,4	50,0	52,8
%		-	33,7	35,7	27,8

Rent allmänt kan man konstatera att de positiva omdömena överväger. I runda tal är hälften av kommentarerna positiva och

en tredjedel negativa. Bäst klarar sig modell C, nomen agentis, vilket kan tänkas bero det på att översättningsuppgiften för den delen inte var lika krånglig som för de två första och att modellen därför var lättare att utnyttja.

Särskilt till exemplen och till den allmänna nyttan av modellerna ställer sig försökspersonerna positiva. I gruppen *allmännytt*a ingår utöver "nyttig" kommentarer som "bra", "mångsidig", "påminde om vissa saker", "pekade på olika alternativ" o.d. Vad gäller åskådligheten går åsikterna mest isär. Å ena sidan finns det många kommentarer i stil med "tydlig", "bra gruppering", "systematisk" o.d. men å andra sidan också "svårbegriplig", "svår att komma underfund med", "svår att tillämpa" o.d. Några försökspersoner säger att de inte läste den förklarande texten utan bara såg på exemplen, men i gengäld efterlyser några mer förklarande text. Några påpekade att exempelsatserna i modellerna var för enkla i relation till satserna i översättningsuppgiften.

Det som också går att utläsa ur tabellen är att modell B, nomen actionis, i högre grad än de övriga verkar att ha innehållit bekanta saker. Tydligt ägnas den problematiken mer uppmärksamhet i undervisningen än egenskapssubstantiv och nomen agentis. Det är säkert också en bidragande orsak till att den deluppgiften gick något bättre än de två andra.

4. Slutsatser

Vad skall man då dra för slutsatser av det här?

En aning bättre än jämförelsegruppen verkar de försökspersoner som haft översättningsmodellerna ur SFSO att ha klarat översättningen, och modellerna har också uppfattats som ett något bättre stödmaterial än Ingos text. Någon korrelation mellan positivt omdöme och bra översättningsresultat finns däremot inte. Synpunkterna på modellerna är övervägande positiva även om de kritiska kommentarerna också är förhållandevis många.

Sammanfattningsvis vågar jag i alla fall påstå, att idén med modellöversättningar i ordboken är värd att bygga vidare på. Det

skulle gälla att ta fasta på den kritik som kom fram i omdömena och utifrån den försöka förbättra modellerna på de punkter där de fick kritik.

Kanske vore det också bra att utveckla modellerna mer i riktning mot kompletterande stödmaterial på så vis, att man i varje enskild ordartikel för de ifrågasvarande ordtyperna ger åtminstone ett eller ett par exempel och därutöver gör en hänvisning till modellerna för den som behöver få veta mer. Problemet är naturligtvis att hitta naturliga finska exempel som också låter sig översättas till idiomatisk svenska. Det var ju just det problemet som var skälet till att modeller av den här typen infördes i ordboken. Men med hjälp av korpusar låter det sig kanske göras i en viss utsträckning. När den första upplagan av SFSO utarbetades fanns det en finsk korpus med några fattiga miljoner ord. Dessutom var den inte tillgänglig för redaktionen förrän under den senare delen av redigeringsarbetet och då inte för hela redaktionen. I och med att det nu föreligger en finsk textbank på drygt 40 miljoner ord (som dessutom hela tiden växer) och en stor mängd annat material i elektroniskt sökbar form blir det lättare att få fram exempel både för själva ordboksartiklarna och för översättningsmodellerna.

Litteratur

- FSS = Cannelin, Knut & Aulis, Hirvensalo, Lauri & Hedlund, Nils. [1976] 1986. *Suomi–ruotsi suursanakirja / Finsk–svensk storordbok*. Borgå: WSOY.
- Ingo, Rune. 1990. *Lähtökielestä kohdekieleen. Johdatusta käännöstieteeseen*. Porvoo-Helsinki-Juva: WSOY.
- Ingo, Rune. 1991. *Från källspråk till målspråk. Introduktion i översättningsvetenskap*. Lund: Studentlitteratur.
- Martola, Nina. 1994. "Översättarhjälp i en ordbok." *Nordiske studier i leksikografi* 2:185–192.
- Martola, Nina. 1995. "Substantiv avledda av adjektiv och verb – en jämförelse mellan några ordböcker med finska som källspråk." *LexicoNordica* 2:89–108.

SFSO = *Suuri suomi–ruotsi-sanakirja / Stora finsk–svenska ordboken*. 1997. Utgivare: Werner Söderström Osakeyhtiö och Forskningscentralen för de inhemska språken. Helsinki-Juva: WSOY.

Suuri suomi–ruotsi-sanakirja. Stora finsk–svenska ordboken. CD-ROM Win/Mac. 1996–98. Publ: Werner Söderström Osakeyhtiö och Forskningscentralen för de inhemska språken. Progr. WSOY – Timehouse/Teklacap Oy 1998.

Appendix: Utdrag ur översättningsmodell A

(Hämtat ur den elektroniska utgåvan av SFSO)

1.1 Egenskapssubstantiv med genitivattribut (eller possessivt suffix)

Ett egenskapssubstantiv med ett framförställt genitivattribut är en vanlig kombination på finska. På svenska motsvaras den oftast av andra konstruktioner.

1.1.1 Nominalfras där substantiveringen motsvaras av ett adjektivattribut på svenska

Hän kehui koiransa viisautta.

Han skröt med sin kloka hund.

Voitaisiin harkita jäsenäänestysten pakollisuutta.

Man kunde överväga obligatoriska medlemsomröstningar.

Den första meningen är ett exempel på att man på svenska undviker flera genitiver efter varandra, om det inte är en person som är egenskapsbäraren. Man kan alltså tala om *hans hustrus klokh*et men ogärna om **hans hunds klokh*et och än mindre om **hans bils snabbhet*. Däremot går *hundens klokh*et bra.

1.1.2 Infinitivfras

Ofta motsvaras den finska nominalfrasen av en svensk infinitivfras, vars verbdel inte har någon representation i den finska texten. Infinitivformen föregås vanligen av *att* eller av *genom att*, *för att* m.fl.

Voitaisiin harkita jäsenäänestysten pakollisuutta.

Man kunde överväga att göra medlemsomröstningarna obligatoriska.

... att införa obligatorisk medlemsomröstning.

Hän ärsyttää monia epäsovinnaisuudellaan.

Han förargar många genom att vara så okonventionell.

1.1.3 Att-sats

Mycket vanligt är att egenskapssubstantiv föregångna av genitiv motsvaras av en att-sats på svenska. Ett lämpligt predikatsverb tillkommer, vars form styrs av huvudsatsen.

Hän vastusti jäsenäänestysten pakollisuutta.

Han tyckte inte att medlemsomröstningarna skulle vara obligatoriska.

Asian kiireellisyyden takia pidettiin ylimääräinen kokous.

På grund av att ärendet var brådskande hölls ett extra sammanträde.

Hän ärsyttää monia epäsovinnaisuudellaan.

Det förargar många att han är så okonventionell.

1.1.4 Andra bisatser

Hän kehui koiransa viisautta.

Han skröt med hur klok hans hund var.

Asian kiireellisyyden takia pidettiin ylimääräinen kokous.

Eftersom ärendet var brådskande hölls ett extra sammanträde.

Vaarallisuutensa takia uusi malli joutui myyntikieltoon.

Den nya modellen förbjöds eftersom (då) den visade sig vara farlig.

Eftersom den nya modellen visade sig vara farlig förbjöds den.

Hän ärsyttää monia epäsovinnaisuudellaan.

Han förargar många eftersom (på grund av att) han är så okonventionell.

1.1.5 Ett lämpligt ord sätts in

Asian kiireellisyyden takia pidettiin ylimääräinen kokous.

På grund av ärendets brådskande natur hölls ett extra sammanträde.

Rahamarkkinoiden kireys ...

Det strama läget på penningmarknaden ...

Hän ärsyttää monia epäsovinnaisuudellaan.

Han förargar många med sitt okonventionella sätt.

Valensinformasjon i NorKompLeks

This article addresses how valency (or argument structure) information is described in the *NorKompLeks* lexicon for Norwegian. We discuss the relation between modern syntactic theories and a practical description of all verbs and their valency properties in a language. We conclude that the lexical description could use an intuitive notion of constructions in order to cover the relevant properties of the verbs, and linguistic theories should be able to interpret these lexical properties. A number of examples from the material are discussed. The relevance for lexicography is that dictionaries could use simplified versions of the valency information, or that alternative publication formats will make it possible to include such information in future dictionaries.

1. Innledning

Denne artikkelen er et bidrag til hvordan valensinformasjon, også kalt argumentstruktur, kan beskrives i norsk på en nokså teorinøytral måte og hvordan slik informasjon kan integreres i ordbøker. Utgangspunktet er prosjektet *Norsk komputasjonelt leksikon* (NorKompLeks) og spesielt dette prosjektets beskrivelse av valens.¹

2. Argumentstruktur – generelle betraktninger

Det er en vanlig oppfatning blant språkforskere at verb kan klassifiseres som intransitive, transitive, ditransitive ("dobbeltransitive"); osv. Men forskerne er ikke enige om hvordan disse termene skal gis en teoretisk fortolkning. I *Nordisk leksikografisk ordbok* defineres valens som "hvilke typer komplement som er knyttet til et ord, oftest et verb". Denne definisjonen lar fortolk-

¹ Det praktiske arbeidet med argumentstruktur i NorKompLeks er i hovedsak utført av Bodil Aurstad og Eli Sætherø Andenes.

ningen av "komplement" stå åpen og flytter på den måten forklaringsoppgaven til et annet nivå av den grammatiske beskrivelsen.

Man kommer ikke langt i klassifiseringen av valens, eller argumentstruktur, uten å spesifisere hva man skal forstå med komplement eller argument, dvs. det som valensen "viser til". Den syntaktiske definisjonen er *setningsledd*, noe som i syntaksteorier kan være "potensiell posisjon med tematisk rolle" som i styrings- og bindingsteori og minimalisme, jf. Chomsky (1981, 1995), "syntaktisk funksjon" som i leksikalsk-funksjonell grammatikk, se Bresnan & Kaplan (1982/1995), eller "elementer i bestemte posisjoner i en subkategoriseringsliste" slik man gjør i hodedrevet frasestrukturgrammatikk, jf. Pollard & Sag (1987, 1995). Alt dette er sterkt teorifargede definisjoner, men de dekker likevel mye av det som språkforskere ønsker å referere til, nemlig hvilken syntaktisk form eller kategori som argumentet skal ha og hvordan argumentet er ordnet i forhold til andre argumenter.

Den rent syntaktiske definisjonen av et argument tilfredsstiller ikke språkforskeren. Man ønsker også informasjon om hvordan de syntaktiske argumenter er koblet til semantiske argumenter. Noen verb har agentive subjekter (f.eks. *sparke*, *løpe*), andre tar subjekter med "mottaker-rolle" (f.eks. *arve*, *motta*), atter andre tar tematisk subjekt og objekt med "opplever-rolle" (f.eks. *irritere*, *berolige*). I argumentstrukturen må syntaktisk argument knyttes til semantisk argument. Igjen vil vi se at lingvistiske teorier foretar slike koblinger på ulikt vis, og det bør bemerkes at det er ingen enighet blant språkforskere om hvilke semantiske argumenter man bør anta. Men det er likevel enighet om at visse semantiske distinksjoner må gjøres i argumentstrukturen fordi verbene oppfører seg ulikt med hensyn til visse grammatiske prosesser, f.eks. at verb av typen *irritere* og *berolige* ikke kan brukes i upersonlig passiv, jf. ugrammatikaliteten til setninger som (1) og (2), mens verb av type *arve* og *motta* kan brukes med denne typen passiv, jf. (3) og (4):

- (1) *Det blir beroliget en student.
- (2) *Det blir irritert en professor.
- (3) Det blir arvet mange penger.
- (4) Det blir mottatt enkelte tvilsomme gaver.

I styrings- og bindingsteorien blir det antatt at syntaktiske argumenter, forstått som posisjoner i trestrukturer, kobles til "abstrakte" tematiske roller i leksikon. Man tilstreber maksimalt generelle koblinger som abstraherer bort "irrelevante" detaljer (som f.eks. det å spesifisere hvilke tematiske roller som lingvistisk teori har behov for). En uheldig konsekvens av denne form for abstraksjon er at få eller ingen er opptatt av å spesifisere koblingene i detalj, og det eksisterer ingen detaljert teori om leksikalske koblinger innen nyere transformasjonell grammatikk som forskerne innen disse rammeverkene kan enes om.

I leksikalsk-funksjonell grammatikk kobles syntaktiske funksjoner til tematiske roller i leksikon dels vha. en lenkings-teori og dels ved at spesialtilfeller listes opp. Resultatet av disse koblingene er f.eks. at subjekt-funksjonen kobles til AGENT-rollen i verb som tar agentive subjekter, og for kontrollverb vil subjekt-funksjonen i infinitivsfrasen kobles til subjektet (eller objektet) i oversetningen. Til forskjell fra styrings- og bindings-teorien, er koblingene spesifisert i mer detalj.

Hodedrevet frasestrukturgrammatikk har en strategi som ligner på leksikalsk-funksjonell grammatikk idet posisjoner i argumentstrukturlister relateres til semantiske roller, jf. Manning & Sag (1999). Til forskjell fra leksikalsk-funksjonell grammatikk har hodedrevet frasestrukturgrammatikk klarere anknytningspunkter til situasjonssemantikk, men denne dimensjonen av hodedrevet frasestrukturgrammatikk lar seg også inkorporere i leksikalsk-funksjonell grammatikk. I likhet med leksikalsk-funksjonell grammatikk, er de leksikalske koblingene spesifisert i rimelig detalj.

Generelt ser vi at ulike teorier implementerer leksikalske koblinger på forskjellig vis. Lister, funksjoner, syntaktiske frasekategorier brukes av noen, men ikke alle. Det er viktig å være klar over at ingen av teoriene er i nærheten av omfattende leksikalske

beskrivelser av noe naturlig språk. En konsekvens av dette faktum er at teoriens prediksjoner på ingen måte er testet ut i full skala, og en empirisk orientert beskrivelse av norsk bør ikke stole på kun en av dem – antakelig ikke på noen.

Det er vanskelig å se for seg at en lingvistisk teori om argumentstruktur kan anvendes på et språks totale vokabular. Til det har lingvistiske teorier for mye å si om hvert enkelt leksem, noe som i praksis er uforenlig med finansielle prosjektbegrensninger. Dessuten er språkvitenskapelig teori i konstant endring, og det betyr at sterkt teori-impregnerte beskrivelser kan vise seg å være av liten varig verdi. Men lingvistisk teori har gitt viktige bidrag til forståelsen av hva argumentstruktur er, og det er derfor ønskelig å integrere slik informasjon i valensbeskrivelser. Løsningen blir at man anviser "løse" beskrivelser som er innenfor rammene av hva de fleste (helst alle) moderne lingvistiske teorier sier om argumentstruktur, og man beskriver deretter alle verb innenfor disse rammene. Dermed blir spørsmålet hva man skal forstå med løse beskrivelser.

Det har alltid vært en ikke uttalt sammenheng mellom konstruksjonstyper og verb som tar del i dem (i denne artikkelen forstås konstruksjonstyper på en mer intuitiv måte enn i Goldberg 1995, men der er likhetstrekk). Det er bare et fåtall verb som kan ta del i "dobbelt objekt-konstruksjonen", og mengden av verb som kan inngå i "kausativ-konstruksjoner" er enda mindre. De fleste verb er transitive på en eller annen måte og kan derfor brukes i "transitivkonstruksjoner". Men de fleste moderne lingvistiske teorier ønsker ikke å referere til "konstruksjonsspesifikke egenskaper", noe som er en arv fra avvísningen av konstruksjonsspesifikke transformasjoner til fordel for et fåtall transformasjoner og generelle begrensninger på dem. Sagt på en annen måte, man ønsker å abstrahere bort konstruksjonstyper, som gjerne oppfattes som "stipulative", og heller bruke mer generelle teoretiske begreper som kan brukes til å utlede de egenskapene som gripes av konstruksjonstyper. Ulempen med denne tilnærmingen er, som nevnt ovenfor, at den i og for seg kan være teoretisk interessant, men den er ikke forsøkt verifisert mot et språks totale vokabular.

Men selv om språkvitenskapelig teori anser konstruksjonstyper som "epi-fenomener", dvs. sekundærobservasjoner som kan forklares på annet vis, er det ingenting i veien for at konstruksjonstypene kan listes opp i et leksikon, og teoriene overlates oppgaven å fortolke dem, altså vise at de faktisk er epi-fenomener. Fordelen med denne tilnærmingen er at vi kan bruke nokså teorinøytrale termer i oppstillingene av argumentstrukturinformasjonen, samtidig som vi kan utnytte ulike teories beskrivelsesapparat til å formalisere denne informasjonen.

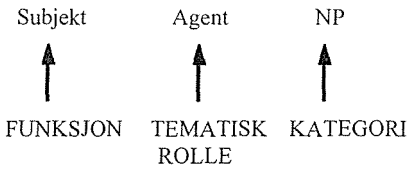
3. Argumentstruktur i NorKompLeks

I prosjektet *Norsk komputasjonelt leksikon* (NorKompLeks) har man forsøkt å beskrive hvilke konstruksjonstyper norske verb kan inngå i. Beskrivelsene antas å være intuitivt forståelige for alle språkforskere med en viss formell trening.

La oss se nærmere på ontologien som er benyttet. Hvert syntaktisk argument markeres for hvilken syntaktisk formkategori det skal ha, dvs nominalfrase (NP), preposisjonsfrase (PP), osv. Dette er motivert i tradisjonell grammatikk, transformasjonsgrammatikk og hodedrevet frasestrukturgrammatikk. I tillegg brukes syntaktiske funksjoner som subjekt og objekt, noe som er vanlig i leksikalsk-funksjonell grammatikk og delvis i hodedrevet frasestrukturgrammatikk. Den tredje dimensjonen er tematiske roller (agent, recipient, osv.), som i større eller mindre grad brukes i all generativ grammatikk. Denne ontologien er sårbar på to sentrale områder, nemlig klassifikasjonen av tematiske roller og syntaktiske funksjoner: det eksisterer ingen alment akseptert enighet om inventaret av syntaktiske funksjoner og tematiske roller. Derfor måtte prosjektet foreta en del vanskelige og til dels kontroversielle valg med hensyn til syntaktiske funksjoner og tematiske roller, noe vi kommer tilbake til nedenfor.

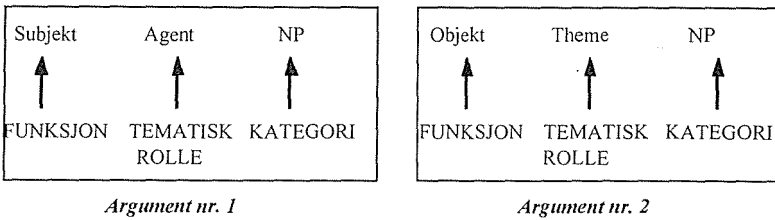
Hvert argument i beskrivelsene er en trippel bestående av syntaktisk funksjon, tematisk rolle og syntaktisk kategori. Et eksempel er:

(5)



En argumentstruktur er en sekvens av slike tripler, som i (6):

(6)



Legg merke til at sekvensen er eksplisitt ordnet, angitt ved argument nummer 1 etterfulgt av argument 2. Strukturen som er beskrevet i figur (6) er en standard transitivkonstruksjon, og den skal knyttes til verb som kan opptre i setninger som

(7) Bjørn Dæhlie slår alle svensker.

Fortolkningen av (6) er en konstruksjonstype med agentivt subjekt, realisert som nominalfrase, samt et tematisk objekt som også er en nominalfrase. Den vanlige intransitivkonstruksjonen er som i figur (5), dvs. at et agentivt NP-subjekt er eneste obligatoriske argument.

I NorKompLeks-basen er hvert verb utstyrt med en liste av de konstruksjonstypene de kan ta del i. Verbet *kondemnere* har listen [trans1], dvs. at det er et verb som bare kan brukes transitivt. Et annet eksempel er *punktere*, som har realisasjonsmulighetene [intrans2,trans1]. Vi kommer tilbake til typen intrans2 nedenfor.

4. Argumentstrukturtyper i NorKompLeks

La oss se litt nærmere på ulike argumentstrukturtyper i NorKompLeks. De enkleste er de intransitive typene, innebefattet to konstruksjoner med ekspletivsubjekt (dette er formalisert på en annen måte i NorKompLeks, men for å gjøre lesningen enklere, er beskrivelsene i denne artikkelen nokså uformelle. I appendikset er de faktiske beskrivelsene i NorKompLeks listet opp, og en mer formalisert beskrivelse av innholdet i NorKompLeks finnes i Nordgård 1998):

(8) Ekspletivkonstruksjon #1:

Subjekt	Ingen_rolle	NP
---------	-------------	----

Argument 1

(9) Ekspletivkonstruksjon #2:

Subjekt	Ingen_rolle	NP
---------	-------------	----

Argument nr. 1

Adv	Lokativ	PP
-----	---------	----

Argument nr. 2

Eksempler på den første typen er setninger som i (10), mens ekspletiver med lokativt argument er vist i (11):

- (10) a. Det snør.
 b. Det buldrer.
 c. Det klarner.

- (11) a. Det kvakk i henne.
 b. Det kribler i beina.

Forskjellen mellom disse to argumentstrukturtypene er at den siste har et obligatorisk lokativt argument. Det betyr at denne varianten er mer restriktiv enn den første, som ikke spesifiserer et obligatorisk lokativt argument, men som naturligvis kan opptre sammen med et fakultativt lokativt argument, som i:

- (12) a. Det snør på fjellet.
b. Det snør ute.

Merk at ekspletivsubjektet er kodet som en nominalfrase uten tematisk rolle. Legg forøvrig også merke til at det lokative argumentet er kategorielt realisert som en preposisjonsfrase i argumentstrukturen, men PP må fortolkes som preposisjonsfrase eller lokativt adverb.

Den mest vanlige intransitive typen har vi allerede sett på, jf. (5) ovenfor. En annen og mer interessant type for syntaktikere er intransitive verb som tar tematiske subjekter, dvs. verb som *ankomme*, *blekne*, *blomstre*, osv.

- (13) *Intransitivkonstruksjon #2:*

Subjekt	Theme	NP
---------	-------	----

Argument 1

Det er viktig å adskille denne typen intransitive verb fra dem med agentive subjekter fordi de normalt ikke kan passiviseres. Denne distinksjonen mellom intransitive verb forsvinner dersom man ikke deler dem opp i undertyper. En tredje type intransitive verb er de som tar "opplevelsessubjekt", dvs. verb som *fryse*, *kaldsvette*, *klø*, osv. De er beskrevet slik:

- (14) *Intransitivkonstruksjon #3:*

Subjekt	Experiancer	NP
---------	-------------	----

Argument 1

Den mest frekvente typen transitive verb har vi allerede sett på, jf. (6) ovenfor. Verb som tar setningskomplementer (*anta*, *si*, *påstå*, osv) er beskrevet som i (15), mens (16) og (17) viser kodingen av transitive verb med henholdsvis opplevelsesobjekt (altså verb som *irritere*, *berolige*, *bekymre*) og opplevelsessubjekt, dvs. verb som *like*, *foretrekke*, *oppleve*:

(15) *Transitivkonstruksjon #2:*

Subjekt	Agent	NP
---------	-------	----

Argument nr. 1

Objekt	Theme	setning
--------	-------	---------

Argument nr. 2(16) *Transitivkonstruksjon #7:*

Subjekt	Theme	NP
---------	-------	----

Argument nr. 1

Objekt	Experiancer	NP
--------	-------------	----

Argument nr. 2(17) *Transitivkonstruksjon #8:*

Subjekt	Experiancer	NP
---------	-------------	----

Argument nr. 1

Objekt	Theme	NP
--------	-------	----

Argument nr. 2

En notorisk vanskelig verbtype er de som synes å ta både tematiske subjekt og tematisk objekt, dvs. verb som *gjelde* og *mangle*. I NorKompLeks er de kodet med to tematiske roller, som en praktisk nødløsning i mangel av bedre alternativer. Men det bør likevel bemerkes at denne typen verb er identifisert i materialet, og syntaktikere kan ved å få dem listet opp danne seg et bilde av hvilke verb som har slike "problematiske" egenskaper:

(18) *Transitivkonstruksjon #6:*

Subjekt	Theme	NP
---------	-------	----

Argument nr. 1

Objekt	Theme	NP
--------	-------	----

Argument nr. 2

Det er på sin plass å si noe om konstruksjonstyper som ikke er såpass enkle å beskrive som dem ovenfor, f.eks. refleksive verb, verb med preposisjonsobjekt og partikkelverb. Typiske refleksive verb som *skamme seg* og *gremme seg* kan kodes slik:

(19) *Refleksivkonstruksjon #1:*

Subjekt	Experiancer	NP
---------	-------------	----

Argument nr. 1

Objekt	Uspesifisert_rolle	NP[kjerne=seg]
--------	--------------------	----------------

Argument nr. 2

Det er uklart om *seg* i disse konstruksjonene har noen tematisk rolle, og det er vanskelig å spesifisere hvilken det i så fall skulle være. Derfor er den markert som uspesifisert. Legg merke til at det refleksive elementet behandles som et argument, selv om grammatikere ofte beskriver det som en partikkel til verbet. Preposisjonsobjekt finner vi i konstruksjoner som:

- (20) a. Han kjefter på læreren.
b. Hun agiterer for moderne syntaksteori.

Preposisjonsfrasene har ingen adverbial funksjon i slike setninger og bør heller analyseres som objekter. Konstruksjonsbeskrivelsen i (21) gir riktig resultat for verbene *kjeft* og *agitere* når de er kodet som i (22a,b):

(21) *Transitivkonstruksjon #11:*

Subjekt	Agent	NP
---------	-------	----

Argument nr. 1

Objekt	Theme	PP[kjerne=VAR]
--------	-------	----------------

Argument nr. 2

- (22) a. *agitere* : [intrans1,trans11[kjerne=for]]
b. *kjeft* : [intrans1,trans11[kjerne=på]]

Valg av preposisjon er ikke vilkårlig og styres av verbene, noe som er indikert ved at kjernen er en variabel (VAR). I refleksivkonstruksjonen (19) er det konstruksjonen selv som bestemmer kjernen i frasen, i dette tilfellet *seg*.

Verbet *brekke* har nokså mange realisasjonsmuligheter, og en av dem er at det kan brukes som partikkelverb:

- (23) a. *brekke* : [part1[av,over,om], ...]
 b. Han brekker av greinene. / Han brekker greinene av.
 c. Han brekker om sidene. / Han brekker sidene om.

Vi ser at formalismen tillater at flere kjerner kan listes opp ved hvert verb. Koden for denne konstruksjonen er beskrevet slik:

(24) *Partikkelkonstruksjon #1*:

Subjekt Agent NP	Partikkel Ingen_rolle PP[kjerne=VAR]	Objekt Theme NP
<i>Argument nr. 1</i>	<i>Argument nr. 2</i>	<i>Argument nr. 3</i>

Den syntaktiske beskrivelsen av partikkelverbkonstruksjoner er ikke enkel, og lingvistiske teorier behandler dem på ulikt vis. I NorKompLeks analyseres partikkelen som et kvasiargument i argumentstrukturen, dvs. et argument uten tematisk rolle, og det er tilordnet en ad hoc syntaktisk funksjon *partikkel*. Dette er en nødløsning som lingvister neppe vil la seg begeistre av, og koden i seg selv sier ingenting om at partikkelen kan komme foran eller etter objektet. Slike spørsmål må uansett overlates til syntaksteorier fordi distribusjonen av partiklene er begrenset av syntaktiske kriterier og hører ikke naturlig hjemme i leksikon. Men igjen vil vi påpeke at oppgaven til leksikonet å vise til hvilke verb som har disse egenskapene, og det blir opp til språkvitenskapelig teori å analysere dem skikkelig.

Verb som tar infinitivskomplementer må ha med informasjon om hvilket setningsledd som "kontrollerer" infinitivssubjektet. Et eksempel er verbet *love* som kan ta indirekte objekt og subjektskontrollert infinitivskomplement. Koden er:

(25) *Kontrollkonstruksjon #2*:

Subjekt Agent NP	Indirekte Recipient NP objekt	Objekt Theme Infinitivskomplement
<i>Argument nr. 1</i>	<i>Argument nr. 2</i>	<i>Argument nr. 3</i>

RESTRIKSJON: *Argument nr. 1 sitt subjekt kontrollerer infinitivskomplementets subjekt i argument nr. 3.*

Objektskontroll er beskrevet på samme måte, bortsett fra at restriksjonen er at det er det indirekte objektet som kontrollerer infinitivskomplementets subjekt.

Vi har nå gitt en generell beskrivelse av hovedstrategien i valensbeskrivelsene i NorKompLeks, men det er et par viktige momenter som vi ennå ikke har vært innom. Det ene gjelder hvilke verb som kan passiviseres, og det andre er hvilke verb som kan brukes i presenteringskonstruksjoner. Passiviseringsmulighetene til verbene er leksikalsk bestemt, og mye arbeid er i de senere år lagt ned i å teoretisere over hvilke leksikalske egenskaper verb som ikke lar seg passivisere har, jf. f.eks. Åfarli (1992) for en grundig analyse av relevante norske data. Lingvister er spesielt interessert i å vise hvordan passiv og presentering henger sammen, men i praksis viser det seg at passiviserings- og presenteringssegenskaper er mer komplekse enn lingvistisk teori i dag makter å forutsi. Derfor har man i NorKompLeks simpelthen valgt å liste opp alle verb som ikke kan opptre i passivkonstruksjoner og de som kan forekomme i presenteringskonstruksjoner.

5. Leksikografisk relevans

Detaljerte beskrivelser av verbs valenssegenskaper finner man ikke i ordbøker, og enkelte ordbøker har ikke valensinformasjon i det hele tatt, f.eks. *Bokmålsordboka*. Det er vel tvilsomt om så detaljert valensinformasjon som NorKompLeks tilbyr har en naturlig plass i allmennspråklige ordbøker. Men det er nokså trivielt å forenkle NorKompLeks-systemet dersom man ønsker det. De seks forskjellige intransitivkodene kan reduseres til en, og man kan redusere de for tiden 26 transitivkodene til en type. Det samme kan gjøres med partikkelkodene og refleksivkodene. Informasjonsinnholdet i valensinformasjonen vil bli redusert, men slike grovklassifiseringer er uansett nyttig informasjon for de brukerne av ordbøker som har en viss grammatisk skoling.

Nye publiseringsformer for ordbøker åpner for at brukerne selv kan bestemme hvor detaljert informasjon de ønsker. På denne måten blir relevansen til prosjekter som NorKompLeks

tydeligere: Ordboken kan i sin basisform referere til grovklassifisert valensinformasjon, men de som måtte være interessert i mer detaljert informasjon kan hente den ut. Vi kan også se for oss at formaliserte lingvistiske teorier gir sine fortolkninger av argumentstrukturinformasjonen slik at man kan få frem leksikonbeskrivelser i leksikalsk-funksjonell grammatikk, hodedrevet frasestrukturgrammatikk og kanskje en eller annen versjon av transformasjonsgrammatikk. På denne måten kan leksikografi og formell språkvitenskap nærme seg hverandre i den forstand at ordbøkene i større grad kan være med på å definere viktige empiriske sider for lingvistikken, mens lingvistikken på sin side kan synliggjøres i de leksikografiske sluttproduktene.

Litteratur

- Bresnan & Kaplan. 1982/1995. "Lexical-Functional Grammar: A Formal System for Grammatical Representation." I: Dalrymple, Kaplan, Maxwell & Zaenen (red.): *Formal Issues in Lexical-Functional Grammar*. CSLI lecture notes 47, Stanford, California.
- Chomsky, Noam. 1981. *Lectures on Government and Binding*. Foris, Dordrecht.
- Chomsky, Noam. 1995. *The Minimalist Program*. MIT Press, Cambridge Mass.
- Goldberg, Adele. 1995. *Constructions : A Construction Grammar Approach to Argument Structure*. University of Chicago Press.
- Landrø M.I. og B. Wangensteen. 1993. *Bokmålsordboka : definisjons- og rettskrivningsordbok*. Oslo : Universitetsforlaget.
- Manning, Christopher and Ivan A. Sag. 1998. "Argument structure, valence, and binding." *Nordic Journal of Linguistics* 21:107-144.
- Nordgård, Torbjørn. 1998. "Norwegian Computational Lexicon (NorKompLeks)." *Proceedings of the 11th Nordic Conference on Computational Linguistics*. Center for Sprogteknologi, Københavns Universitet.

- Pollard, Carl and Ivan A. Sag. 1987. *Information-based syntax and semantics*. CSLI lecture notes 13, Stanford, California.
- Pollard, Carl and Ivan A. Sag. 1994. *Head-driven phrase structure grammar*. University of Chicago Press, Chicago.
- Åfarli, Tor Anders. 1992. *The syntax of Norwegian passive constructions*. Amsterdam : J. Benjamins.

Appendiks. Implementerte koder i NorKompLeks

Nedenfor er de Prolog-implementerte kodene som artikkelen refererer til, listet opp, sammen med en referanse til nummereringen i teksten.

- (5) `arg_code(intrans1,[arg1:su::ag::np])`.
- (6) `arg_code(trans1,[arg1:su::ag::np,arg2:obj::th::np])`.
- (8) `arg_code(nullv,[arg1:su::norole::np])`.
- (9) `arg_code(nullv1,[arg1:su::norole::np,arg2:adv::loc::pp])`.
- (13) `arg_code(intrans2,[arg1:su::th::np])`.
- (14) `arg_code(intrans3,[arg1:su::exp::np])`.
- (15) `arg_code(trans2,[arg1:su::ag::np,arg2:obj::th::s1])`.
- (16) `arg_code(trans7,[arg1:su::th::np,arg2:obj::exp::np])`.
- (17) `arg_code(trans8,[arg1:su::exp::np,arg2:obj::th::np])`.
- (18) `arg_code(trans6,[arg1:su::th::np,arg2:obj::th::np])`.
- (19) `arg_code(refl1,[arg1:su::exp::np,arg2:obj::unspec::np:[kjerne=seg]])`.
- (21) `arg_code(trans11,[arg1:su::ag::np,arg2:obj::th::pp:[kjerne=VAR]])`.
- (24) `arg_code(part1,[arg1:su::ag::np,arg2:part::norole::pp:[kjerne=VAR],arg3:obj::th::np])`.
- (25) `arg_code(ditrans3,[arg1:su::ag::np,arg2:io::rec::np,arg3:obj::th::inf,arg3:obj:su=arg1:su])`.

Metaordboken – et rammeverk for Norsk Ordbok?

Innledning

På slutten av 1980-tallet vokste det frem en ide ved Det historisk-filosofiske fakultet ved Universitetet i Oslo om å la data-maskinene overta for arkivskapene i en rekke samlingsavdelinger og museer. Dette noe uspesifiserte, men på det daværende tidspunkt svært ambisiøse og fremtidsrettede utopiske målet ble til det noe mer jordnære Dokumentasjonsprosjektet. For de leksikografiske avdelingene skulle prosjektet lette ordboksarbeidet ved å erstatte ordsedler og annet materiale med databaser og ikke minst den måten gjøre informasjonen i arkivene tilgjengelig for et større publikum.

De teknologiske forutsetningene i verden rundt oss har utviklet seg voldsomt i de 7 årene prosjektet varte. Det som kunne virke utopisk i 1990, er i dag hverdagslige realiteter. Vi er alle knyttet til Internett der vi kan hente ut informasjon om de fleste emner. Det foregår også en stadig økende elektronisk publikasjon av ordbøker, tekstarkiv og tekstkritiske kommentarutgaver både via Internett og via CD-ROM. Men det som kanskje er viktigere for leksikografer og forfattere av andre typer oppslagsverker, er at man nå ser de de første økonomiske konsekvensene av Internett. Det foregår for tiden vesentlige endringer i måten de store forlagene publiserer sine store ordbøker og andre oppslagsverk på. Verken *Oxford English Dictionary* (OED) eller *Encyclopaedia Britannica* vil lengre komme i papirform. Sistnevnte finnes nå bare i en reklamefinansiert gratisutgave på Internett. Tilsvarende strukturendringer har nå også kommet i Norge der vårt største forlag for oppslagsbøker, Kunnskapsforlaget, av økonomiske grunner legger om sin utgivelsesstrategi for sitt største leksikon fra papir til Internett.

Hvordan kan så et stort og tradisjonelt ordboksverk som *Norsk Ordbok* klare seg i denne moderne heksegryten? *Norsk Ordbok* er et nasjonalt ordboksverk som redigeres og utgis i tradisjonen fra begynnelsen av 1900-tallet. Internett-alderen vil neppe ha noen direkte økonomiske konsekvenser slik den har for de kommersielle oppslagsverkene idet salgsinntektene ikke utgjør noen vesentlig del av finansieringen. Men indirekte vil det kunne ha alvorlige konsekvenser siden særlig yngre mennesker forventer å finne informasjon gjennom elektroniske kanaler. Dette kan forverres ytterligere av den langsomme publiseringstakten. Arbeidet med *Norsk Ordbok* har til nå vart i nesten 70 år, og med dagens ordartikkelvolum vil man ikke bli ferdig før rundt år 2060. Sett i et historisk lys er dette lange tidsperspektivet langt fra enestående for et nasjonalt ordboksverk. Men en slik trøst setter strenge krav til den idealistiske tålmodighet hos oss nålevende mennesker. Det er derfor ikke urimelig om man ønsker å søke etter løsninger som ikke bør være til hinder for en realisering av det store planlagte verket, men samtidig kan formidle mer stoff, om enn i en noe ubearbeidet form, til dem som venter.

I denne artikkelen prøver jeg å belyse en mulig løsning på problemet gjennom å etablere det jeg vil kalle en metaordbok. Dette er riktignok ingen ordbok eller bok i det hele tatt. Metaordboken er en blanding av en leksikalsk database, et elektronisk seddelarkiv, et historisk ordboksbibliotek og et ordboksredigeringsverktøy med et voksende manuskript. En slik hybrid vil etter min mening være nødvendig for å kunne omplassere de store nasjonale ordboksverkene i en nåtidig omgivelse.

Ideen til metaordboken har vokst frem gjennom arbeidet med Dokumentasjonsprosjektet (1991–1997) og har hentet mye spesielt fra arbeidet med elektronisk utgivelse av eldre tekster som middelalderdiplomer og 1800-talls litteratur i Dokumentasjonsprosjektet. Det største enkeltprosjektet innenfor Dokumentasjonsprosjektet var likevel å lage en elektronisk erstatning for seddelarkivet og annet bakgrunnsmateriale for *Norsk Ordbok*. Dette arbeidet startet i august 1991 og ble i hovedsak avsluttet i 1998.

Metaordboken ble i sin første form foreslått som en mulighet i en artikkel skrevet i forbindelse med et innlegg på seminaret "Norsk Ordbok – nynorskens kanon" i 1996 (se min artikkel i Vikør 1997). Denne artikkelen er en viderutvikling og utdypning av seminarartikkelen. Det har vist seg at utviklingen igjen har gått forttere enn ventet. Metaordboken er også kommet nærmere en realisering siden vi i 1999 har hatt penger til et mindre pilotprosjekt gående for å prøve å bygge opp ryggraden i metaordboken. Det og mulighetene for videre arbeid vil bli behandlet til slutt i denne artikkelen.

Siden Dokumentasjonsprosjektets arbeid er vesentlig for muligheten av å realisere en metaordbok, har jeg gjort kort rede for prosjektet og hva databasene inneholder. Hovedvekten er lagt på en gjennomgang av forholdet mellom arkiv, kommenterte tekstutgaver og annen vitenskapelig publisering i en elektronisk tidsalder og hvilke innvirkninger dette kan og bør ha for et prosjekt som *Norsk Ordbok*.

Norsk Ordbok og seddelarkivet

Arbeidet med *Norsk Ordbok* har til nå vart i nesten 70 år. Prosjektet har vokst langt ut over de forestillingene man hadde i 1930 om å sammenstøpe ordtilfanget i de da eksisterende ordbøkene for nynorsk og utvide dette med nyere ord til et firebindsverk. I dag er målet "ei vitenskapelig ordbok over dei norske målføra og det nynorske skriftspråket" planlagt i 12 store bind.

Det store skiftet i utgivelsesplanen synes å ha kommet rundt andre verdenskrig da man erkjente at det da ferdige Grunnmanuskriptet ikke uten videre egnet seg som grunnlag for et trykkmanuskript, og at det tilfanget av nyere ord utgjorde en adskillig viktigere og større del av ordboken enn det man i utgangspunktet hadde forestilt seg. En del av denne erkjennelsen kan muligens ha kommet fra arbeidet med å bygge opp et leksikografisk seddelarkiv i forbindelse med prosjektet. Et typisk kort i et slikt arkiv inneholder et ord i grunnform, en liten tekstbit som gir et eksempel på en bruk av ordet, hvor ordet er brukt, opp-

lysninger om grammatikk, uttale og eventuelle andre forhold rundt dette eksempelet. Arkivet er sortert alfabetisk etter ordenes grunnform. Om man leser forordet til de ulike heftene som til nå er kommet av *Norsk Ordbok*, slår det en at oppbygningen av seddelarkivet har vært svært sentral hele tiden helt til det siste tiåret. Det har vært et uttalt ønske om å få arkivet størst mulig for nettopp å kunne ha mange eksempler for hvert eneste ord, også de mer sjeldne. Seddelarkivet rommer nå omlag 3,2 millioner sedler. Dette er ikke spesielt mye. Tilsvarende nasjonale prosjekter i Sverige og Danmark har mer en 10 millioner hver, mens arkivet til *Oxford English Dictionary* rommer mer enn 30 millioner sedler.

Seddelarkivet er altså den tradisjonelle systematiske metoden for å samle belegg og opplysninger om ord til bruk i redigeringen av en ordbok. I de siste 40–50 årene har imidlertid datateknologien muliggjort alternative metoder for å fange inn og lagre tilsvarende informasjon. Den nye teknikken har skapt et skille i synet på bruk av materialet i ordboksredigeringen. For noen står seddelarkivet som en tilfeldig samling opplysninger eller deler av en konkordans, og arkivet er kun et eksempel på hva gårdsdagens teknikk kunne produsere. For andre representerer seddelarkivet en skattkiste der hver seddel er valgt med omhu, mens det med et korpus kan være vanskelig få med det spesielle og sjeldne språket.

Dokumentasjonsprosjektet

Formålet med Dokumentasjonsprosjektet var å bidra til å omstille samlingsavdelingenes behandling av informasjon til moderne datateknikk og dermed effektivisere det interne samlingsarbeidet, eksternt samarbeid og utveksling av informasjon samt innhenting av ny informasjon. På denne måten ønsker man både å utløse et forskningspotensiale, men også å avdekke svakheter i rutiner og systemer og forbedre disse. Det var også et viktig mål å tilgjengeliggjøre informasjonen for andre forskere, for studenter, for undervisning, for offentlig forvaltning og for allmennheten

så langt dette var forsvarlig ut fra personvern, sikkerhet, opphavsrettigheter og eventuelle kommersielle hensyn.

Dokumentasjonsprosjektet kan grovt deles inn i en museumsdel og en språklig orientert del. I den museumsrettede delen av prosjektet arbeidet vi med å bygge opp forskningsdatabaser for de arkeologiske museene i Bergen, Oslo og Tromsø. I Tromsø arbeidet vi også med nyere kulturhistorisk materiale. Den språklige delen bestod i hovedsak av tilrettelegging av bakgrunnsmateriale for ordboksavdelingene (bokmål, gammelnorsk og nynorsk) ved Avdeling for Leksikografi i Oslo og for Trønderordboka ved Norges teknisk-naturvitenskapelige universitet (NTNU) og for navnegranskingsmiljøene i Oslo og Tromsø.

Dokumentasjonsprosjektet ble avsluttet i 1997. Museumsarbeidet ble i 1998 fortsatt gjennom et større Museumsprosjekt. Arbeidet med de språklige samlingene blir videreført av en liten etterorganisasjon i perioden 1998–2001. Etter den tid vil antakeligvis samlingene bli ivaretatt av et "Senter for språklige data-samlinger" ved Universitetet i Oslo.

Dokumentasjonsprosjektet og de leksikografiske arkivene

Den opprinnelige målsetningen for delprosjektene ved den daværende Avdeling for leksikografi ved Institutt for nordistikk og litteraturvitenskap på Universitetet i Oslo var "å gjøre seddelarkivene tilgjengelig på elektronisk form". Til sammen er det omlag 9 millioner sedler i avdelingens arkiver. Ved en mer inngående analyse av seddelmaterialet ble det besluttet bare å konvertere nyordsarkivet til Bokmålsavdelingen, seddelarkivet til Gammelnorskavdelingen og hele nynorskarkivet (Seddelarkivet til *Norsk Ordbok*), i alt 4 millioner sedler. De resterende sedlene var stort sett laget ut fra hele forfatterskap (Henrik Wergeland, Bjørnstjerne Bjørnson mm.) og inneholder den samme informasjonen som man vil finne i en vanlig KWIC-konkordans. Det aller meste av Bokmålsavdelingens sedler er av denne typen. Istedet for å arbeide med ordsedlene har vi her brukt optiske leseprogrammer (OCR) for å gjøre de opprinnelige tekstene

elektronisk tilgjengelige. Disse tekstene ble kodet i overensstemmelse med anbefalingene til det SGML-baserte "Text Encoding Initiative" (Goldfarb 1991, Sperberg-McQueen and Burnard 1994). Tekstene kunne da også legges ut til almenheten i et lite elektronisk nettbibliotek på en tilsvarende måte som vi finner i det svenske Runeberg-prosjektet eller det amerikanske Project Gutenberg.

Seddelarkivet er som nevnt tidligere, den tradisjonelle systematiske metoden for å samle belegg og opplysninger om ord til bruk i redigeringen av en ordbok. Det store arbeidet som er lagt ned i oppbygningen og vedlikeholdet av seddelsamlinger vitner om hvilket viktig hjelpemiddel ordsedlene har vært og fremdeles er i ordboksarbeid. I de senere år har datateknikken muliggjort en mye mer effektiv oppbygning av den informasjonen som et seddelarkiv representerer. Elektronisk lesning av tekst (OCR), konkordansprogrammer og hjelpemidler for (halv-) automatisk markering av grammatisk informasjon til ord i løpende tekst kan nevnes (se for eksempel Atkins 1992).

I et leksikografisk seddelarkiv er det for et utvalg ord fra et utvalg tekstbrokker (en til 20 linjer) gitt opplysninger om bøyingsformen, ordets rot og annet. Et tagget tekstkorpus er derimot en samling større tekstfragmenter (fra 25 sider løpende tekst og oppover) der hver ordform i tekst(fragment)ene har fått markert grunnord, ordklasse og aktuell bøyingsform. Informasjonen i et seddelarkiv kan altså sammenlignes med den vi finner i et tagget tekstkorpus. Denne observasjonen lå til grunn da vi planla konverteringen av de leksikografiske arkivene og valgte å erstatte bokmålsarkivet med løpende elektronisk tekst, riktignok uten grammatisk informasjon.

Gammelnorskavdelingens arkiv er også et arkiv av "KWIC-konkordans" typen. Men sedlene er så systematisk bygd opp at vi her har valgt å bruke informasjonen på dem til å bygge opp hele tekster der hver ordform har fått markert grunnord, ordklasse og aktuell bøyingsform. Vi har altså laget et lite tagget tekstkorpus for det gammelnorske materialet.

Arkivet til *Norsk Ordbok* skiller seg ut fra de to andre arkivene ved at det er bygd opp av mange hundre frivillige uten spesiell

leksikografisk utdanning over en lang periode (omlag 60 år). Dette har resultert i et heterogent arkiv som både består av rene ekserptsedler og av sedler med mye ekstra informasjon om blant annet bruk og uttale. En effektiv konvertering av et slikt arkiv fordrer at de rene ekserptsedlene frasorteres og erstattes med elektronisk tekst. De resterende sedlene kan skrives inn og SGML-kodes (Sperberg-McQueen and Burnard 1994) slik at de kunne lastes inn i et databasesystem. Dette var også den opprinnelige planen for vårt arbeid. Etterhånden viste det seg at det var ekstremt ressurskrevende å foreta en sortering av arkivet. Det viste seg også at avskrift og koding av sedlene byr på de samme prinsipielle og metodologiske problemene som enhver transkripsjon av håndskrevet materiale. Vi valgte derfor gå bort fra avskriften av sedlene. I stedet satset vi på en database med faksimiler av samtlige nynorsksedler. Denne samlingen av faksimiler har oppslagsord, ordklasse og uttømmende kildeopplysninger som søkenøkler. Man mister på denne måten mulighetene til å søke i den løpende teksten på sedlene, men har fremdeles muligheten til å finne sedler etter grunnord, ordklasse, sted i landet, kildetype og hvem som har skrevet seddelen. Se (Ore 1996) for en inngående diskusjon av fordeler og ulemper ved denne løsningen.

Dokumentasjonsprosjektet og leksikalske databaser

Med leksikalske databaser menes databaser som inneholder informasjon om ord som deres oppbygning og bøyning (morfologi), grammatisk funksjon, mening, relativ frekvens og så videre. En leksikalsk database er altså datateknikkens svar på en ordbok. Den skiller seg fra en tradisjonell ordbok også ved at det ikke er meningen at all informasjonen skal leses av mennesker. I mange tilfeller kan leksikalske databaser inneholde informasjonen kodet til bruk i ulike språkteknologiske verktøy så som morfologiske analysatorer, syntaksanalysatorer og oversettelsesstøttesystemer. Men en leksikalsk database vil typisk kunne inneholde teksten

fra en eller flere tradisjonelle ordbøker som hjelp og supplement til menneskene som bruker basen.



FIGUR 1. *En skisse av den leksikalske databasen.*

Dokumentasjonsprosjektets nynorske orddatabase (leksikalske database) inneholder informasjon fra mange kilder, seddelarkiv, ordbøker og ordsamlinger. Blant ordbøkene kan vi nevne *Aasens Norsk Ordbok* (1872), *Nynorskordboka* (1994) og *Grunnmanuskriptet* som er et første utkast til Norsk Ordbok fra 1930-tallet (se senere for en nærmere beskrivelse). I tillegg finner vi 20–30 ordsamlinger fra 1600-, 1700- og 1800-tallet. Man arbeider også med å legge inn den såkalte Storm-synopsen som er en oversikt over uttale og bøyning av omlag 1000 ord 1000 steder i Norge. I tillegg til dette historiske materialet inneholder orddatabasen også bøyningsinformasjon og annen grammatisk informasjon hentet fra IBM og fra arbeidet med et "Komputasjonelt leksikon (NorKompLeks)" ved Lingvistisk institutt ved NTNU (Nordgaard 1995). Dokumentasjonsprosjektets nynorske

ordbaser vil dermed være en blanding av en leksikalsk database i ordets mer tekniske betydning og av en samling av elektroniske versjoner av ordbøker, ordsamlinger og annet leksikografisk bakgrunnsmateriale. Figur 1 gir en skjematisk oversikt over den nynorske leksikalske databasen slik denne kan bli.

Et viktig punkt som krever norskfilologisk og helst leksikografisk ekspertise, er å etablere en kobling av grunnord mellom de ulike delbasene. Dette kompliseres av de ulike rettskrivingstandardene som er brukt. De eldre ordsamlingene og ordbøkene følger sin egen rettskriving. Seddelarkivet og Grunnmanuskriptet følger ideelt sett 1938-rettskrivingen, mens det nyere materialet følger moderne rettskriving. Sammenkobbingsarbeidet krever altså at alle oppslagsordene i de ulike samlingene og verkene også får påført en variant som følger den moderne rettskrivingen eller eventuelt 1938-rettskrivingen. Dette er et ikke-trivielt problem siden ingen av de nevnte normene dekker alle grunnordene i databasen. Fagkonsulentene i Dokumentasjonsprosjektet begynte standardiseringsarbeidet. Dette arbeidet har en nå tatt opp igjen i pilotprosjektet til metaordboken. Men med dagens utgivelsesstrategi for *Norsk Ordbok* vil den leksikalske databasen de facto være den eneste samordnede presentasjon av grunnlagsmaterialet fra A til Å i de neste 50 årene.

En frigjøring fra Gutenberg

Mange forskningsprosjekter vil vanligvis bygge på store mengder data som er samlet inn gjennom eksperimenter, gjennom feltarbeid eller som resultat av besøk i biblioteker og arkiver. På bakgrunn av disse dataene utarbeides det artikler og rapporter som presenterer de resultatene man har kommet frem til. Prosjektgenerert bakgrunnsmateriale blir så enten ødelagt eller bevart i et eller annet arkiv eller bibliotek. Et slikt arkivmateriale kan selv bli gjenstand for (meta)forskning ved en senere anledning der det har vokst frem et ønske om å finne ut hvordan prosessen mot det ferdige resultatet egentlig var.

Som den motsatte ytterlighet til formidling av forskningsresultater gjennom artikkelen, kan vi sette temaarkivet eller temabiblioteket. Det er en lang tradisjon innen arkiv og biblioteksverden å samle arkiver og bibliotek over spesielle temaer eller over enkeltstående personers etterlatte papirer. Her kan for eksempel nevnes Wittgensteinarkivet i Bergen, Kildeskriftsavdelingen ved Riksarkivet i Norge eller ved Stiftinga Nynorsk-kultursentrum (tidligere Ivar Aasen-stiftinga) der man vel originalt tenkte seg å samle alle Ivar Aasens etterlatte papirer, hans bibliotek og ideelt sett alt som er skrevet om eller på nynorsk.

Tekstkritiske utgaver av ulike verk og verk som prøver å etablere "en beste tekst" ut fra eksisterende versjoner, vil på denne tenkte skalaen kunne plasseres nærmere den vitenskapelige artikkel enn bibliotek og arkiv. Som eksempler på slike arbeider kan jeg nevne så ulike ting som utgivelser av Homer og bibelforskning. Den tradisjonelle publiseringen av større tekstsamlinger ved bruk av mikrofilm ligger derimot nær bibliotek eller arkiv på denne skalaen.

Denne klassifikasjon kan kanskje virke noe uinteressant om ikke introduksjonen av ny teknologi hadde snudd opp ned på en lang rekke tilvante publiseringsmetoder. Det har til nå vært høyst uvanlig at forskningsprosjekters bakgrunnsmateriale er blitt publisert selv i de tilfeller der institusjonen som har huset prosjektet, ikke har noen kommersielle eller sikkerhetsmessige grunner til å holde materialet for seg selv. Mye av årsaken ligger selvfølgelig i kostnadene. Men i en tid hvor det meste av ny informasjon finnes i elektronisk form, og hvor lagringsmediene (f.eks. CD-ROM) er svært billige, vil en slik publisering av bakgrunnsmateriale sammen med resultatene av forskningen kunne bli alminnelig.

Innenfor de humanistiske fag har det i løpet av 1990-tallet blitt mer allminnelig å gi ut selve kildemateriale i elektronisk form utstyrt med søkeverktøy og/eller et elektronisk note- eller kommentarapparat. Denne trenden var tydelig allerede i 1996 (se Ore 1997). Slike publikasjoner er på mange måter svært like mikrofilmversjoner av et arkiv eller bibliotek, men synes likevel å være nærmere intensjonen bak tekstkritiske utgaver. Den publi-

seringsmetoden er også et svært godt alternativ til det tradisjonelle filologiske arbeid der kildematerialet kun refereres gjennom noteapparatet. Ved å legge selve bakgrunns materialet sammen det vitenskapelige arbeidet, endres forholdet mellom kildematerialet og den vitenskapelige artikkelen. Som George Landow påpeker, får fotnoten eller henvisningen en annen rolle. Fotnotene er ikke lenger underordnede tekstfragmenter, men vil bli pekere inn i andre komplette tekster (Landow 1991). I stedet for å stå som en egen overordnet enhet vil på mange måter den vitenskapelige artikkelen kunne bli et kommentarverk som eksisterer i parallell med sine kilder. Artikkelen vil således bevege seg i retning av en spesialisert kommentert tekstutgave.

Muligheten for å publisere hele arkiver med et noteapparat vil på den annen side kunne viske ut forskjellen mellom arkiv/bibliotek og de filologiske tekstkritiske samlingsutgavene. Dette vil stoppe utgivelsen av de store trykte kommentert tekstutgavene med omfattende tekstkritisk noteapparat. Av åpenbare økonomiske årsaker, men også av praktiske årsaker vil de bli erstattet av utgivelser av arkiver der grunnmaterialet vil være mer eller mindre bearbejdede rådata i form av avskrifter eller faksimiler supplert med ulike fortolkede utgaver av de samme tekstene. Man kan også stille seg spørsmålet om i hvilken grad de svært forseggjorte noteapparatene i de tekstskritiske utgavene virkelig brukes av forskere til å rekonstruere alternative tekstvitner, eller om disse forskerne heller vil søke til (faksimiler av) originalmanuskriptene i det de ikke stoler på sin egen evne til å gjennomføre dette puslespillet nøyaktig nok (se Vanhoutte 1999 for en interessant debatt om dette). Vanhoutte opererer også med en arkiv/museums-modell, der arkivdelen svarer til arkivklassifikasjonen over, museumsdelen, eller i min tolkning utstillingsdelen, er den aktuelle tekstkritiske utgave.

I dag finnes det uttallige eksempler på elektronisk utgitte tekstarkiv og tekstarkiv koplet mot et kommentarapparat. Det fantes allerede i 1996 tre ulike Shakespeare utgivelser på CD-ROM. Det første store og gode eksempelet på elektroniske kommenterte tekstutgaver var "The Wife of Bath's Prologue" publisert som CD-ROM gjennom "The Canterbury Tales Project"

(Robinson 1996). Her er omlag 40 ulike manuskripter lemmatisert og sammenstilt med felles kommentarapparat og søke-system. I Norden vil vel både den nye August Strindberg-utgaven, Søren Kierkegaard-utgaven og Henrik Ibsen-utgaven komme som elektroniske produkter.

Norsk Ordbok og de nye publikasjonstrendene

I det foregående avsnittet ga jeg en kort diskusjon om forholdet mellom edisjonsfilologi og elektroniske metoder for publisering. Har dette edisjonsfilologiske aspektet noen relevans for historisk orientert leksikografi?

Prosjektet "Norsk Ordbok" slik vi kjenner det i dag, er et tradisjonelt nasjonalt ordboksprosjekt med sterk vekt på historisk leksikografi. Verk av denne typen blir oppfattet som en autoritativ og uttømmende beskrivelse av betydning og bruk av ordene i et språk. Enhver filologisk skolert person vet at dette ikke er tilfellet. Men de fleste vil akseptere påstanden om at disse historiske nasjonalensyklopediene er det nærmeste man kommer en slik altomfattende beskrivelse av et språk. Mange moderne ordbøker er laget mot en bestemt målgruppe som for eksempel skoleelever eller (fremmedspråklige) studenter. Målgruppen for de store nasjonalensyklopediene er ofte noe mer diffus. En ordartikkel i de store verkene representerer resultatet av redaktørens analyse av sitt kildemateriale der de ulike betydninger og ordets utbredelse er underbygget med kildehenvisninger på en tilsvarende måte som noteapparatet i et hvilket som helst annet vitenskapelig filologisk arbeid. For å forstå mange av artiklene fullt ut forutsettes det derfor et relativt høyt kunnskapsnivå hos leserne. *Norsk Ordbok* presenterer seg selv som en "vitenskapelig ordbok" som åpenbart henviser både til metode for redigering og for anvendelsesområde. På dette grunnlaget kan det være riktig å sjangerbestemme *Norsk Ordbok* som vitenskapelig publikasjon.

Hvis man nå kan betrakte *Norsk Ordbok* som en slags antologi av vitenskapelige artikler, hvilke resultater vil vi få dersom vi appliserer de påståtte trendene fra forrige avsnitt? Prosjektet slik

det er drevet i dag, baserer seg direkte på et seddelarkiv, en målføresamling (Storm-synopsen se over) samt en rekke eksisterende ordbøker. Indirekte gjennom seddelarkivet baserer det seg på tusenvis av skriftlige kilder. Det er ikke urimelig å betrakte seddelarkivet som resultatet av den datainnsamling som foregår i ethvert empirisk prosjekt. Man skulle dermed kunne tenke seg en samlet publisering av både bakgrunnsmateriale og artiklene som presenterer resultatene, det vil si ordboksartiklene. I dette tilfellet vil dermed artiklene være systematiseringer og kommentarer til bakgrunnsmateriale. Kildehenvisningene vil være pekere inn i det originale materialet som er brukt i redigeringen. Man ville dermed få en enestående mulighet til å kunne studere hvordan ordboksredaktøren har arbeidet frem artikkelen. Siden Dokumentasjonsprosjektet har gjort hovedmengden av kilde-materialet elektronisk tilgjengelig, er forutsetningene til stede for en slik kobling av ordboken mot kildematerialet.

Den andre påståtte trenden ovenfor er at det vil skje en økt publisering av arkivmateriale i elektronisk form utstyrt med søkeverktøy og et kommentarapparat. Slike utgivelser vil etter all sannsynligvis overta for de store mangebinds kommenterte tekstutgavene fordi de vil være billigere å utgi, kunne nå et større publikum, inneholde mer informasjon og være mer hensiktsmessig i bruk. Nå kan vel neppe *Norsk Ordbok* sies å være en kommentert tekstutgave, men den har klare islett av edisjonsfilologi. Men bakgrunnsmateriale for arbeidet er et arkiv som det vil være interessant å publisere i sin egen rett. Jeg tenker her både på seddelarkivet som eksempler på bruk av ord, men også på alle de ordlister, ordsamlinger og ordbøker som direkte eller indirekte gjennom seddelarkivet brukes i arbeidet med *Norsk Ordbok*. Her ligger det klart i dagen et stort utgivelsesarbeid og venter: en fortrinnsvis kommentert utgave av alle kjente ordlister, ordsamlinger og ordbøker over (ny)norsk og norske dialekter supplert med arkivet til *Norsk Ordbok*. Mesteparten av grovarbeidet er allerede gjort gjennom Dokumentasjonsprosjektet.

Et eksempel på en slik blanding av ordbok og arkiv er "OED On Line" (OED 1996). Lik utgivere av andre gigantverk har Oxford University Press forsøkt å finne alternative utgivelses-

former. "OED On Line" ble i 1996 presentert som en abonnementtjeneste der man får tilgang til ordboken, men også til deler av bakgrunnsmateriale som er brukt. Oxford University Press har også satt i gang et gigantisk revisjonsarbeid av OED, anslagvis 800–1000 årsverk. I 2010 vil det foreligge en fullstendig ny utgave av ordboken. Denne ordboken vil *foreligge bare i elektronisk form* og med muligheter for å se dagens ordartikler og utvalgt bakgrunnsmateriale som et supplement.

En ny strategi for utgivelsen av Norsk Ordbok – metaordboken

Norsk Ordbok-prosjektet ble startet i 1930. Arbeidet går jevnt fremover, men verket vil med dagens hastighet neppe være fullført før om 60 år. Arbeidet er svært tidkrevende. I andre land har det vært alminnelig å bruke mellom 50 og 100 år på tilsvarende verk. *Norsk Ordbok*-prosjektet er således på ingen måte unormalt i sin klasse, og den lange utgivelsestiden kan synes nødvendig. Jeg mener imidlertid at man ikke kan slå seg til ro med at det er nødvendig å vente i 60 år på fullføringen av *Norsk Ordbok*. At det er ressurskrevende å redigere en slik ordbok, er hevet over enhver tvil. Oxford University Press bruker hundrevis av årsverk på sin nye utgave. Men det mangler i Norge i dag en stor ordbok over nynorsk og dialektene. Jeg tror at det vil styrke nynorsksaken og dialektene å ha mest mulig tilgjengelig informasjon. Jeg mener videre at det er viktig å nå yngre mennesker gjennom de samme kanalene der de finner annen informasjon nemlig Internett. Spørsmålet er om en ikke kan kombinere det elektronisk tilgjengelige bakgrunnsmateriale med modern teknologi og metoder hentet fra elektronisk edisjonsfilologi til både å øke produksjonstakten av den tradisjonelle utgaven, men samtidig lage et (elektronisk) verk som kan gjøre stoffet mer tilgjengelig.

Den nåværende redigeringsmetoden er den tradisjonelle sekvensielle gjennomgang av alfabetet der hver artikkel gjøres helt klar for trykk og hvor det er viktig å gjøre artikler mest

mulig stenografisk for at antall sider ikke skal bli for stort. De to største ulempene er den lange tidsperioden mellom første og siste hefte (100 år) og at artiklene kan være tunge å lese for legfolk. Metoder for å presse mest mulig informasjon inn på minst mulig plass har vært et sentralt tema i tradisjonell leksikografi. Til tider kan en spørre seg om en i leksikografien har evaluert denne komprimerte presentasjonen mer ut fra trykkøkonomiske hensyn enn ut fra pedagogiske hensyn. Satt litt på spissen: Det er en stor prestasjon å skrive Koranen på et frimerke, men det er dårlig formidlingskunst. Det er grunnlag for å diskutere om det er riktig å fortsette å bruke en redigerings og utgivelsesmetode som er så tett knyttet til papiret og trykkerkunsten og den skriftlige presentasjonsmetoden slik vi har kjent den helt siden de hellenistiske filologene.

For å unngå misforståelser vil jeg understreke at det ikke et spørsmål om man skal utgi *Norsk Ordbok* i sin nåværende form, men heller om hvorledes utgivelsesprosessen best kan organiseres for at flest mulig nålevende skal få adgang til mest mulig informasjon om norsk på en måte som vil bli brukt. I mitt foredrag i 1996 brukte jeg som et 'memento mori' at det prestisjetunge *Encyclopedia Britannica* hadde en så sterk nedgang i salget av papirutgaven på 1980-tallet og begynnelsen av 1990 tallet at utgiverne besluttet å satse utelukkende på en nettbasert abonnementsordning. Høsten 1999 ble verket til og med relansert som en reklamefinansiert gratis netttjeneste. OEDs tredje utgave kommer som nevnt bare som elektronisk abonnementsutgave. Her i Norge ser det ut til at vårt største konversasjonsleksikon ender som et elektronisk abonnementsprodukt. De bindstore verkene nyter åpenbart ikke lenger den respekt de før fikk bare i kraft av sin fysiske størrelse. De ansees vel rett og slett for å være for uhåndterlig for folk flest.

Er denne dreiningen vekk fra de bindsterke verker et problem, og i så fall hvilke konsekvenser har det for utgivelsen av *Norsk Ordbok*? Man kan vel neppe forestille seg dette eller andre tilsvarende nordiske verk skal ende som reklamefinansiert netttjenester. Til det er nok den tradisjonelle lesergruppen for smal. Men burde man ikke vurdere om redigeringsstrategi kan endres

slik at den bedre utnytter de muligheter teknologien gir oss til å nå et større publikum samtidig som en sikrer det forsvarlig vitenskapelig nivå i redigeringen?

Ovenfor hevder jeg at vi ser en begynnende trend til en elektronisk publisering av kildemateriale med kommentarer og en elektronisk publisering av vitenskapelige essays og monografier sammen med kildematerialet. Kunne man for *Norsk Ordbok* foreta et bevisst valg og publisere verket i bredde først istedenfor som nå i dypde først, det vil si å tilby en ordbok fra A til Å i en gradvis økende finhetsgrad for hver utgave, istedet for å starte på A og lage hver artikkel helt ferdig? Kan man bygge opp en norsk nasjonalordbok ved å flette sammen hovedkildene til et verk for deretter å legge til systematiserte målføreopplysninger, etymologi og definisjoner?

Tanken virker besnærende. Men det kan innvendes at ideen har vært prøvd i 1930-årene under oppstarten av prosjektet *Norsk Ordbok* og i 1990-årene i forbidelse med prosjektet "Dansk Ordbok". Ingen av forsøkene ga særlig oppløftende resultater og planene ble oppgitt. Er det da noen hensikt å prøve på nytt? La oss først se på hva som ble gjort på 30-tallet og hvorfor planen ble oppgitt.

I innledningen til det første heftet av *Norsk ordbok* gir redaksjonen en kort skisse av den originale planen. Her heter det: "I samsvar med den planen som var lagd for ordboka, vart arbeidet skipa såleis at det fyrst vart laga eit grunnmanuskript, der alt tilfanget i Aasens og Ross' ordbøker (med alle tillegga), Schjøtt's "Norsk Ordbok" og sume mindre ordbøker og ordsamlingar vart samanstypt. Samstundes vart ordtydingane overførde til norsk. Grunnmanuskriptet var det så tanken å fylla ut med det nye tilfanget frå målføra og litteraturen, og såleis nå fram til det endelege prentemanuskriptet" (Førebels innleiing til *Norsk Ordbok*, 1950). Grunnmanuskriptet ble også laget og finnes i en maskinskrevet versjon på 13 500 sider. Når man studerer Grunnmanuskriptet, slår det en også at mange av definisjonene bærer preg av de danske originalene, samt at mye av særlig Aasens stringente oppstilling av ordartiklene har gått tapt. Stilistisk er vel heller ikke Grunnmanuskriptet helt på topp. Det

er åpenbart at store deler av manuskriptet måtte ha blitt renskrevet og nyere ord føyd til. Det var dermed ikke store spranget til dagens bruk av Grunnmanuskriptet. Det brukes nå som en viktig kilde når en ny ordartikkel skal skrives.

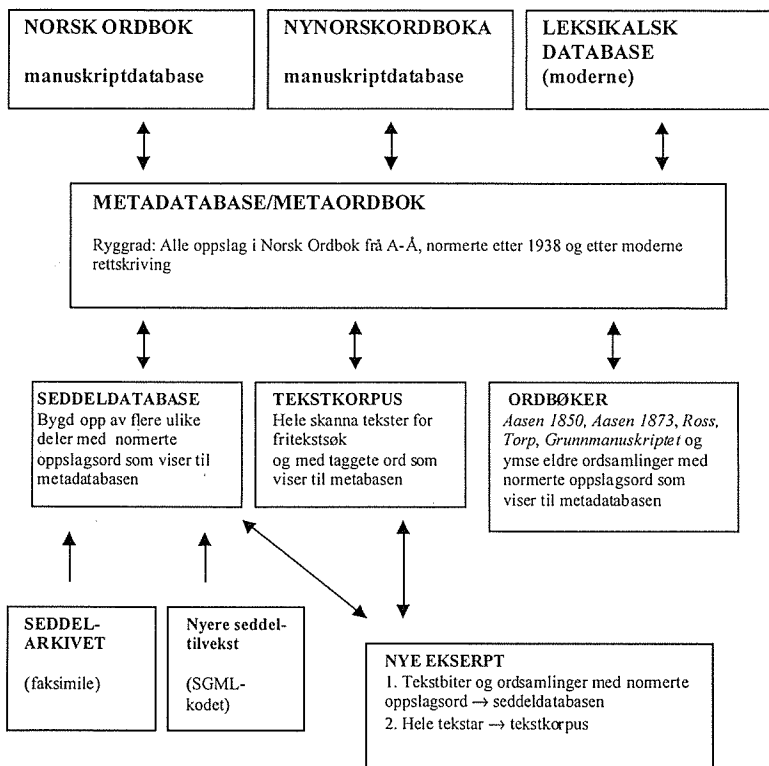
Prosjektet "Den danske ordbog" ble i 1991 (Hjort 1999) ble på sin side planlagt som et gjenbruksprosjekt eller sammenstøpingsprosjekt der man skulle gjenbruke ordboksdata fra ulike ordbøker for hurtig å lage en ordbok. Her kom redaksjonen til den erkjennelse at ordutvalg, uttale og rettskrivning var foreldet og dels feilaktig i forhold til det innsamlede korpusmaterialet. Man ga i følge Hjort opp gjenbruksideen og bestemte seg for å skrive ordboken fra grunnen. Den optimistiske ideen bak "Den danske ordbog" viste seg altså å fungere like dårlig som den tilsvarende tanken bak "Grunnmanuskriptet til Norsk Ordbok" selv med moderne datateknologi.

I tillegg til disse to eksemplene på mislykket "resirkulering" av leksikografisk materiale synes det også å være en god del skepsis eller usikkerhet rundt bruken av datateknologi i leksikografisk arbeid. I *Nordisk leksikografisk ordbok* presiseres det under oppslaget "leksikalsk database" at "slike programmer [knyttet til en leksikalsk database] innebærer selvsagt ikke at ordboksarbeidet blir automatisert. Tvert imot er leksikografens aktive innsats svært viktig". Videre formulerer en stor kapasitet som Lars Vikør at "...for hovedsaka for oss [NOs redaksjon] er det intellektuelle grunnarbeidet, som ikkje all verdas datateknologi kan rasjonalisere bort om det skal bli ei ordbok" (Vikør 1999).

Det er altså veldokumentert skepsis til å bruke eldre ordboksmateriale til å lage nye ordbøker særlig ved hjelp av datamaskinelle metoder. Med metaordbokideen griper man imidlertid saken noe anderledes an enn i de nevnte eksempler. Metaordboksiden springer ut fra de nye trendene i edisjonsfilologien og vil nok ikke være egnet til arbeidet med en moderne definisjonsordbok. Den er ment som et hjelpemiddel til å gradvis kunne publisere materialet det Lars Vikør kaller en "flergenerasjonsordbok" (Vikør 1999) eller Matti Vilppula kaller enn "evighetsordbok" (Vilppula 1999). Som nevnt tidligere, har det de siste 5-

6 årene begynt å vokse frem en forståelse for at datateknologiens tilnærmet ubegrensede lagringskapasitet kan brukes til å publisere (elektronisk) grunnlagsmaterialet for et vitenskapelig prosjekt sammen med den vitenskapelige publikasjonen. Dette kan være interessant dels fordi man da kan se hva forskerne, i vårt tilfelle ordboksforfatterne, har brukt som grunnlagsmateriale. Dette gjør det mulig for andre for å forske i det samme materialet eller etterprøve resultatene. Denne tanken har særlig fått aksept innen edisjonsfilologien i de senere årene. Det store Henrik Ibsen-prosjektet som startet i 1998, er lagt opp på denne måten. OED har i sin nettversjon lagt inn en mulighet til å søke i bakgrunns materialet. Tanken om å gjøre noe tilsvarende for *Norsk Ordbok* ble lansert allerede våren 1996 i et foredrag jeg holdt på et jubileumsseminar for Norsk Ordbok i Oslo, se Ore 1997.

Ryggraden i metaordboken vil være en liste med normerte (oppslags)ord. Den gjør det mulig å skaffe seg oversikt over ordtilfanget og å planlegge størrelsen på *Norsk Ordbok*. Ved hjelp av denne ryggraden vil det også være mulig å henge på de ulike kildene: Dette vil i og for seg svare til sammenstøpings-tanken fra 1930 tallet. Men med "sammenstøping" i forbindelse med Grunnmanuskriptet mente man en sammenskriving og oversettelse av høyst ulike forfatteres verker. Dette er en vanskelig oppgave dersom man ikke kan stoffet svært godt og klarer å frigjøre seg fra språket i de ulike originalene. "Sammenstøping" eller påhekingen av originalkilder i metaordboken vil være en helt annen ting. Alle de originale kildene vil være med komplette, i sin opprinnelige form. Metaordboken vil i sin mest basale form bestå i at denne samlingen av kilder får felles inngang gjennom grunnordene. For hvert grunnord vil man få tilgang til alle de ordartikler eller avsnitt der ordet er behandlet eller forekommer. I tillegg vil det selvfølgelig være andre søkemuligheter som datateknikken gir oss så som tid, sted, type og generelt fritekstsøk. Dette er selvfølgelig ikke en ordbok – det er mange. Men i motsetning til Grunnmanuskriptet danner den et fundament for en ordbok. Det trenges i seg selv ikke renskrives til et trykkmanuskript.



FIGUR 2. Skjematisk fremstilling av metaordboken.

Å sette opp listen over ønskede oppslagsord for *Norsk Ordbok* er imidlertid ingen triviell oppgave. Som et pilotprosjekt er det i 1999 satt av 9 månedersverk til dette. Pilotprosjektet går ut på å gå gjennom det elektroniske seddelarkivet fra og med bokstaven *i* og etablere en foreløpig oppslagsordliste på bakgrunn av oppslagsordene på sedlene. Pilotprosjektet er ennå ikke avsluttet: Men arbeidet er interessant også fordi det setter de eksisterende redaksjonsreglene på en hard prøve og avslører problemer man ikke har tenkt på eller ikke ønsket å ta opp (ennå) i den tradisjonelle A-til-Å-redigeringsmetoden. Vi håper nå å få finansiering i 2000 og 2001 til å fullføre oppbyggingen av denne.

Ordartiklene i *Norsk Ordbok* slik den i dag redigeres, vil komme i tillegg til metaordboken og vil eksistere som et eget separat verk i parallell med sine kilder. Ordartiklene kan nå skrives i det omfang og i den rekkefølgen som er ønskelig. Metaordboksprosjektet gjør det også mulig å benytte maskinenes ubegrensede lagringskapasitet til å skrive ikke fullt så kompakte ordartikler. Man kan altså legge noe mer vekt på formidlingsaspektet og ikke så mye på antallet spaltemillimeterer man har til disposisjon. Det er også mulig for en enkelt redaktør å vie seg til enkeltfelter som målføreopplysninger eller etymologi. Det bør være mulig å kunne vurdere utvalget av opplysninger i den trykte versjonen av *Norsk Ordbok* idet man kan lettere henvide til den elektroniske metaordboken for supplerende opplysninger og/eller legge inn for eksempel detaljerte dialektopplysninger i en senere utgave. En kan dermed øke tempoet gjennom alfabetet slik at en (første) gjennomgang kan gjøres i løpet av en 10–15 års periode.

Det store prosjektet "Norsk Ordbok" kan dermed deles opp i overkommelige delprosjekter som utgjør noen titalls årsverk hver. Et avsluttet delprosjekt vil resultere i en ny og større utgave. Men i motsetning til dagens situasjon vil hver utgave vil dekke hele alfabetet. Ordboken vil på denne måten vokse frem og alltid være temmelig ajour. Denne gradvise utviklingen av ordboken vil dermed være i samsvar med det den skal beskrive, nemlig et levende språk som hele tiden endres gjennom bruk.

Henvisninger

- Atkins, S. 1992. "The Hector Project." I: *Proceedings of Complex '92*. Budapest.
- Goldfarb, C. 1991. *The SGML Handbook*, Oxford University press.
- Delary P. and Landow G. (eds.). 1991. *Hypermedia and Literary Studies*. The MIT Press, London.
- Nordgaard T. 1995. "NordKompLeks – et norsk komputasjonelt leksikon." Prosjektbeskrivelse. Trondheim.

- Oxford English Dictionary On Line*. 1996. Oxford University Press, <http://www.oed.com>.
- Ore, C.-E. 1991. "Dokumentasjonsprosjektet ved Det historisk-filosofiske fakultet, Universitet i Oslo." I: *Nordiske studier i leksikografi*, konferanserapport, Oslo.
- Ore, C.-E. 1996. "Korpus og seddelarkiv, fredelig sameksistens mellom det beste og det gode?" I: *Nordiske studier i leksikografi*. 3. Reykjavik.
- Robinson P. (ed.). 1996. *The Wife of Bath's Prologue*. Cambridge. University Press, Cambridge.
- Sperberg-McQueen, C.M., Burnard, L. (eds.). 1994. *Guidelines for the Encoding and Interchange of Machine-Readable Texts (TEI P3)*. Chicago and Oxford.
- Vanhoutte, E. 1999. "Where is the Editor? Resistance in the creation of an electronic critical edition." I: *Human IT* nr 1, 1999. Institutionen Bibliotekshögskolan, Högskolan i Borås.
- Vikør, L.S. 1999. "Fleirgenerasjonsordbøker og tida." I: *Nordiska studier i leksikografi*. 4. Helsingfors.
- Vilppula, M. 1999. "Ordbok över finska dialekter och evigheten." I: *Nordiska studier i leksikografi*. 4. Helsingfors.

Riitta Palkki

Lagring av ordbok över äldre finskt skriftspråk i strukturerad form

The Dictionary of Old Finnish (VKS, *Vanhan kirjasuomen sanakirja*) is written using a structural program, FrameMaker+SGML (SGML = Standard Generalized Markup Language). The program is used in Unix environment. A Frame document consists of parts called elements, which are hierarchically organized in fields according to the Document Type Definition (DTD). In one field there can be several parallel or subordinate fields. The basic structure of the articles in VKS is as follows: index word with its variants, word class code, definition, data from old dictionaries, examples with sourcecodes, list of compound words. When writing an article for the VKS, several windows are needed. The window "Elements" contains all the elements of the dictionary. The "Structure View" shows a graphic outline of the document's structure. The Document window is used when writing the text. Other windows, "Terminal" and "Text Editor", which do not belong to the Frame program, are also available. The written text can also be viewed on the screen in SGML format. When finished, the dictionary can be used in research for example as a text corpus.

Bakgrund

Perioden för *äldre finskt skriftspråk* började med *Mikael Agricolas Abc-bok* ungefär år 1540 och räckte till slutet av svenska tiden, dvs. ungefär till år 1810. Materialet för *Ordbok över äldre finskt skriftspråk* (*Vanhan kirjasuomen sanakirja*, i fortsättningen kallad VKS) utgörs av tryckalster från den tiden och omfattande handskrifter, till exempel finska översättningar av lagar, som senare har blivit tryckta.

Ordbokens första del (bokstäverna *a* till *i*), som utkom år 1985, skrevs på gammaldags skrivmaskiner. Den andra delen (bokstäverna *j* till *k*) skrevs med ordbehandlingsprogrammet *WordPerfect* och kom ut år 1994. När arbetet på den tredje delen inleddes övergick vi samtidigt till ett program för strukturerad redigering, *FrameMaker+SGML* (i fortsättningen kallad *Frame*).

Varför Frame?

Frame fungerar i *Unix*-miljö, så redaktörerna har terminaler anslutna till en *Unix*-centralenhet. Bildskärmen fungerar som *arbetsbord*. Programmen öppnas som s.k. *arbetsfönster* på arbetsbordet. Utöver Frame erbjuder *Unix* flera användbara hjälpprogram för ordboksarbete. När det gäller VKS är det främst *Terminal* och *Text Editor* som är viktiga.

En strukturerad redigering av VKS diskuterades för första gången redan för ett tiotal år sedan. Då var ordboken exempelprojekt i ett forskningsprogram som gällde strukturering av texter och behandlingen av strukturerade texter. En av ledarna för projektet var *Erja Nikunen*, dåvarande chef för adb-avdelningen vid Forskningscentralen för de inhemska språken. Hon utarbetade en s.k. *kontextoberoende grammatik* för VKS, som i hög grad påminde om strukturbeskrivningen för Frame, och ordboken blev av denna orsak pilotprojekt för Frame vid forskningscentralen.

Det fanns flera praktiska skäl till att det nya systemet togs i bruk. Forskningscentralens *textkorpus* med äldre finskt skriftspråk omfattar över 2,5 miljoner ord, och en betydande del av nyckeltexterna i ordbokens stora källmaterial föreligger i elektronisk form. Det var därför viktigt att göra det så lätt som möjligt att utnyttja korpusmaterialet. En annan orsak var att man ville att redaktörerna skulle slippa tänka på sådana tekniska detaljer som maskinen sköter felfritt. Det var vidare önskvärt att ordboken skulle få en sådan form att den å sin sida skulle gå att utnyttja för så mångsidiga sökningar som möjligt. Av allt detta följde att artiklarnas struktur måste göras mer konsekvent än tidigare. När man skrev på maskin eller wp behövde så stor uppmärksamhet inte fästas vid att strukturen var absolut logisk.

Lagring av ordboken på Frame

Ett Frame-dokument består av delar som kallas *element*. Dessa ordnas i enlighet med den logiska *strukturbeskrivningen* för en ordboksartikel som *fält* som ligger i följd, jämsides eller innanför

varandra, så att de tillsammans bildar textens stomme. Med tanke på SGML (förkortning för *Standard Generalized Markup Language*) talas det om en DTD (*Document Type Definition*), en *strukturgrammatik*, som klargör vilka element som är tillåtna och i vilken ordning dessa kan användas (se t.ex. Lehtinen / Karvonen / Rahikainen 1995:55). Strukturbeskrivningen för VKS utarbetades ursprungligen av Erja Nikunen. Redaktörerna har i över två års tid prövat och finlipat strukturen tillsammans med forskningscentralens adb-expertter, och nu börjar det fungera någorlunda felfritt. Strukturbeskrivningen har lagrats i *master*-filen, som tas fram som underlag för varje nytt dokument.

1 2 3 4

hyypiö, -ä s. huuhkaja (S *hyypiä* bubo, buff, der
 Franck | VR *hyypiö* bubo, wff | F *hyypiä* = VR |
 J, JllsK *hyypiä* aboens. ulula, uf; lisäksi J *hyyp-*
piö | G 223a *hyypiä, hyypiö, hyypiä* aboens. uuf,
 ulula; stenuggla; ~ *huhkaja*) | A III 359 (B¹⁻⁴ Ps 5
 102:7) Mine olen ninquin iocu Hypie njisse häuite-
 tyis Caupungeissa (A I 300 hwhkaija) | ib. 707 (B¹
 Sef 2:14) Cacarit ia Hypiet pite heiden Tornisans
 asuman (B²⁻⁴ tarhapöllöit) | B¹⁻⁴ 3 Ms 11:17 ei
 teidän pidä syömän [lintuja], kuin owat -- Kiwi-
 rautia, Luicoi, Hyypiä | B¹ 5 Ms 14:16 Hyypiö,
 Riutahyypiö, Yöcköi (B² Linduhaucka, Hyypiö,
 Yöcköi, B¹⁻⁴ Huhkaja, Hyypiö, Yöcköi) | Pet-
 raeus 1649 A3a *hyypiä, hyypiän, bubo* | Pacch
 1667 N3b [Monien pitää olla] niijn quin Hyypiän
 nijsä häwitetyisä Caupungeis | RWL 1759 106
 (RWL 1808 112) haucan, *hyypiän, cockolinnun, ja*
muita raatelewia linduja, mahta joca mies syytöin-
nä ambua (Forseen 1738 137 tarhapöllön) | Ran
 1803 35 Se pyhä maa ei ole -- kärmetten koto eikä
 hyypiän pesä. — *Yhd. mäki-, riutahyypiö.*

6

BILD 1. Strukturen för ordartikeln i VKS.

Strukturen för ordartiklarna i VKS är i huvudsak följande (som exempel ordet *hyypiö* (berguv, *Bubo bubo*) (bild 1): (1) *uppslagsordet* med varianter; (2) *ordklassangivelse*; (3) *betydelsebeskrivning*; (4) uppgifter ur *gamla ordböcker*; (5) *exempel* med källhänvisningar; och (6) förteckning över *sammansatta ord*. Det finns många variationsmöjligheter, och en ordartikel behöver inte nödvändigtvis innehålla alla dessa delar. Ordklassangivelse och

turdiagram, som åskådliggör den inbördes hierarkin mellan artikelns delar. Redaktören väljer ett element och placerar det på rätt plats i strukturdiagrammet. De lodräta strecken visar på vilken nivå i diagrammet man rör sig. Elementförteckningen förändras beroende på var i strukturdiagrammet eller artikeln man befinner sig. Överst i förteckningen finns alltid de element som är de rätta på det aktuella stället. Programmet förhindrar inte att man väljer fel element, men ett felaktigt val syns i strukturdiagrammet som en streckad röd linje. En röd fyrkant i diagrammet visar att något element saknas.

Stora fönstret till vänster är *dokument*-fönstret, det egentliga skrivfönstret. I översta balken finns namnet på dokumentet och uppgifter om i vilken katalog det finns. Strukturtaggar eller elementhierarki syns i allmänhet inte i detta fönster. I ordboken används specialtecken som inte finns på tangentbordet, och dem tar man fram med musen i det lilla fönstret *Frame*. *Terminal* och *Text Editor* är hjälpfönster som inte hör till *frame*. I *Terminal* ges kommandon, bl.a. när det gäller korpussökning. Resultaten av sökningen sammanförs i allmänhet till en fil, och det är lätt att bearbeta dem i texteditorn.

Några exempelartiklar

Som exempel tar jag artikeln *mela* (bild 2). Först väljer vi elementet *sana-artikkeli* (ordartikel). Vi får då fram ett fönster där artikelns längd ska bestämmas. Det vanliga värdet är *#lyhyt#* (kort). Om artikeln kommer att innehålla flera betydelsemoment och bli så lång att den måste indelas i stycken väljer vi här alternativet *#pitkä#* (lång). Valet leder till att elementet *hakusana* (uppslagsord) framträder på strukturdiagrammet. Uppslagsordet *mela* skrivs nu i dokumentfönstret. Sedan väljer vi ordklass (*sanaluokka*), här alltså *s.* (substantiv). I VKS intar exempelfraserna en nyckelställning. Beläggen för exempelfraser på ordet *mela* visar att ordet inte är entydigt, och artikeln måste därför indelas i betydelsemoment. Då väljer vi alltså i detta skede *vanhat leksikot* (gamla lexikon). Uppgifterna ur varje ordbok skrivs

som ett eget lexikonexempel (*leksikkoesimerkki*). Det finskspråkiga materialet markeras med taggen *korostus* (markering) och blir kursiverat.

Sedan går vi över till ordets betydelser. Då vi väljer elementet *merkitysryhmä* (betydelsemoment) får vi automatiskt beteckningen *1*. När samma element väljs på nytt får vi *2* osv. Efter siffran följer betydelsebeskrivningen (*selite*); vi väljer element och skriver in beskrivningen i dokumentfönstret. Därefter följer exempelblocket (*esimerkkilohko*) med exemplen i kronologisk ordning, dvs. så att det äldsta belägget kommer först. För att hitta exempel gjorde jag en korpussökning med programmet *egrep*. Söksträngen syns i terminalfönstret. Den teckensträng som jag sökte var *'mel[ao]'*. Jag skrev *[ao]* därför att *a* i vissa former av ordet *mela* ersätts av *o*. Resultatet av sökningen fick utgöra en fil i min egen hemkatalog, och denna fil tog jag in i texteditorn för bearbetning. På bilden syns bara en del av resultatet av sökningen. Endast Mennander har använt ordet i den egentliga betydelsen, 'styransordning i båt'. Frosterus har använt det för att beskriva en fågelstjärt. Hos Aken avser det ett paddelliknande redskap att röra om med. Korpusexemplen har mekaniskt förts in i artikeln med hjälp av musen. Exemplet från Kettu har tagits ur det gamla ordsedelmaterialet. Sådana exempel skriver redaktören själv in i texten och granskar samtidigt ortografin.

Det sista elementet i artikeln är förteckningen över sammansatta ord (*yhdys-sanaluettelo*). Här finns det bara ett sammansatt ord, *mäskimela*.

Uppslagsformen av *mela* är entydig. Av många andra ord finns det desto fler varianter (bild 3). Till exempel i artikeln *melkein* ingår den nuspråkliga uppslagsformen *melkein* och dessutom formerna *melkeen*, *melke*, *melkeinkin*, *melkäinkin*. När det gäller formen *melkeen* är det endast slutet som avviker från huvuduppslagsordet. Då väljer man elementet *päätevariantti* (ändelsevariant) och skriver *-en*. Man får inte glömma lodstrecket som avskiljer ändelsen. Programmet för in kommatecknet. De övriga formvarianterna skrivs sedan som *variantelement* (*variantti*); programmet ger automatiskt parentes och tilde. I elementförteckningen ingår också elementet *vanha muoto* (gammal form).

Så kallade varianterna nämligen om nuspråkets form av ordet, uppslagsformen, inte alls har förekommit. Ett exempel på detta är artikeln *melkeittäisin*: i alla exempelfraserna står det *melkeettäisin*. Varianter och gamla former skrivs in i ordboken som hänvisningsartiklar (*viiteartikkeli*) på sin egen plats i alfabetet, till exempel *melke* och *melkee-*. När man väljer elementet *viitehaku-sanat* (hänvisningslemman) lägger programmet automatiskt in en hänvisningspil.

The screenshot shows a FrameMaker SGML Structure View. The left pane displays a text document with the following content:

1 #Lyyhyt melata v. J brevi remo utor, brukar styrären | JlisP 210 pian
 2 soottu soucka salmi. melattu (melattu) matala meri prov. | G II 167 a - J;
 3 myndär, håller styr på båten, guberno cymbam. - melata, meloita.¶
 4 #Lyyhyt melke adv. -> melkein.¶
 5 #Lyyhyt melkee -> melkei.¶
 6 #Lyyhyt melkein, -en (- melke, melkein, melkänkin) ¶
 7 #Lyyhyt melkeittäisin (melkeettäisin) adv. suurin piirtein, jotenkin,
 8 jokseenkin | Liz 1756 56 Carl Fleischer toisi melkettäisin lukea Kirjaa
 9 sisältää, ja lupais nyt virittää opetella ulcoa | Wirz 1768 735 Kusa
 10 orjantappurat joukkoittain kasowawat, siinä on melkettäisin hedelmätöin
 11 maa.¶
 12 #Lyyhyt melkeä, -iä a. (J melkeä aliguantus, tämlig | JlisP 210, JlisK
 13

The right pane shows the SGML structure view with the following elements and their attributes:

- Hukusana melke
- Sivupää PerusSana
- Sanalukka adv.
- Viitehaku-sanat melkein
- SanaArtikkelit [] melkee-
 - APLus ID = Elyhyt
 - ID = eno vales
- SanaArtikkelit []
 - APLus ID = Elyhyt
 - ID = eno vales
- Hukusana melkein
- Sivupää PerusSana
- Pääkirjain -en
- Vainnit melke, melkein, melkänkin
- SanaArtikkelit []
 - APLus ID = Elyhyt
 - ID = eno vales
- Hukusana melkeittäisin
- Sivupää PerusSana
- Vainnit melkeittäisin
- Sanalukka adv.
- Selin suurin piirtein, jotenkin, jokseenkin
- Esimerkki [] Liz 1756 56 Carl Fleischer lätin

BILD 3. Varianter, gamla former och hänvisningsartiklar.

Pronomenet *me* (vi) är ett bra exempel på en lång artikel. På bilden (bild 4) ses strukturen i huvuddrag. Trots att exempelfraserna har utelämnats är det bara en del av artikeln som har fått plats. Indelningen börjar med övermomentet (*suurmerkitysryhmä*) *I*; sedan kommer betydelsemomenten med arabiska siffror och undermomenten med små bokstäver. Det finns också avdelningar för användningsområden och sublemman på olika nivå.

Till höger på samma bild finns texten i SGML-form, uppsjälkad i element eller fält, och taggarna är som synes många. I

VKS är taggarna hela ord och alltså mycket åskådliga. Starttaggen för ett fält pekar till höger och sluttaggen till vänster. Det kan finnas flera fält innanför varandra, och inuti ett fält kan det finnas flera jämbördiga fält. På bilden syns betydelsemomentet *l* 2 ur artikeln *me*. Inom den finns användningsområdet (*käyttöalaryhmä*) *erik*. (specialanvändning), som har indelats i undermomenten *a*, *b* och *c*. Inom *c* finns en sublemmagrupp (*alihakuryhmä*), som innehåller ett sublemma (*alihakusana*), dvs. *meitin*, en förklaring (*selite*) och ett exempelblock (*esimerkkilohko*), som alla är på samma nivå i hierarkin. Under *b* finns en textjämförelse (*tekstivertailu*) inom ett exempel (*esimerkki*). Längst ner finns sex sluttaggar efter varandra, först taggen för det innersta fältet och sist taggen för det yttersta fältet. Det är lätt att ta fram strukturfönstret på arbetsbordet, och man har stor hjälp av det om man misstänker att man har gjort något fel. När man skriver en artikel behöver man naturligtvis inte hålla hela strukturhierarkin i minnet; det räcker med att man har överblick över det avsnitt som man håller på med.

Bearbetningen och finslipningen av artiklar

En artikel som är skriven i strukturerad form är lätt att bearbeta. När man gör ändringar har man nytta av strukturdiagrammet som visar på vilken nivå man arbetar. Om ett element flyttas följer alla element med som finns lägre i hierarkin. Det är därför också lätt att flytta hela artiklar. Ibland måste man flytta betydelsemomenten, och då sörjer programmet automatiskt för att numreringen blir riktig. Innan artiklarna trycks måste naturligtvis taggarna och andra överflödiga detaljer avlägsnas, och texten måste bearbetas för tryckning. Denna finslipning sköter programmet om. Som exempel på färdiga artiklar finns här artikeln *kummitus* (bild 5) ur ordbokens andra del som redan har kommit ut.

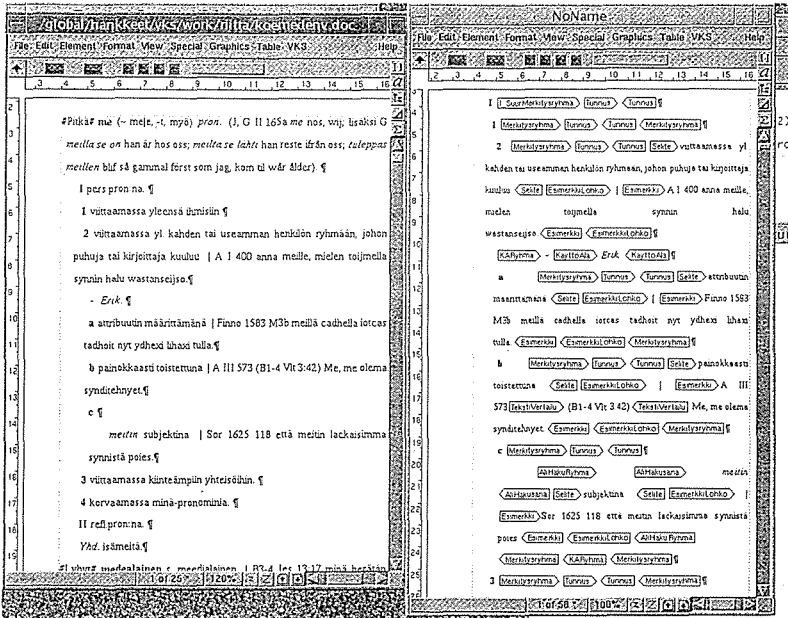


BILD 4. Pronomenet me som exempel på en lång artikel.

kummitus s. (JlisP 62 spöke, Sotk. | G 496b spö-
kande, spökerin ~ kummitteleminen). I outo ja
selittämätön olento tai ilmiö, aave; myös kuv. |
Fors 1704 83 [Maailman menot] silmisiä sinulla
ovat culla kijldäwäinen, Kijldäwäiset cummi-
tuxet | Weg 1747 867 cummituxet ja Piruin näöt
enimmitten aina näkywät Kirkoig aidoisa | Weg
1749 953 taita hän sairautten parannuxisa ja mui-
sa cummituxisa paljon ihmedeldäwiä asioita toi-
mitta | SAchr 1766 20 Caicki cuollon cummituxet
Jätti julkinen Jumala | JAFro 1791 28 senkaldai-
set [tulet metsässä] ei ole mitkän cummituxet,
mutta heillä owat heidän luonnolliset syynsä.
2 tarpeeton, kauhistuttava teko tai ilmiö, hullu-
tus, turhuus | GJCal 1734 16 [Maanpetturij] Teki
wisi willitystä, cuusi muuta cummitusta | Weg
1747 esip. 9 wierasten waatetten ylöllisyydesä ja
muodottomisa keweisä cummituxisa ja waihetuxi-
sa ja muusa lihan menosa | ib. 415 [Velhot] Isä
meidän Rucousta edes tacaisin lukewat? ja mon-
da muuta edes tacaista cummitusta sijnä edesäns
pitäwät. — Yhd. ihmekummitus.

BILD 5. Ordet kummitus ur ordbokens andra del.

Till slut

En ordbokstext som har lagrats i strukturerad form är en ypperlig källa för fortsatt forskning. Som undersökningsobjekt kan man välja vilket enskilt element eller vilken elementkombination som helst. Man kan till exempel plocka ut enbart uppslagsorden eller betydelsebeskrivningarna eller varför inte alla verbartiklar. Möjligheterna är många, och det är viktigt. Forskningsarbetet blir ju avsevärt lättare och ändamålsenligare när det finns material och program som kan bidra med detaljerad och mångsidig information.

Litteratur

A. ORDBÖCKER

VKS = *Vanhan kirjasuomen sanakirja*. Kotimaisten kielten tutkimuskeskuksen julkaisuja 33. 1. Päätoim. Esko Koivusalo. Helsinki 1985. 2. Päätoim. Maija Länsimäki. Helsinki 1994. Valtion painatuskeskus.

B. ÖVRIG LITTERATUR

- Adobe FrameMaker+SGML*. Using FrameMaker+SGML. 1996.
 – Quick Reference. 1996.
 – Getting Started. 1996.
- Hietanen, Risto / Syrjänen, Seppo. 1995. *Unix-käyttäjän alkeisopas*. 3. painos. Helsinki: Yliopistopaino.
- Lehtinen, Marja / Karvonen, Pirjo / Rahikainen, Tarmo. 1995. *Tekstikorpuksset*. Kotimaisten kielten tutkimuskeskuksen julkaisuja 81. Helsinki.
- Nikunen, Erja. 1988. *Vanhan kirjasuomen sanakirjan rakenne*. Moniste.
- Nikunen, Erja / Rahikainen, Tarmo / Lounela, Mikko. 1998. *Vanhan kirjasuomen sanakirjan rakennekuvaus*. Uusin versio elokuu.

Vanhan kirjasuomen sanakirjan artikkelinkirjoitusohjeet 1977.

Toim. Esko Koivusalo / Irene Hellemaa / Riitta Suhonen /
Anja Lehtimäki. Helsinki.

Vanhan kirjasuomen sanakirjan toimitusperiaatteet 1997. Toim.
Riitta Palkki. Moniste.

Översättning från finska av Monica Äikäs.

Bo Ralph

Orden i ordning

Den historiska framväxten av en lexikografisk tradition i Sverige

För några år sedan, 1995, fick den isländske författaren Einar Már Guðmundsson Nordiska rådets litteraturpris, huvudsakligen för romanen *Englar alheimsins*, som på svenska har fått titeln *Universums änglar*. När han höll sitt tacktal gjorde Einar Már såsom islänningar och finnar brukar i nordiska sammanhang. De avstår ju från att använda sina egna modersmål, eftersom vi andra är så okunniga, och anpassar sig till den nordiska språkliga majoriteten. Einar Már kom naturligtvis in på Norden och det nordiska samarbetet och sammanfattade sin syn på läget med följande kärnfulla påstående: "Norden er i orden!" Detta uttalade han, som den islänning han är, med mycket tydligt rullande r och korta å-ljud på båda ställena, ungefär: "Nårdden er i årdden".

Nu var detta ett tal som framfördes muntligt, såsom brukligt är, och därför borde det kanske egentligen inte vara så mycket mer att tillägga. Men hade Einar Már på författares normala sätt presenterat den sentensartade formuleringen endast i skrift, så hade den kunnat bli föremål för vidare filologisk analys. Den skrivna strängen "Norden er i orden" kan till nöds läsas med slutna o-ljud – och med en vridning av innebörden: /nu:rdən ɛ:r i u:rdən/, i grov fonematisk återgivning; framför allt faller detta sig naturligt för en svensk. Det uttalet inbjuder i sin tur till en filosofisk fundering, som kanske leder fram till en parafras i stil med: "Det är i orden vi finner Norden" e.d. För en del år sedan var några forskare vid vår institution i Göteborg involverade i en bok som heter *Orden speglar samhället* (Allén, Gellerstam & Malmgren 1989), vars syfte var just att demonstrera relationen mellan ordförråd och omvärld. Vid en nordisk konferens om lexikografi kan det väl vara klädsamt att inledningsvis konstatera såväl att

”Norden är i ordning” som att ”Norden är i orden”, dvs. att ”orden speglar Norden”.

Nästa tanke blir givetvis då att perspektivet lätt kan vidgas till ett globalt påstående med lydelsen ”Jorden är i orden”, eftersom vi ju är väl medvetna om att den s.k. encyklopediska världen generellt återspeglas i språket. Dessvärre är det betydligt svårare att hävda, i samklang med Einar Már, att ”Järden er i ården”!

Däremot vill vi kanske gärna – återigen som lexikografer – att ”orden er i orden”, och därmed är jag framme vid mitt egentliga ämne: ”Orden i ordning. Den historiska framväxten av en lexikografisk tradition i Sverige”. Jag vill då gärna skynda mig att påpeka att valet av svensk lexikografihistoria inte är ett utslag av någon allmän storsvensk imperialistisk ambition, även om jag senare kommer in på detta ämne – det är helt enkelt en nödvändig inskränkning framtvungad av mina starkt begränsade kunskaper.

Lexikografisk tradition

Uttrycket *lexikografisk tradition* kan uppfattas på åtminstone två olika sätt. I det följande avses i första hand ’lexikografisk verksamhet alltifrån äldsta tid’, alltså en ganska allmän innebörd, som är den ena möjliga. En annan, mer specifik, som vi kan återge med ’gängse lexikografisk beskrivningsteknik’, får här stå tillbaka.

I mitten på 1980-talet tog jag initiativet till ett projekt som var avsett att åtminstone översiktligt kartlägga den svenska lexikografins historia. Det var då tämligen jungfrulig mark. Adolf Noreen gällde som den stora – om inte enda – auktoriteten genom översikten i *Vårt språk* (1903; se även Holm & Jonsson 1990:1933). Projektet kom i gång så smått 1986, i större omfattning året därefter, då Sven-Göran Malmgren och jag tillsammans fick ett anslag från Humanistisk-Samhällsvetenskapliga Forskningsrådet. Projektet har strängt taget hållit på sedan dess, i växlande omfattning, och nu vet vi betydligt mer om detta spännande ämne. Så sent som hösten 1998 kom en doktorsavhand-

ling av Anna Hannesdóttir, *Lexikografihistorisk spegel. Den enspråkiga svenska lexikografins utveckling ur den tvåspråkiga*, som spänner över vida fält, men som framför allt utreder den på 1800-talet slutgiltigt etablerade enspråkiga lexikografins historiska beroende av den tvåspråkiga. Anna Hannesdóttir bidrar själv till denna volym på annan plats.

Våren 1998 disputerade Lena Rogström på en avhandling som beskriver *Jacob Serenius lexikografiska insats*. Serenius var egentligen mest känd som biskop och engagerad politiker, men för oss är han framför allt den som lade grunden till den engelsk-svenska och svensk-engelska lexikografen. Serenius var i sin ungdom legationspräst i London och har betytt mycket för kontakterna mellan England och Sverige även bortsett från de tre ordböckerna av hans hand (Serenius 1734, 1741, 1757). Lena Rogström visar att många svenska ord som idag är självklara inte har några kända belägg förrän i Serenius ordböcker. I den mån dessa förstabelägg är äkta, tycks biskopen med andra ord ha hållit igång sin egen lilla skapelseprocess. Förutom att han direkt kan ha berikat det svenska ordförrådet, har genom hans förmedling brittisk och fransk lexikografi tidigt kommit att lämna bidrag också till den svenska beskrivningstraditionen, dvs. genom de utländska arbeten som Serenius använt som förebilder för sina ordböcker. Serenius ordböcker var i sin tur länge inflytelserika både inom och utom landet (bl.a. har de använts inom dansk 1700-talslexikografi). Lena Rogström kommer in på förstabeläggen och deras betydelse i sin presentation i denna volym.

Monica Johansson lade våren 1997 fram sin doktorsavhandling *Lexicon lincopense. En studie i lexikografisk tradition och svenskt språk vid 1600-talets mitt*. Den behandlar den första latin-svenska ordboken i verklig mening. Vi återkommer mer till denna ordbok i det följande.

Förutom dessa monografier på sammanlagt 1400 sidor har ett antal studier i uppsatsform publicerats av de nämnda tre projekt-deltagarna och andra, främst Sven-Göran Malmgren, Kerstin Norén, Lars Wollin och mig själv. Det är omöjligt att exakt redovisa vad som skett inom projektets ram i strikt mening och vad som mer berott på fri företagsamhet, men sammantaget har

ganska många aspekter av den svenska lexikografihistorien fått sin belysning. Kerstin Norén har mest sysslat med metodfrågor och de senare skedenas lexikografi, Lars Wollin har koncentrerat sig mer på inledningsstadiet och undersökt latintraditionen, i synnerhet ett handskrivet latin-svenskt glossar från 1400-talet, medan Sven-Göran Malmgren och delvis jag själv varit inne på bl.a. den tysk-svenska och svensk-tyska lexikografin. Malmgren har också undersökt samspelet mellan grammatisk och lexikalisk beskrivning, vilket i synnerhet även Anna Hannesdóttir har gjort. Det kan slutligen nämnas att en tills vidare avslutande samlingsvolym med olika kortare studier för närvarande planeras. Därför, och av platskäl, ges här ingen fullständig bibliografisk redogörelse för projektets skriftliga produktion.

Vi får här inskränka oss till denna blixtsnabba rapport från ett mer än tioårigt projekt – den som är speciellt intresserad av olika enskildheter hänvisas till själva skrifterna med vidare referenser. Låt mig bara understryka att Noreens och andra tidigare auktoritetens uppgifter visat sig diskutabla eller rentav felaktiga på åtskilliga punkter. Jag skulle emellertid nu vilja fortsätta med att anlägga ett mer principiellt perspektiv på lexikografihistorien. Till att börja med kan man fråga sig vad en kartläggning av lexikografins historia skall tjäna till.

Man kan naturligtvis alltid svara att ett historiskt förlopp i sig säkert kan vara intressant att följa, men många nöjer sig inte med det svaret. Det är antagligen fler som accepterar att man vill fylla på det idé- och lärdomshistoriska vetandet i allmänhet med lexikografins utveckling i synnerhet. Vidare brukar specialister på språkhistoria föra en ojämn kamp med bristfälligt källmaterial, och de välkomnar naturligtvis kvalificerade bedömningar av äldre ordböckers tillförlitlighet som språkhistoriska källor. Men även den lexikografiska disciplinens egen verksamhet kan dra nytta av lexikografihistorien. Vi har vant oss vid att finna vissa typer av uppgifter i våra ordböcker, medan annan, logiskt lika möjlig information saknas. Urvalet beror till stor del på den tradition som växt fram, och kunskap om denna hjälper oss således att förstå "the state of the art". När man, som nu för tiden, dessutom söker nya vägar inom lexikografarbetet, kan eventuellt

ett studium av lexikografins pionjärer paradoxalt nog bidra till förnyelse. Exempelvis lanserade Carl Jonas Love Almqvist såsom företrädare för den enspråkiga lexikografen en del idéer som fortfarande uppfattas som fräscha lösningar, som ännu inte vunnit allmänt insteg i våra ordböcker.

Lyckligtvis för nordisk lexikografi är både Norden och orden numera i ganska så god ordning, bland annat tack vare en stark lexikografisk tradition. Vi kan då för ett ögonblick t.o.m. låta "lexikografisk tradition" åsyfta inte bara själva verksamheten generellt utan också det mer specifika 'lexikografisk beskrivningsteknik'. Det är förvisso sant att man för de nordiska språkens del sedan länge bedriver organiserad lexikografisk verksamhet. Detta har så småningom lett till ett mer i detalj utvalt och systematiserat ordförråd, och man har också utvecklat väl inarbetade tekniker för grammatisk och annan (t.ex. stilistisk) beskrivning av detta ordförråd. Det nyss beskrivna projektet syftade just till att ge en bild av hur det svenska språket efterhand fångats upp av lexikograferna, hur ordförrådet samlats, etablerats och ordnats till en ordmängd möjlig att redovisa, definiera, klassificera och bestämma i olika avseenden. Det är denna process som Einar Haugen kallar ett språks *lexisering* (Haugen 1987:60, 64).¹ I flera avseenden verkar utvecklingen ha förlöpt på ett jämförbart sätt också i de andra nordiska länderna, vilket framgår av exempelvis rapporterna i *LexicoNordica* 7.

Lexiseringen är enligt Haugen en aspekt av ett språks kodifieringsprocess, som man kan säga inleds med utvecklandet av ett skriftspråk (grafisering) och formulerandet av de grammatiska reglerna (grammatisering). Varje fas beskriver ett mödosamt och problematiskt förlopp, som man har svårt att föreställa sig i hela dess vidd när allting väl fallit på plats. Idag tar vi den välutvecklade lexikografen för given. Detta gäller då särskilt lexikografiska specialister, men det är nog – i olika utsträckning – en

¹ Haugen använder i sin engelska text termen "lexication", vilket i Dag Gundersens norska översättning blir "leksisering" (Haugen 1990:50, 54). I språkdiskussionen i Sverige har man egentligen inte förrän alldeles nyligen urskilt "grammatisering" och "lexisering" som särskilda aspekter av språkets kodifieringsprocess (se vidare Hannesdóttir 2000).

ganska utbredd attityd, särskilt bland folk som arbetar professionellt med språk. Det har inte alltid funnits skäl till en sådan inställning. Man kan lätt konstatera att lexikografi som akademisk disciplin vid våra universitet stabiliserats märkbart bara de senaste tio–tjugo åren, åtminstone ämnesmässigt (jfr Gundersen 1992). Det är alltså i stort sett under den tid som det organiserade nordiska lexikografiska samarbetet ägt bestånd, senast manifesterat i tillkomsten av Nordisk forening for leksikografi. I ett något längre perspektiv har det hänt oerhört mycket på några få decennier, och går vi bara 100–150 år tillbaka hamnar vi i rena pionjärtiden. De närmaste ca 400 åren dessförinnan rymmer den lexikografiska förhistorien.

Kodifieringsprocessen är givetvis en väsentlig del av ett språks historia. I traditionell språkhistorisk beskrivning har man inte alltid tillräckligt noga skilt mellan det talade språkets historia och det skrivna språkets historia. Givetvis finns det ett nära samband mellan dessa, men distinktionen måste principiellt upprätthållas. Det är först när ett språk regleras i sin skriftliga form som det föreligger en förutsättning för att ett standardspråk skall utvecklas. Lexikografihistorien får företrädesvis räknas som en del av det skrivna språkets historia. Det har vidare blivit allt vanligare inom språkhistorisk beskrivning att göra en distinktion mellan inre och yttre språkhistoria. Med *inre* språkhistoria menas därvid själva språkssystemets utveckling, medan den *yttre* språkhistorien tar sikte på förändringarna i de externa villkor som kan gälla för språkets förändring. Det kan vara svårt att konsekvent ta ställning till om ett fenomen hör till den yttre eller den inre språkhistorien. Så vetter exempelvis förändringar i språkbruket och förutsättningarna för olika texttypers uppbyggnad och användning åt båda hållen. Även lexikografihistoria visar denna dubbelhet; en lexikografihistorisk studie kan bidra till såväl den yttre som den inre språkhistorien.

En lexikografikonferens idag präglas av teori- och metodmedvetenhet. I takt med att lexikografen utvecklats till en självständig, vital akademisk disciplin, har den interna metod- och teordiskussionen intensifierats, och en genomtänkt, egen terminologi etablerats (se fr.a. NLO 1997). Den lexikaliska beskrivningen

som en del av den totala språkbeskrivningen, helst på grundval av genomtänkta teoretiska modeller, har också blivit en viktig angelägenhet. Detta självmedvetande skiljer alltså dagens lexikografi från bara obetydligt äldre verksamhet på området, som inte utmärktes av explicit teoretiserande i någon större utsträckning. Att idag åter konfronteras med exempelvis Ali Al-Kasimis beskrivning från 1977 av förhållandet mellan språkvetenskapen och lexikografin är mycket tankeväckande. Man kan sålunda notera en viktig brytpunkt i den tradition, vars utveckling det är min uppgift att här skissera. Då skall man förstås märka, att frånvaron av explicit teoridiskussion inte behöver betyda teorilöshet i arbetet. Därför blir kartläggandet av den (vanligen implicita) teoretiska basen för arbetet under olika tider en viktig del även av den historiska analysen.

Ladislav Zgusta har i olika sammanhang presenterat en typologi över ordböcker med hänsyn till deras inflytande över standardformen av ett språk (sammanfattad i t.ex. Zgusta 1992:70ff.). Han talar om (1) "standard-creating dictionaries" med underavdelningarna (1a) "new standard languages", (1b) "revived languages", (1c) "varieties raised to standards"; vidare (2) "modernizing dictionaries"; (3) "antiquating (or archaizing) dictionaries"; (4) "standard-descriptive dictionaries". Ordböcker som inte är standardpåverkande är vanligen de som har främst pedagogiska syften. Snarast i en annan dimension kan ytterligare en viktig faktor urskiljas: ideologi i vid eller inskränkt mening. Någon form av ideologi kan naturligtvis utläsas ur de flesta texter: i vid mening som den enskildes eller sociala grupperns reaktion på den allmänna tidsandan, i form av vanligen implicita och osystematiska attityder, åsikter och reaktioner, eller i mer inskränkt mening som uttryck för explicit och systematisk tankebyggnad.

En typologi med dessa komponenter hjälper oss att urskilja vissa funktioner som den svenska lexikografin har haft under olika perioder. Dessa funktioner är olika starkt framträdande vid olika tidpunkter. De avlöser delvis varandra och bildar så till vida tillsammans ett historiskt förlopp, men när de väl etablerats kan

de uppträda sida vid sida, med växlande tyngd. Låt oss begrunda de viktigaste.

(1) *Lexikografarbetet inom latintraditionen* är grundläggande i Sverige liksom för övrigt inom Västerlandets lexikografihistoria. Latinets centrala ställning som gemensamt överstatligt kommunikationsmedel under långa tider, och därtill spritt över stora områden, ger det en särställning i de flesta sammanhang. Latinet har emellertid under större delen av sin långa historia inte varit modersmål för flertalet användare, utan det har varit ett inlärt språk. Även om det fyllt mångahanda praktiska funktioner har det ändå stått den lärda världen särskilt nära. Detta gäller inte minst studiet av metafysiken. Härigenom har exempelvis bibelstudiet tidvis inlemmats i latinstudiet; det har sidoställts med studiet av de latinska auktorerna, såväl de klassiska som de senare. Latinstudiet har varit en nödvändighet, med tyngdpunkten lagd på pedagogik och bildning övergående i lärdom. Listningar av ord i olika former, från enkla gloslistor till långa bibelkonkordanser, har varit integrerande delar av denna tradition. Uppkomsten av lexikografi i någon mening är överallt i västvärlden intimt förknippad med latintraditionen.

(2) Om förtrogenhet med latintraditionen länge var en nödvändighet för ett mer avancerat socialt liv, var *kunskap i andra främmande språk* – de ”moderna” språken, som vi brukar säga nu för tiden – snarast förenad med praktisk nytta. Det har naturligtvis alltid förekommit möten mellan olika språkvarieteter, ibland hos en och samma språkbrukare (dvs. hos två- eller flerspråkiga individer), men efterhand har mer systematiskt behov av kunskaper i främmande (standardiserade) språk uppstått, varvid ordkunskap är en viktig del. När grammatiska framställningar över enskilda språk börjar ta form, ingår vanligen långa listningar av ord, som oftast är kategoriserade på något särskilt sätt, och som skall tjäna som exempel på grammatiska egenskaper eller beteenden som dessa ord gemensamt uppvisar. När sådana ordlistor frigörs från de grammatiska strukturbeskrivningarna får de karaktären av ordböcker. I själva verket hör många av de äldsta

påträffade listningarna av lexikografisk natur hemma i det här sammanhanget.

(3) Språk ingår ofta i staters grundläggande uppsättning av betydelsefulla symboler. Staters autonomi markeras ytterligare genom valet av gemensamma tecken, t.ex. speciellt avpassade lagar som stipulerar statsskicket, en flagga, en nationalsång (i nyare tid) etc. Hit hör som sagt också vanligen ett eller flera preciserade språk. Som ett extremt exempel kan Frankrike anföras, som bara för några år sedan skrev in trikoloren, marseljäsen och det franska språket i grundlagens fundamentala stadganden. Härigenom fylls språket med nationella symbolvärden, och språket ägnas en medveten planering parallellt med att nationsbygget i stort förs framåt enligt en likaså medveten plan. När det gäller *språkplanering* i denna mening rör det sig om grader på en skala. Den kan skärpas till utpräglad ideologiska eller politiska dimensioner, såsom då de patriotiska strömningarna på 1600-talet i Sverige förde upp den s.k. göticismen som ledande idé. I ordböckernas värld kan det bli fråga om att lansera ord och uttryck med viss prägel som ersättning eller komplement till det spontant etablerade ordförrådet.

(4) Allmän *språknormering*, som en aspekt av språkets standardisering eller upprätthållande av språkets etablerade status, brukar vara mindre spektakulär. I ett språksamhälle är det givetvis praktiskt om det gemensamma språket fungerar med så lite gnissel som möjligt. Ett sätt att säkerställa en sådan situation är att reducera språkliga olikheter så långt möjligt genom att främja vissa varianter på andras bekostnad. Det brukar sägas att varje ordbok som får spridning har normerande effekt, oavsett om detta varit ett uttalat syfte eller ej. Som en särskild form av språknormering får man betrakta terminologiarbete.

(5) En ordbok kan möjligen också fungera som en ren *språkbeskrivning*, även om det är ovanligt i praktiken (bl.a. med hänsyn till vad som nyss påpekades om att existerande ordböcker lätt får normerande effekt). Moderna ordböcker uppges trots allt ofta ha som primärt syfte att objektivt beskriva ett visst språks ordförråd. Det ändamålsenliga i att medvetet utforma ordböcker avsedda för praktiskt bruk så att de även fungerar som språkmodeller har jag

diskuterat i ett annat sammanhang (Ralph 1995). Frågan behandlas i allmänna termer av Harras (1989).

De angivna funktionerna kan säkerligen kompletteras med andra liknande. Vissa av dem är nära relaterade till varandra, men det kan vara lämpligt att hålla isär dem, när ordböcker beaktas i det historiska perspektivet. Som redan påpekats kan också olika funktioner urskiljas i kombination med varandra.

Lexikografins utveckling som en del av språkvetenskapens historia

Eftersom lexikografen hör hemma inom språkvetenskapen, måste man för att rätt bedöma lexikografihistorien ta hänsyn till språkvetenskapens utveckling i stort. Även om inte de frågor som dominerar den specialiserade lexikografiska debatten idag är alldeles desamma som de som sysselsätter språkvetenskapen i allmänhet, finns det ändå en nära relation dem emellan. Så har det varit även under tidigare skeden, men från vår nutida utgångspunkt kanske vi inte i varje enskilt fall lika tydligt urskiljer parallelliteten mellan äldre lexikografiska arbetens utformning och den samtidigt bedrivna lingvistikens. Det är således en viktig uppgift i det lexikografihistoriska perspektivet att fastställa ordböckernas relation till samtida språkvetenskap.

Inom språkvetenskapen räknar man med ett närmast förvetenskapligt stadium fram till slutet av 1700-talet (Robins 1979: 133f.). Länge dominerades språkdebatten av *urspråksproblematiken*. Språksynen var given genom Bibelns berättelser om Babels förstöring och den påföljande språkförbistringen. En fråga som debatterades i anslutning till denna grundsyn var om hebreiskan var identisk med urspråket (dvs. det språk som Gud använde då han talade med Adam) eller om hebreiskan senare utvecklats ur ett äldre språk. Som en följd av resonemanget ville man igenkänna hebreiska ord i andra språks ordförråd, mer eller mindre vanställda. Enskilda språks förhållande till hebreiskan ansågs kunna specificeras genom graden av överensstämmelse. Synsättet motiverade sålunda den vanliga ordboksuppgiften –

oavsett språk – om hur motsvarigheten på hebreiska till en given form lyder. Man får helt enkelt betrakta detta som etymologisk information. En annan fråga gällde möjligheten att vissa folk tidigt hade sökt sig egna vägar, så att dessa undgått språkförbistringen och deras språk därmed stod i ett ursprungligare förhållande till urspråket än de flesta andra, nämligen alla de som efter förbistringen successivt hade utvecklats vidare.

Denna spekulering kring urspråket och enskilda språks förhållande till detta gav upphov till eviga *språkjämförelser*, som kan verka ometodiska, men som ofta har just hebreiskan eller ett annat givet språk, som ansetts mera ursprungligt än andra, som referensram. Språkjämförelsen sofistikerades efterhand, och det är i själva verket denna grundläggande metod som slutligen leder fram till epokgörande upptäckter, såsom den germanska ljudskridningen. Det avgörande tillkommande momentet består egentligen endast i den explicita formuleringen av *hypotesen om genetisk språksläktskap*, sådan den möter oss mot slutet av 1700-talet och kring 1800. (Se vidare t.ex. Agrell 1955.)

Just i samband med svensk lexikografihistoria kan det vara värt att notera, att Johan Ihre bland annat med sina lexikografiska arbeten intar en särställning mellan de senare företrädarna för den äldre språkspekulationen och de tidigaste representanterna för den mera strikt vetenskapliga komparistiken. Ihre har ibland, framför allt av svenskar, betraktats som en föregångare till Rask och Grimm. Hans *Glossarium Suiogothicum* (1769) är en säriling i den svenska ordboksproduktionen, en stor foliant sprängfylld av filologiskt vetande. Ihres syfte är att sätta in den samtida svenskan i ett språkligt sammanhang. Han går bakåt i tiden, över fornsvenskan och isländskan, vidare till anglosaxiskan och gotiskan för att slutligen hamna i keltiskan, latinet och grekiskan. Metoden är systematisk, helt annorlunda än det tidigare mer eller mindre slumpmässiga etymologiserandet.

Även Ihres ordbok är emellertid främst etymologisk; den är avsedd att ge en bild av det svenska ordförrådets ursprung och position inom den europeiska språkfamiljen, om vars inbördes släktskapsförhållanden Ihre har en ganska klar uppfattning. Vad som skiljer honom från de något senare banbrytarna inom den

jämförande språkforskningen är egentligen endast en explicit formulerad släktskapsteori. Intressant nog påpekar han ändå persiskans likhet med de europeiska språken, och detta ett par decennier före Sir William Jones berömda inplacering av sanskrit i den indoeuropeiska familjegemenskapen. Ihres klara blick för släktskapsförhållandena tycks emellertid ha skärpts av Erik Benzelius föregående insats. Från denne övertog Ihre mängder av arbetsmaterial. Det bestående värdet med Ihres lexikografiska arbete har väl annars varit de långa förteckningar av systematiserade iakttagelser om ljudmotsvarigheter i en lång rad språk som Rask och Grimm säkerligen utnyttjade senare. Ihre fick egentligen ingen efterföljare i Sverige. Hans ordbok var länge, trots många brister, den oöverträffade källan till information om svenska ords ursprung, men någon direkt betydelse för den senare svenska lexikografin har den knappast haft. Först på 1900-talet utkom en verklig etymologisk ordbok för svenskan (Hellquist 1922). Då hade det språkvetenskapliga klimatet ändrats radikalt.

En konsekvens av allt jämförande är fokuseringen på språkens *ordförråd*, endast indirekt på ords ljudlikheter, varvid man i praktiken normalt är omedveten om den grundläggande skillnaden mellan bokstäver och ljud. En hel del tidiga ordböcker har uppenbarligen som åtminstone sekundärt syfte att tjäna som underlag vid språkjämförelser. Endast undantagsvis läggs språkens systematiska morfologiska likheter i övrigt till grund för jämförelserna. Ett undantag utgör den finsk-ugriska språkforskningen, där en del avancerade jämförelser av hela språkssystemet genomförs tidigt. Normalt ligger tonvikten annars vid ordens fonetiska eller fonologiska egenskaper, såsom redan framhållits. Ett morfologiskt betraktelsesätt i vår mening förekommer egentligen inte – alltsedan den grekiska grammatiken är ju "etymologia" den sammanhållande termen för språkets formella sida med särskild hänsyn till härledningsaspekten (där distinktionen synkroni–diakroni inte medvetet iaktas i den äldre lingvistik).

Språkvetenskapen som helhet är i äldre tid ofta inriktad på att bidra till ett nationalspråks stabilisering. Ibland förutsätter man också att det enskilda språket kan uppvisa ett antal grundläggande karakteristika som bör tas till vara vid språkets regle-

ring. Med en lång urspråksdebatt bakom sig inom språkvetenskapen blir man på ett naturligt sätt benägen att hävda särskild ålderdomlighet för det egna språket, om den allmänna tidsandan är patriotiskt laddad. Då anlägger man också gärna en grundläggande *normativ* syn på språket, och normen är *konservativ* eller rentav *rekonstruktiv*. Följande citat från Johan Erik Rydqvist ([1849] 1863:7) är typiskt för sin tid:

Men den historiska språkforskningens åliggande är, icke blott att tyda det förflutna såsom sådant, utan också att ur det förflutna förklara det närvarande, uppställa språkets genomgående lagar: visa vägen och sättet, hvarpå det nu bestående uppkommit: hvad deruti är ren och naturenlig utveckling eller missbildning: hvad af det missbildade fastvuxit med det äkta, ursprungliga, eller hvad som blott löst fladdrar på ytan, och utan fara och utan skada kan afsöndras. [- - -] Under vissa tidevarf, så länge folken stå naturen närmare, äro de ställda under hennes omedelbara ledning, och verka liksom med hennes magt. Språk-instinkten är då friskare, lifligare, renare. Men med en framskriden odling förlorar naturdriften något af sin verksamhet. Odlingen måste då, med sina egna medel, återställa det förlorade: vetenskapen och konsten måste hela såren, som de sjelfve slagit. Hvad förr var blott dunkel drift, skall derföre bringas till medvetande. Först sedan man kommit till klar uppfattning af sitt modersmåls sanna art och väsende, kan man med insigt, sans och smak verka för dess förnyring och tillväxt. Språkforskaren åligger, ej mindre att gå medlande och jemkande emellan, än att sjelf träda i spetsen både för rörelsen och återhållet, på en gång väcka slumrande anlag och välbetänt använda de redan utvecklade. Mer än till en viss grad skall det icke lyckas honom. Det är likväl hans pligt, att ega det stora till ögonmärke.

Såväl grammatiken som språkhistorien står således primärt i *språkvårdens* tjänst.

Språket är varken nu eller i äldre tider bara analysobjekt utan i första hand medel. Såväl de förmedlade åsikterna och idéerna som de språkliga uttrycksmedel som väljs för att överföra dem varierar. Mediet, dvs. språket, är också utsatt för växlande ideal som gör sig gällande i det här sammanhanget. Allt detta måste också hållas i minnet då äldre ordböcker undersöks. Under större delen av den tid som lexikografihistorien spänner över, åtminstone fram till början av 1800-talet, leddes de lärda i sin verksamhet av ett grundläggande ideal, *eloquentia*. Detta betydde långt

mer än vad den direkta översättningen 'värtalighet' ger vid handen. Språket uppfattades som en av kulturens eller rentav samhällets viktigaste förutsättningar och skulle därför helst behärskas till fulländning. Bland skolämnena och vetenskaperna blev följaktligen *artes dicendi*, dvs. grammatiken, dialektiken (eller logiken) och retoriken, fundamentala. Kurt Johannesson sammanfattar den rådande språksynen så här (1982:59):

Grammatiken påvisar den ordning, som måste existera mellan orden och deras skilda former, för att man därmed skall kunna skapa meningsfulla språkliga uttryck. Dialektiken söker förena denna ordens inneboende ordning med en annan, den rationella ordning bland tillvarons fenomen som det mänskliga förnuftet genom bestämda operationer kan återfinna eller upprätta. Retoriken slutligen visar hur själva kommunikationsprocesserna följer vissa lagar och mönster, som ger människan möjlighet att via språket nå och påverka skilda mottagare på skilda sätt. Här förbinds språket därför med samhällets egen ordning. När dessa tre konster förenas, betraktas språket samtidigt ur tre perspektiv: dess meningsfullhet, dess halt av sanning och slutligen dess effektivitet som individuellt kommunikationsmedel och socialt instrument.

Men *eloquentia* förblir inte ett statiskt begrepp. Dess olika komponenter betonas olika starkt under olika perioder, och även varje komponent kan skifta karaktär. Exempelvis har retorikens uppgift anpassats till olika tidsideal, utan att detta inverkat på den retoriska tekniken och arsenalen – men däremot på utnyttjandet av dessa. Under medeltiden har retoriken snarast en ordnande funktion; den skolastiska utläggningen är tidstypisk. Under renessansen blir retorikens uppgift mer att tolka och tillrättalägga, dvs. den får en övertygande funktion. Så småningom ställs retoriken alltmer i det sublimas tjänst, så att dess funktion blir att sammanlänka den sinnliga världen med en andligare värld. Så småningom – under nyare tiden – blir retorikens funktion alltmer att tillföra ren prydnad.

Motsvarande tendenser kan iakttas inom exempelvis diktningen respektive historieskrivningen, där retorikens regler ofta återspeglas i valet av uttrycksform. Den företrädesvis episka diktningen (bl.a. i historieskrivningens tjänst) dominerar under medeltiden, medan lärodikten och den lyriska betraktelsen så småningom vinner terräng, varefter formen betonas alltmer och

dikten som estetik tar över. Inom en texttyp som historieskrivningen är annalistisk, diariet och ämbetsmannalängder samt enkelt krönikeberättande länge allenarådande, medan mera moraliserande exemplar och rentav elaborerade allegorier efterhand ströms in. Med renässansen favoriseras den propagandistiska historieskrivningen som skulle tjäna politiska syften, gärna till den begynnande nationalstatens fromma.¹ Med upplysningstiden utvecklas kriterier för en mer rationell, vetenskaplig historieskrivning, samtidigt som man försöker uppställa lagar för kulturhistoriens förlopp och tilldelar historieskrivningen en förutsägande kraft.

Det är i en växlande *idéhistorisk rymd* med sådana inslag som här beskrivits som lexikografin tar form. Ordboksförfattarnas bildning och utbildning genomsyras av ideal av den typ som exemplifierats. Det måste man hålla i minnet även vid analysen av lexikografins utveckling.

Svensk lexikografihistoria

Här följer nu i enlighet med uppdraget en kortfattad redogörelse för den svenska lexikografihistoriens förlopp. I en översikt över framväxten av den lexikografiska traditionen i Sverige kan det vara naturligt att inledningsvis fråga sig vilken som är den första svenska ordboken. Även om frågan är legitim, är det lättare att formulera frågan än att ge något klart, entydigt och enkelt svar. Låt oss rådfråga traditionen. "Först i början av 1700-talet erhålla vi en något så när fullständig ordbok, som betecknar ett framsteg utöfver Lincopensens index, i och genom ärkebiskop HAQVIN SPEGELS (d. 1714) *Glossarium-Sveo-Gothicum Eller Svensk-Ordabook*, Lund 1712, där de svenska orden äro öfversatta till latin, stundom äfven till andra språk, och som innehåller strödda etymologiska anmärkningar." Orden är Adolf Noreens, när han i

¹ Det brukar hävdas att nationalstaten är ett fenomen som egentligen uppstår först på 1800-talet, som en följd av franska revolutionen. Det är naturligtvis en definitionsfråga. I vissa avseenden visar exempelvis det svenska riket i sin opposition mot Danmark tidigt upp nationalstatsdrag.

första delen av *Vårt språk* går igenom den nysvenska språkforskningens historia (Noreen 1903:189). Det är tydligen på hans omdöme andra påståenden går tillbaka, direkt eller indirekt, om att Spegels *Glossarium* är den första svenska ordboken. ”Som språkvetenskapsman var Spiegel den förste, som utgav ett svenskt lexikon: ’Glossarium — Sveo-Gothicum eller Swensk-Ordabook’ 1712”, meddelar exempelvis Jan Agrell i sin utmärkta sammanfattning av den äldre språkjämförelsens allmänna och svenska historia fram till 1827 (Agrell 1955:129).

Annars skulle man kunna föra fram andra kandidater. Lincopensens index, som Noreen nämner i det citerade stycket, avser en svensk ordsamling principiellt ordnad alfabetiskt. ”Lincopensen” eller ”Lexicon lincopense” är det gängse smeknamnet på *Dictionarium Latino-Sveco-Germanicum*, som påbörjades av Nicolaus Nicolai Grubb, färdigställdes och publicerades år 1640 av hans efterträdare i ämbetet (och i äkta sängen), Jonas Petri Gothus, så småningom biskop i Linköping. Ordboken som sådan är en märklig skapelse, en mäktig foliant, till synes sprungen direkt ur Zeus huvud, utan några egentliga föregångare inom landet. Det rör sig emellertid om en gren på det livaktiga latinska lexikonträdet, som har sina rötter djupt i den kontinentala myllan. Monica Johansson har i sin avhandling (1997:120ff.) övertygande visat att den viktigaste förlagan är *Lexicon Trilingve*, som kom ut i Strasbourg 1587.

Till denna latin-svenska ordbok fogades sålunda ett svenskt ordindex (och dessutom ett tyskt), som är den första längre svenska ordsamlingen (192 foliospalter), av allt att döma utförd av Jonas Petri Gothus själv. Indexet har senare återanvänts i olika sammanhang, framför allt i olika upplagor av Comenius *Janua Lingvarum* i svensk översättning, den första ombesörjd 1640 av Ericus Schroderus men redan året därpå publicerad med ”Register på alla the Swenske Ord”, som också detta tycks vara Schroderus verk. Från och med upplagan 1642 innehåller *Janua Lingvarum* Lincopensens index, nu kallat *Index svecicus secundus* (se vidare Johansson 1997, särsk. s. 369ff.).

Eventuellt skulle man kunna känna sig frestad att utse ett ganska exakt hundra år äldre arbete av lexikalisk natur till första

svenska ordbok: *Variarum rerum vocabula cum sueca interpretatione*, som publicerades i Stockholm 1538. Det skulle kunna finnas olika skäl till en sådan upphöjelse. För det första är *Variarum rerum vocabula* det första lexikografiska arbete som trycks i Sverige, dessutom innehåller det svenska, och vidare tar det delvis upp ord av en typ som sällan förekommit i texter dessförinnan och inte förrän långt senare. Men för att återigen anknyta till Noreens uppfattning: ”Detta anonyma arbete, väsentligen en öfversättning från tyskan, är dock icke en produkt af svenskt, utan af latinskt språkstudium, utgörande en efter sakliga kategorier ordnad latinsk glosbok med mot de latinska glosorna svarande svensk öfversättning, således indirekt en liten nysvensk ordbok med realuppställning” (Noreen 1903:184). Här kan väl också tilläggas att bakgrunden till detta lexikografiska förstlingsarbete inte är restlöst klarlagd, men att flera som yttrat sig velat förknippa det med någon av de svenska reformatörerna, i första hand någon av bröderna Laurentius eller Olaus Petri. Einar Haugen satsar särskilt på Olaus och framhåller som en möjlig parallell, att den första liknande vokabelsamlingen i Danmark, *Vocabularium ad usum dacorum* (1510), brukar tillskrivas den danske reformatorn och bibelöversättaren Christiern Pedersen (Haugen 1986:100). I det sammanhanget kan det förtjäna att påpekas att den svenska *Variarum rerum vocabula*, förutom att den återuppstod i två nya upplagor och sammanlagt ett tiotal nytryck till in på 1700-talet, också lades till grund för en dansk upplaga 1576: *Vocabulorum variorum expositio atqve collectio ex lingua svetica*.

Går vi ytterligare ett steg tillbaka i tiden, till 1400-talets senmedeltid, kan vi konstatera att vi där finner det första kända lexikografiska arbetet tillkommet på svensk mark. Ett glossarium med beteckningen Codex Upsaliensis C 20 föreligger i handskrift från 1470-talet, därtill ett troligen äldre fragment, som kan härröra från en förlaga; det anses nämligen säkerställt att C 20 är en avskrift av ett äldre svenskt arbete. C 20 anses vara en lätt bearbetning – eller snarast en ganska trogen kopia – av Johannes Balbus de Januas *Lexicon Catholicon*, som färdigställdes år 1286 och varav två exemplar anses ha införts i Sverige under

tidigt 1400-tal. Den svenska texten kan ha kommit till i birgittinsk klostermiljö, och man har gissat på Jöns Budde, flitig översättare i Vadstena och Nådendal, som den som försvenskat lexikonet. Lars Wollin, som senast undersökt glossaret, lämnar frågan om upphovsmannen öppen (Wollin 1989).

Om vi ansluter oss till Adolf Noreens argumentation diskvalificeras såväl det handskrivna glossaret C 20 som *Variarum rerum vocabula* och den därnäst följande regelrätta latinordboken med svensk uttolkning, *Lexicon lincopense*, som första svenska ordbok. Kvar står den ännu icke nämnda *Synonymorum libellvs*, som utgavs av Elaus Petri Helsingius år 1587. Dess betydelse verkar klart underskattad i svensk lexikografihistoria.

Synonymorum libellvs återgår förvisso på utländska förlagor, liksom alla de tidigaste lexikografiska arbetena tillkomna i Sverige. Man har pekat på Johannes Serranus lågtyska ordbok med samma namn, tryckt i Nürnberg 1567. Denna utgår i sin tur från holländaren Pelegromius *Synonymorum silva* (1546). Men Helsingius presterar trots allt den första lexikografiska framställningen någonsin som har svenska som källspråk. Det är sant att den mer eller mindre kalkerar en utländsk förlaga, men det finns faktiskt svenska uppslagsord hos Helsingius som inte kan återföras till någon av de allmänt erkända källorna. Dessutom förses de svenska uppslagsorden med såväl latinska som grekiska ekvivalenter, och boken användes länge i undervisningen i grekiska (Lindroth 1975:271). Det framstår som märkligt att boken, låt vara ganska anspråkslös, så totalt negligerats av forskningen.

Helsingius ordbok kan, liksom de andra här behandlade, med lite god vilja föras till latintraditionen. Låt oss hastigt bara följa detta spår ytterligare framåt i tiden. I och med *Lexicon lincopense* etableras alltså en av de två latinbaserade traditioner som alla senare latinordböcker kan återföras på. Den andra, som så småningom utkonkurrerar Lincopenstraditionen, startar med Petrus Schenbergs *Lexicon latino-svecanum*, som utkom 1739. Denna ordbok är i sammanhanget särskilt intressant, eftersom Schenberg också presenterar en *Swensk Ordabok, ther hvart och ett Swenskt ord och ordalag til thes skrifart, på Latin gifwes och vttydes*. För denna svensk-latinska del har Schenberg excer-

perat Jesper Swedbergs handskrivna, aldrig publicerade svenska ordbok, som Lars Holm studerat (Holm 1986). Visserligen tillkommer en del ytterligare framstående lexikografiska arbeten inom den latinska traditionen, och somliga bidrar rentav, som Anna Hannesdóttir visat (Hannesdóttir 1998), till det lexikografiska raffinemanget, men efter Schenberg är det egentligen inga brott i traditionen fram till Cavallin (1871 och 1875–76) och våra dagars svenska latinlexikon. Därmed lämnar vi latintraditionen.

Ytterligare ett arbete har utpekats som vår första svenska ordbok, nämligen av dess utgivare Bengt Hesselman: Ericus Schroderus Upsaliensis "Dictionarium quadrilingue Svedicum Germanicum Latinum Græcum". Hesselman motiverar sin ståndpunkt på följande sätt: "Dictionarium quadrilingue, som måste ha tillkommit på 1630-talet, är som titeln anger ett fyrspråkslexikon, med de svenska orden i första spalten, därefter tyska, latin och grekiska. Det är alltså i första rummet ett svenskt lexikon och utgör i själva verket det första avslutade ordboksarbete i Sverige, för vilket huvudintresset har varit ett så vidt möjligt fullständigt insamlande av det i bruk varande svenska ordförrådet" (Hesselman 1929:VI). Detta arbete trycktes aldrig och har därför inte haft något inflytande på den svenska ordbokstraditionen. Detta hindrar inte att boken är ett intressant vittnesbörd om sin upphovsmans lexikografiska förmåga.

Däremot utkom verkligen av samme författare år 1637 ett *Lexicon latino-scondicum, Latinum, Svecicum, Germanicum & Venedicum seu Finnonicum*, som alltså likaledes innehåller fyra språk: latin-svenska-tyska-finska; inte med svenskan först, men med *finskan* som ett av de fyra språken. Egentligen är detta inte så förvånansvärt, eftersom Finland utgjorde den östra rikshalvan. Den senare så seglivade myten om Sveriges enspråkighet hade ännu inte fått fotfäste inom landet.

Sådana flerspråkiga sammanställningar är långt vanligare i äldre tid än senare (se t.ex. Noreen 1903 och Hannesdóttir 1987). De olika två- och flerspråkiga ordböckerna värderas inte särskilt högt av Noreen. Under beteckningen "praktiska hjälpreddor, bland hvilka visserligen några mycket värdefulla, för öfversättning till främmande språk" (Noreen 1903:209), sorterar han

in Petrus Fiellströms *Dictionarium sueco-lapponicum* (1738), Olof Linds *Teutsch-Schwedisches und Schwedisch-Teutsches Dictionarium* (1738, utökad uppl. 1749), Schenbergers nämnda latinordbok (1739, ny uppl. 1747), Jacob Serenius *Dictionarium Suethico-Anglo-Latinum* (1741), Levin Möllers *Nouveau dictionnaire françois-svedois et svedois-françois* (1745, ny uppl. 1755), Daniel Juslenius *Finsk Orda-Boks Försök* (1745), med ett register på svenska, Dähnerts *Liber memorialis germanico-latino-svecicus* (1746, ny uppl. 1775) samt samme författares *Kurzgefasstes Deutsches und Schwedisches Hand-Lexicon* (1784).

Andra deltagare i det ovannämnda lexikografihistoriska projektet har noggrant beskrivit de flesta av dessa ordböcker och visat på vad sätt de kan sägas ha bidragit till utvecklingen av den svenska lexikografiska traditionen. Anna Hannesdóttir och Lena Rogström kommer också in på det i sina bidrag nedan. Framför allt kan man kanske lägga märke till vilket spektrum av språk som finns representerat i dessa ordböcker, ett intryck som förstärks om man gör en fullständig genomgång.

Med detta tänkte jag ändå lämna även de tvåspråkiga ordböckerna och övergå till att konstatera, att Noreen med rätta ger Abraham Sahlstedt en särskild rangplats i lexikografihistorien. Efter åtskilliga lexikografiska och grammatiska förarbeten av växlande omfattning och efter mängder med förvecklingar utkom 1773 hans epokgörande *Swensk Ordbok Med Latinsk Uttolkning, På Kongl. Maj:ts allernådigste Befallning Författad*. Ibland har man betraktat denna ordbok som ett svensk-latinskt lexikon, vilket är fel. I allt väsentligt är det en ganska uttömmande svensk ordbok med latin som metaspråk för de semantiska specifikationerna. Dessutom står ordboken i ett systematiskt och komplementärt förhållande till Sahlstedts grammatiska utredningar på annan plats. I vilken utsträckning Sahlstedt utnyttjat Samuel Schultzes utgivna manuskript – som han fick tillgång till och som äger stora lexikografiska förtjänster, såsom förre redaktören vid Svenska Akademiens ordbok, Sture Hast, visat (Hast 1990) – går väl inte att avgöra. Själv kom Sahlstedt under alla omständigheter – eller rättare sagt hans ordbok – att

bli ett föredöme för många senare lexikografer (Hannesdóttir 1998, särsk. s. 238ff.).

Med Sahlstedt är vi egentligen framme vid den enspråkiga ordbokstraditionen, fullt utvecklad hos Anders Fredrik Dalin i mitten på 1800-talet (Hannesdóttir 1998, särsk. s. 462ff.). Denna senare utveckling har också belysts väl i tidigare arbeten. Jag hänvisar generellt till Anna Hannesdóttirs avhandling (1998) med de hänvisningar som ges där. Man kan säga att såväl de lexikografiska arbeten som tillkommit inom latintraditionen som de tvåspråkiga främst tjänar pedagogiska syften, om vi tänker på Zgustas typologi som återges ovan. Så till vida försvarar de sin plats även i pedagogikhistorien. Därtill kommer en del tvåspråkiga ordböcker efterhand att bidra till standardiseringen av svenskan, en uppgift som senare överflyttas helt på de enspråkiga ordböckerna, varvid det snarast blir fråga om ett bibehållande och stärkande av standarden. Här skulle jag emellertid ännu en gång vilja återvända till den äldre ordbokstraditionen för att peka på några omständigheter, som torde vara betydelsefulla men som inte framhållits tillräckligt starkt i äldre översikter (t.ex. Noreen 1903). Därvid koncentreras intresset mer till den ideologiska snedbelysning som också antyds i typologin.

Den patriotiska lexikografin

När man undersöker Spegels ordbok från 1712 slås man i första hand av hur underlig den är, sedd med våra ögon. Den innehåller vitt skilda typer av artiklar, alltifrån vad vi nu uppfattar som helt vanliga ordboksartiklar till långa utredningar av närmast encyklopedisk karaktär. Till många ord fogas oändliga rader med ekvivalenter eller på annat sätt liknande ord hämtade från andra språk. Syftet med ordboken uppges vara att underlätta för de nya sydsvenskarna, som före Roskildefreden 1658 hade varit danskar, att läsa den heliga Skrift på svenska. Hur väl lämpad för denna uppgift den har varit är det svårt att avgöra; visst finns det ord ur Bibeln bland uppslagsorden, men inte alls så konsekvent listade som exempelvis i Jesper Swedbergs opublicerade ord-

boksmanus. Däremot finns det åtskilligt hämtat från dialekternas ordförråd och – framför allt – ålderdomliga ord och uttryck från de fornsvenska texterna samt antagligen för de flesta samtida svenskar tämligen obegripliga norröna ord.

På sätt och vis är sålunda Spegel, liksom så många av de första lexikograferna, ute i patriotiska ärenden. Ändå måste nog Spegels ordbok uppfattas som ett brott med en tidigare tradition. Han tar exempelvis uttryckligen avstånd från den form av högstämd patriotism som i svensk kulturhistoria brukar kallas *göticism*, och som hade varit rådande under den föregående perioden. Låt oss ett ögonblick fundera över denna tidsanda såsom den naturliga idéhistoriska kontexten för tidens lexikografiska prestationer.

”Sin ööverträffade höjdpunkt nådde göticismen i Rudbecks *Atlantica*”, skriver Sten Lindroth i sin likaledes ööverträffade framställning av svensk idé- och lärdomshistoria (Lindroth 1975: 284). Olof Rudbeck var egentligen läkare och naturforskare, från 1658 professor i medicin i Uppsala. Inom anatomin var han världsberömd, med den epokgörande upptäckten av lymfsystemet bland sina främsta bragder, men han förnyade också medicinundervisningen i allmänhet och var trägen inom de dagliga universitetsrutinerna. På äldre dagar blev han en framstående botanist och grundade tillsammans med sonen med samma namn en naturalhistorisk tradition, som betytt avsevärt mer för exempelvis Linné än man brukar uppmärksamma.

Även om Rudbeck således genom en fördomsfri hållning och öppen attityd till det mesta i världen och därtill med en stark tro på vetenskapliga experiment representerade en ny intellektuell era, så markerar han i sitt intresse för antikviteter och fornhistoria just en höjdpunkt i de göticistiska spekulationerna. Dessa kan vid ett ytligt betraktande verka helt galna och närmast genanta för en svensk idag. ”Atlantica är ett monument över stormaktstidens vilseförda patriotiska forskaranda, oroande, nästan skrämmande i sin förening av snille och vanvett”, för att än en gång citera Lindroth (1975:284).

Men det finns anledning att se närmare på den förkättrade svenska stormaktsgöticismen, både i allmänhet och såsom den

företrädde av Rudbeck i synnerhet. Spåren av den levde nämligen kvar länge, och den har haft en icke-trivial betydelse för lexikografin. Even Hovdhaugen har i en studie valt Rudbecks *Atlantica* och den svenska göticismen under stormaktstiden som exempel på hur forskares subjektiva idéer och det ideologiska klimat som råder i det samhälle som de är en del av generellt styr vår uppfattning om vad som kan anses vara en vetenskaplig förklaring (Hovdhaugen 1996). Rudbecks uppburna forskning, som vi nu uppfattar som fullständigt ovetenskapliga, vilda fantasier, låt vara skarpsinniga, illustrerar på ett slående sätt hur relativt begreppet "förklaring" är. Med risk för att falla under samma dom som den vi utfärdar över Rudbeck skall vi här se något närmare på hur han arbetade och hur de göticistiska luftsloppen byggdes upp.

Det är ganska symptomatiskt, men långt ifrån självklart, att Rudbeck skulle gripas av den snedvridna patriotism som var vida spridd i landets ledande kretsar under 1600-talets senare del. Han hade som sagt sedan länge en framskjuten position inom lärdomsetablissemangen, och han uppvisade utpräglade drag av mångkunnig renässansgestalt, men var länge alltså inte särskilt engagerad inom fornforskningen, såsom de något äldre renässansfigurerna Bureus och Stiernhielm hade varit. Det var först när Olof Verelius skulle ge ut den fornvästnordiska *Hervarasagan* (som ansågs vara särskilt intressant för svenskarna) som Rudbeck blev inkopplad. Han förväntades åstadkomma en Sverigekarta, som skulle skänka åskådlighet åt de händelser sagan återberättar. Rudbeck orkade bara med två kartor över Mälardalen, men hans intresse för fornforskningen var väckt, så småningom med häpnadsväckande resultat. Den tändande gnistan var de mängder av överensstämmelser som Rudbeck tyckte sig kunna konstatera mellan å ena sidan namn på orter, personer och folkslag i den isländska sagan, som i sin tur kunde återfinnas i det svenska namnbeståndet, och – å andra sidan – namn i den antika litteraturen.

Under ett par år utvecklar Rudbeck en febril verksamhet och har 1679 åstadkommit ett helt enastående verk: *Atland eller Manhem*, på svenska och med latinsk parallelltext. Detta första

band var ursprungligen avsett att vara ett avslutat helt, men 1689 kom ett ännu tjockare band, följt 1698 av ett tredje, och ett rentryckt fjärde förstördes nästan helt vid branden i Uppsala 1702. Rudbeck fullföljer en tradition från bröderna Johannes och Olaus Magnus, Johannes Messenius, Johannes Bureus, Georg Stiernhielm, Johannes Loccenius m.fl., som tidigare bidragit till att utveckla bilden av svenska rikets sagolika förflutna. Rudbecks mest originella bidrag består i identifieringen av Sverige med Platons antika Atlantis (som omtalas i dialogen Timaios), allt annat är en extrem konsekvens av det redan etablerade fantastieriet. Han tar upp hyperboréerlegenden, införd av Bureus, som handlar om ”de yverborne”, ättlingarna till kung Bore längst i norr, och så är det vildsinta etymologiserandet igång. Täby i Stockholmstrakten identifieras med grekernas Thebe. Ett fullständigt magnifikt jämförande på bred front tar vid. Rudbeck målar upp en myllrande ”Urgeschichte”, som bitvis för tankarna till Prologen i Snorri Sturlusons Edda, som han för övrigt kände väl till.

Rudbeck följdes av några generationer trogna lärjungar och uttolkare, som underbyggde den sanktionerade statslögnen.¹ Rudbecks eget incitament återgick alltså på kontakten med Verelius. Denne var tolv år äldre och hade redan en lång och uppskattad verksamhet bakom sig. Men även han vanns först sent för antikviteternas utforskning. Vid knappt 50 års ålder hade han 1666 övertagit riksantikvarieämbetet efter Laurentius Bureus, en släkting till den förste innehavaren av detta ämbete, den legendariske Johannes Bureus. Dessförinnan hade han varit mycket framgångsrik räntmästare vid Uppsala universitet men blev genom direkt ingripande från rikskanslern, Magnus Gabriel De la Gardie, anmodad att med denna tjänst förena en nyinrättad professur i ”fäderneslandets antikviteter”. Detta uppdrag innebar att han skulle sammanställa allt som kunde belysa den svenska fornhistorien, i praktiken ett i första hand filologiskt arbete.

¹ Man bör i den svenska kulturhistorien egentligen skilja mellan ”göticism” och ”rudbeckianism”, såsom Mats Malm visat (1996:15).

Verelius brann tydligen av iver inför uppgiften. Han gav föreläsningar om de gamla lagarnas språk och de isländska sagorna, om runologi, Wulfilas gotiska bibel, som kommit till Uppsala genom De la Gardies försorg och – inte minst – historieskrivningen, framför allt efter Johannes Magnus. Han gav ut en handbok i runkunskap, där han levererar batalj med dansken Ole Worms, som han genomgående refererar till som ”doktor Orm”. Han skrev en mycket patriotisk historia, som länge förblev otryckt. Men framför allt gav han sig i kast med det isländska materialet.

Under 1600-talet inträffar som bekant – till en början främst i Danmark – det man ibland kallat ”den isländska renässansen”. Den mångsidige Ole Worms var en av dem som gick i täten. Han var god vän med islänningarna Arngrímur Jónsson och Brynjólfur Sveinsson, båda präster, den senare biskop, med vilka han uppehöll en flitig brevkontakt. Dessa hade stor del i att den lärda världens uppmärksamhet kom att riktas mot de fornisländska handskrifterna, även utanför Island. År 1658 kom Jón Rugman mera av en tillfällighet till Sverige, snart följd av andra islänningar och, framför allt, en hel del isländskt handskriftsmaterial. Detta ledde också successivt till en utgivningsverksamhet, där svenskarna stod för en pionjärinsats.

Med Rugmans och andra islänningars hjälp anskaffades sålunda en del isländska handskrifter. En rangplats i det materialet intog den så kallade Uppsala-Eddan, ännu idag betraktad som en av huvudhandskrifterna till Snorri Sturlusons Edda. Handskriften hade De la Gardie ställt till förfogande. Av denna gjorde Verelius en avskrift, och Rudbeck studerade också intensivt detta arbete. Det var genom dessa båda entusiasternas förenade ansträngningar som Verelius *Index lingvæ veteris scytho-scandicæ sive gothicæ* slutligen kom ut.

För Verelius var – liksom något tidigare för Stiernhielm – den gamla svenskan och den gamla isländskan samma språk, dvs. ”götiska”, som en gång talats av förfäderna, de gamla nordborna. I ett register till Hervararsagan, som alltså publicerades 1672, hade ett ganska fylligt ordregister fått förklaringar, eller snarare ekvivalenter, på latin. Under sin senare levnad kämpade Verelius

med en regelrätt ordbok över det nordiska fornspråket, i samma stil men betydligt mer omfattande. Han fick aldrig se den fullbordad, men Rudbeck lyckades till sist få verket utgivet 1691, nio år efter Verelius död (1682). *Index lingvae veteris scytho-scandicae sive gothicae* upptar glosor och språkprov från svenska medeltidslagar och fornisländska sagor och ger motsvarigheter på tidens svenska och latin. Syftet med detta lexikon (Verelius använder själv genomgående den modesta beteckningen "index") är detsamma som Stiernhielm haft dessförinnan: att säkerställa ett gammalt fosterländskt ordförråds fortbestånd, så att det inte skulle gå förlorat sedan det räddats till eftervärlden genom återupptäckten av den isländska litteraturen.

Verelius index ger källuppgifter för de ord som upptas. På en slumpvis vald sida får man sådana källuppgifter till 72 ord. Av dessa hänför sig 2 till Bibeln, 9 till svenska landskapslagar, 1 till (den danska) Skånelagen, 1 till den norska Frostatingslagen, 3 till Wulfilas gotiska bibel (Verelius skriver *Ulphila*), 1 till Snorris Edda, 3 till den norska kungaspegeln, 1 till den svenska, 7 till olika fornvästnordiska handskrifter, 13 till de norska kungasagorna och resten, dvs. 31, till utgivna isländska sagor. Detta kan väl ge ett intryck av vad för sorts ordförråd det är som ingår i indexet.

Även om Verelius index är det enda större göticistiska arbete av rent lexikografisk art som färdigställs och publiceras, var den föreställningsvärld det representerar väl spridd bland landets elit under stormaktstiden. Georg Stiernhielm är viktig som idégivare på det här området, men liksom sin lärofader Johannes Bureus fullbordade han ganska lite själv. Han ger däremot Rudbeck impulser till *Atlantican* genom att vidareutveckla hyperboréertemat från Bureus. Annars är hans stora projekt urspråksproblematiken. Den blir väl belyst i utredningen "De linguarum origine", tryckt som förord till Stiernhielms utgåva av den gotiska silverbibeln 1671.

Stiernhielms urspråksteori går ut på att det äldsta språket bevarats i vårt götiska modersmål. Uppslaget hade han, som så mycket annat, fått från Bureus. Denne hade i sin tur tänkt vidare med utgångspunkt i belgaren Goropius Becanus arbete *Origines*

Antwerpianae (1569). Becanus markerar ett brott i den dittills rotfasta föreställningen att hebreiskan var identiskt med urspråket. Motivet till den egentligen rätt utmanande idén var patriotiskt. Becanus såg cimbriskan, dvs. nederländskan, som det oförfalskade urspråket. Förklaringen var helt enkelt att hans förfäder, cimbrerna, aldrig skulle ha varit med vid Babel och därför inte blivit drabbade av språkförobustringen. Resonemang av sådant slag skulle för stormaktstidens svenska tänkare erbjuda helt nya möjligheter. För Stiernhielms del var det basen för ett vidlyftigt teoretiskt system.

Stiernhielms språkspekulation innehåller naturligtvis många drag som är helt främmande för oss. Det götiska urspråkets primitiva enkla ordrötter hade tillsammans med allting annat i kosmos uppkommit genom harmonisk samverkan mellan de tre världsprinciperna Förnuft, Ljus och Materia, och varje element var betydelsebärande. Så ansågs exempelvis bokstaven A, vokalernas moder, betyda 'ljus', medan U fick symbolisera 'natt' och 'mörker'. I sådana banor länkades resonemanget. Ordrötterna tänktes rymma mer komplexa betydelser och ha allmängiltigt värde. Genom att analysera dem kunde man avslöja förborgade hemligheter om tillvarons innersta väsen. Språkspekulationen ledde över i metafysiska funderingar. Därmed började ett ändlöst jämförande mellan språken, således inte alltid bara i vad vi uppfattar som språkvetenskapliga syften. Men Stiernhielm var alltså härvidlag i samklang med de internationella tongångarna.

Resultaten av alla mödor samlades bland annat i form av långa listor med götiska ord. Därur extraherades de förment urgamla svenska ordstammarna, som i bästa fall också återfanns i hebreiskan och andra fornspråk. Urspråket blev en svensk egendom, de fornnordiska orden ansågs återge tingens verkliga natur.

Givet det nordiska fornspråkets sublimes status blir det för Stiernhielm naturligt att företräda språklig purism, och han samlar sina excerperade ord till ett lexikon: *Gambla Swea- och Götha-måles fatebur*, som dock aldrig kom längre än bokstaven A. Men han inventerar som en av de första det material som för Verelius och andra efterkommande blir oundgängligt i jakten på gamla ord, som av olika skäl måste räddas till eftervärlden.

Stiernhielms berömda företal till *Gambla Swea- och Göthamåles fatebur* (1643) innehåller ganska lite explicit metafysisk spekulaton. Visserligen framställs människans språk som ett Guds under, men Stiernhielm ger ändå mest rationella skäl för att man skall återinföra ord från det gamla språket. Han blev betydligt mer spekulativ senare. Också därvidlag förvaltade han arvet från Bureus väl.

Johannes Bureus kan i många stycken betraktas som Stiernhielms lärare och föregångare. Han var också minst lika begåvad och mångsidig, men hans vetenskapliga gärning är oupplösligt förenad med det mest avancerade fantasteri. I det mesta av sitt tänkande tilldelar han språket en avgörande roll. Bureus var som riksantikvarie djupt engagerad i de gamla nordiska texterna, och han kan betecknas som Sveriges förste filolog. Han är också den i särklass skickligaste av de tidiga runologerna, liksom Verelius ständigt i fejd med Ole Worms. Hans intresse omfattade även samtidsspråket: han arbetade på såväl en grammatik över svenskan som dialektframställningar med funderingar om fonetik. Runt 1630 höll han också på med ett "Lexicon gothicum", som helt enkelt var en svensk ordbok, det mesta nu förlorat. Allt försökte han förena med teosofiska spekulatoner och religiös mystik, apokalyptisk bibeltolkning och varm patriotism. Möjligen var det hans pålitlighet som glödande patriot som ibland räddade den extreme och oortodoxe religiöse mystikern från kyrkans onåd. Omkring 1634 tillkommer en "Afhandling om språkens vpkomst", vari Bureus framställer det gudomliga Ordet, *Logos*, som makten bakom skapelsen och den senare därför som en i grunden språklig entitet. Bureus var mycket väl förtrogen med den judiska kabbalan, och denna blev både i traditionell och i vidareutvecklad form vägledande för hans tänkande. Så blev runorna, språket och hela skapelsen i hans världsbild sprängfyllda med esoterisk kunskap, som han tog som sin uppgift att klarlägga.

Redan hos Bureus möter således mycket av det fantastiska tankegods som efterhand skulle omfattas av alltfler svenskar och bland annat driva dem till orimliga anspråk för nationens del. Bureus själv förblev jämförelsevis stillsam i sin framtoning. "Den

fromme mystikern och fornforskaren brydde sig föga om härnadståg och vapenbrak, för honom var den svunna götiska storhetstiden framför allt en andlig guldålder. Våra hedniska förfäder hade kunskap om den sanna religionen och uttryckte genom runtecknen, brukade som adulrunor, dess djupaste hemligheter” (Lindroth 1975:242).

När Spegel i sin ordbok 1712 tar avstånd från den extrema göticismen, hindrar inte det att hans verk ändå kan sägas i vissa avseenden beteckna höjdpunkten på den patriotiska lexikografin. Den underrubrik han ger: ”Them til en wällmeent Anledning/ som om thet härliga Språket willia begynna någon kunskap inhämta”, påminner om formuleringarna i Stiernhielms förord till den projekterade ”Fataburen”. Spegel försöker, som den kyrkans man han är, att sammanjämka den bild som Bibeln ger med vetenskapens, och han anmäler, inte minst, rationella skäl för purism:

Men icke des mindre hoppas jag at alla rätsinte besinna både nödwändigt och hederligt wara.

1. At man så fulkomligen kan thet språket/ hwilket man dageligen bruka måste/ at man icke behöfwer inblanda några ord/ som man af andra språk lånt hafwe. Ty the som thet giöra/ äre intet klokare/ än the som lappa ihoop en klädning utaf allehanda klutar/ ther the kunde hafwa en mycket bättre utaf et stycke/ eller enahanda tyg.” (Spegel 1712, ”Företal” [s. 9])

Hans purism får emellertid snarast sägas vara ett utslag av en strävan att troget återge det samtida språket mer än en idealiserad, fornspråksrelaterad rekonstruktion. I samma riktning pekar hans närmast häpnadsväckande öppenhet för vad som är språkliga realiteter, då han tillfogar som ett appendix en ”Förteckning på några Fransöska Ord som äro nu inblandade uti Swenskan/ och kan hende af allom icke rätt förstås” (s. 583).

När man i efterhand analyserar en ordbok ihop med alla andra som tillkommit senare, är det lätt att orättvist bedöma förstlingen. Spegels ordbok är förvisso motsägelsefull. Den uppger ekvivalenter i andra språk till svenska uppslagsord, men den är ingen två- eller flerspråkig ordbok i vanlig mening. Den återger en del citat från latinska auktorer, men den hör under inga omständig-

heter hemma i latintraditionen. Den har kallats den första svenska ordboken, men den innehåller inga konsekvent genomarbetade definitioner. Den innehåller ett slags etymologisk information, men den har inte ordens härledning som sitt huvudärende. Många artiklar sväller ut till encyklopediska utredningar, men helhetskaraktären av ordbok kvarstår. Många menar att den avses spegla det samtida svenska ordförrådet, men det är inte hela sanningen.¹

Sant är däremot att Spegels ordbok är den första svenska ordbok som inte på något tydligt sätt ansluter till latintraditionen. Men det är ändå missvisande att betrakta den som den första ”moderna” svenska ordboken. Som ett slags embryo till ordboken kan man se det ”Register Och Förklaring öfwer några gamla Ord som äre uti thenna Skriften införde” som Spegel tillfogade första upplagan av den stora lärodiikten *Gudz Werk och Hwila / Thet är Hela Werldenes undersamma Skapelse* (1685). Detta register har tycke av Verelius index (1691), låt vara inte lika utpräglat göticistiskt.

Så kan denna genomgång av den svenska lexikografihistorien få sluta där den började: med Haqvin Spegel. Om hans ordbok 1712 kan sägas beteckna början på en inhemsk svensk ordbokstradition, måste man samtidigt ha i åtanke att han snarast fullföljer en stormaktstida lexikografisk tradition i patriotisk anda. Den leder sitt ursprung till Bureus allomfattande religiösa mystik, över Stiernhielms djärva språkspekulationer och den göticistiska vurm som företräds av Verelius och andra. Spegel är med sin tids mått oerhört lärd, men han är i alla avseenden den gamla tidens företrädare, som kyrkans tjänare, som skald, som språkman och lexikograf. Hans ordbok representerar med vårt synsätt ett förvetenskapligt stadium. Det är först med den intensifierade, regelrätta tvåspråkiga lexikografin som utvecklingen mot modern lexikografi sätter fart.

¹ Egenskaperna hos Spegels ordbok har fått en noggrann genomlysning i Anna Hannesdóttirs avhandling (1998, särsk. s. 128ff. jämte vidare referenser; samma författare förbereder för närvarande en ny utgåva av Spegels ordbok på uppdrag av Svenska Vitterhetssamfundet).

Lexikografi i kontext

Inledningsvis ställdes frågan om lexikografihistoriens *raison d'être*. Kanske är det motiverat att nu påminna om den efter genomgången rapport.

Det är välbekant att traditioner som från början vilar på ett stadigt fundament av välmotiverade skäl, så småningom kan övergå i rutiner, som gör att man inte längre ställer grundläggande kritiska frågor utan gör som man blivit lärd. Det har ibland hävdats att lexikografen har sin styrka i att man inte hela tiden teoretiserar och ifrågasätter gällande konventioner utan troget följer väl utprovade praktiska regler (t.ex. Al-Kasimi 1977). Genom ett historiskt perspektiv på lexikografins framväxt kan man nå fram till en djupare förståelse av det informationsurval som brukar förekomma i ordböcker, man inser varför vi har den typ av ordböcker som vi har vant oss vid. Men dessutom kan det vara värt att fundera över vilka kriterier som gällde när informationsurvalet en gång gjordes, och man kan behöva ompröva den ursprungliga argumentationens aktualitet idag. Dessutom kan vi lära av historien att en hel del potentiell lexikografisk information genom den framväxande traditionens tyngd trängts undan. Är det rimligt, sett med våra ögon, eller var det olyckligt, att vissa alternativ gick förlorande ur gamla strider?

Lexikografihistoria är givetvis en del av språkhistorien. Därmed kan man med lexikografihistorien konfrontera samma frågor som man bör ställa till all språkhistorisk beskrivning. En sådan fråga är om det undersökta och redovisade språket naturligen är talat eller skrivet språk; en annan om det fokuserade språket är offentligt eller privat, med andra ord: något slags (officiellt) standardspråk eller snarare individuellt eller regionalt färgat, alltså lite mer speciellt språk; en tredje om språket verkligen är autentiskt eller snarare tillrättatlaggt eller idealiserat. Med den senare frågan sammanhänger problemet att avgöra hur representativa de olika ordböckerna från svunna tider är för sin tids ordförråd. I språkhistoriska framställningar åberopas ofta information i äldre ordböcker och grammatikor som argument för en ståndpunkt. En lexikografihistorisk analys kan således för språkhisto-

riker fungera som varudeklaration på ordböckerna. På annan plats har jag illustrerat denna punkt (Ralph 1992). Exempelvis kan en utpräglad normativ ordbok förväntas vara mindre representativ för samtidens spontana språk än en mindre normativ, som nöjer sig med att registrera språkbruket som det är. Saken blir inte mindre komplicerad – och riskabel för språkhistoriker – om en starkt normativ och tillrättalagd ordbok har varit framgångsrik: former som den förespråkar, men som endast har haft svag förankring i samtiden, kan ha blivit allena rådande senare. Från en modern utgångspunkt förefaller därför uppgifterna om formernas existens under en viss period rimliga, när deras autenticitet i själva verket snarare råkat bli till synes bekräftade av den senare utvecklingen. En av de frågeställningar som Lena Rogström tar upp i sin avhandling gäller just vilken relevans första-beläggen hos Serenius har vid en beskrivning av det svenska ordförrådets historia (Rogström 1998). Redan en medvetenhet om att den här typen av problem kan dyka upp gör eventuellt att vi håller en välbehövlig varningsflagg hissad under vårt studium.

Just i fråga om undersökningar av äldre tiders ordförråd respektive ett språks lexikaliska utveckling ligger det en konflikt inbyggd i att undersöka utgivna ordböcker, som ju alltid har ett syfte. Det är exempelvis inget krav på äldre lexikografer att de bara skall ta upp språkgods som är lexikaliserat, vilket kanske trots allt är ett mål för många allmänt hållna ordböcker nu för tiden. Under en period då ett standardspråk ännu är under uppbyggnad, blir ordboken lika mycket ett viktigt instrument i denna process som en pålitlig reflex av språkssystemet ("standard-creating dictionaries" i Zgustas terminologi; se ovan). Det kan som sagt innebära att mycket av det som lanseras i ordboken i själva verket har en ganska dubiös status i förhållande till det verkliga bruket. I efterhand kan sådana asymmetrier vara omöjliga att utläsa ur ordboksinformationens enbart.

Såsom redan påpekats, har det blivit normalt i modern språkhistorisk forskning att göra en distinktion mellan yttre och inre språkhistoria, varvid beskrivningen av den yttre tar fasta på de utomspråkliga villkoren för språkets förändring, medan den inre språkhistorien rör själva språkssystemets utveckling. Denna

distinktion är lika viktig vid en bedömning av det språk som kommer till uttryck i äldre ordböcker som i andra texttyper. Den kan exempelvis ge oss nyckeln till Noreens relativa ointresse för äldre lexikografi. Noreen var i grunden junggrammatiker och lade därför störst vikt vid språkssystemets förändring, den inre språkhistorien. Lexikografins framväxt kan givetvis kasta ett visst ljus över språkssystemets förändring, framför allt det kodifierade, tillrättalagda standardskriftspråkets utveckling. Men i första hand resulterar rimligen en lexikografihistorisk kartläggning i en djupare förståelse av den yttre språkhistoriens växlingar. Lexikografihistorien blottlägger processer som leder till att ett språks ordförråd hamnar inom lexikografernas kontroll, och den avslöjar vilka tekniker som efterhand utvecklas för att beskriva språkets lexikaliska sida. Härvid måste även olika icke-språkliga faktorer vägas in, såsom kommersiella hänsyn, förlagstekniska villkor osv.

Framför allt måste man beakta den ideologiska och kulturella kontext som lexikografen rör sig inom, såsom redan understrukits upprepade gånger. Al-Kasimi har framhållit att den drivande kraften bakom lexikografins utveckling varierar från kultur till kultur. De äldsta lexikografiska hjälpmedlen kom till för rent praktiska ändamål i tvåflodslandet för 3000 år sedan; arabisk lexikografi tjänade ursprungligen religiösa syften enbart; de första engelska tvåspråkiga glossarierna genererades som i Sverige av latinstudiet; amerikansk lexikografi uppstod som en nationalistisk reaktion mot att brittiska engelskspråkiga ordböcker ignorerade specifikt amerikanska förhållanden (Al-Kasimi 1977:1f.). Man kan tillägga att även inom en nationell tradition kan bevekelsegrunderna till att man framställer ordböcker variera från tid till annan.

Lexikografins historiska incitament i Sverige är, som vi sett, latinstudiet, men efterhand hamnar modersmålet alltmer i fokus för lexikografernas intresse. Dessa båda olika beskrivningsuppgifter kräver olika grepp, och länge kan man märka ett spänningsförhållande mellan de skilda syftena i de resulterande ordböckerna. I viss mån fortlever denna spänning oavbrutet fram till den enspråkiga lexikografins genombrott, alltifrån renässansens

ambivalens inför latinet som den självklara förebilden och nationalspråket som det eftersträvansvärda målet. Samma paradox karakteriserar för övrigt Johan Ihres hela verksamhet. Som innehavare av den skytteanska professuren i Uppsala hade han som primär uppgift att vårda den latinska vältaligheten. Latinet hade emellertid under 1700-talet förlorat mark till franskan som det internationella diplomatspråket och till nationalspråken i en rad andra sammanhang. I enlighet med denna utveckling lade Ihre an på svensk språkriktighet och övade med sina studenter orationer på svenska, givetvis efter latinska mönster. Härtill kom hans vetenskapliga insatser att främst gälla gotiskan, isländskan och den jämförande germanska språkforskningen. Han blev i praktiken snarare professor i germanska språk än i latin.

De tidigaste lexikaliska undersökningarna som utanför det rena latinstudiets ram ägnades svenskt material ingick i en esoterisk språkspekulation, som påvisats ovan. Den modernare utvecklingen inleds med tvåspråkiga ordböcker för rent praktiska ändamål ett stycke in på 1700-talet. I skarven häremellan faller två ordböcker, som båda är särlingar. Den ena är Haqvin Spegels, som nyss behandlades. Den andra är Jesper Swedbergs, som tidigare nämnts i förbigående. Den senare publicerades aldrig men utnyttjades i Schenbergs svensk-latinska lexikon. Ordförrådet är emellertid hämtat ur Bibeln, vilket återigen aktualiserar frågan om representativitet och autenticitet. Å andra sidan har 1541 års bibel och dess modifieringar 1618 och 1703 i många stycken fungerat som riksligare även i språkliga sammanhang, så frågan är komplicerad.

Swedberg var biskop liksom Spegel, före dem Jonas Petri Gothus och efter dem Serenius. Dessa biskopar och lexikografer i kombination hade skyldighet att ta många hänsyn, inte minst till metafysiken. Det är svårt för en nutidsmänniska att föreställa sig hur symboltyngt redan själva ordet *ord* var för bara 250 år sedan för den lärde, som även om han inte var biskop mycket ofta var teologiskt utbildad. Ordet, *Logos*, hade en betydelse som svävade mellan 'ord' och 'lära'. Att samla dessa enheter som kallas *ord* och betrakta dem som enbart något profant är för oss en självklarhet, men om vi räknar med samma attityd i det förgångna gör

vi oss skyldiga till ett grovt anakronistiskt synsätt. Få äldre lexikografer delade exempelvis Bureus extrema föreställningar, men även, låt säga, Leibniz uppfattning om språket är annorlunda än vår. Lexikografins pionjärer levde de facto i en helt annan värld än vi, och när det gäller deras världsbild var den ännu mer olik vår. De hade förvisso aldrig hört talas om Jurassic Park, varken filmen eller det ”faktaunderlag” för den som vi i någon mening räknar med. Däremot var jordens skapelse, som inte direkt uppfattades som någon Big Bang, för människorna under 1700-talet så att säga inom räckhåll – den låg enligt deras beräkningar några tusen år tillbaka, i stort sett bara hälften så lång tid som vi idag anser att det har bedrivits jordbruk i Europa.

Nutidens människor har en tendens att både underskatta och överdriva avståndet till våra förfäder och deras tänkande. Det är viktigt att komma ihåg skillnaderna, men vi skall för den skull inte glömma bort den kontinuitet som idéutvecklingen samtidigt uppvisar. Swedbergs son Emanuel Swedenborg – en av de internationellt mest kända svenskarna någonsin – utvecklade under påverkan av främst Bureus och Rudbeck ett metafysiskt system, som förts vidare in i vår tid av William Blake, Novalis, Balzac, Baudelaire och symbolisterna, Strindberg osv. Traditioner är starka, i diktningen som i lexikografen, och insatta i nya sammanhang får de ny betydelse. Strindberg fascinerades av Swedenborgs korrespondenslära (som påverkats av Rudbeck och i mycket påminner om Bureus idéer), som han försökte tillämpa i sina egna sena ”språkvetenskapliga” studier. Om Einar Már Guðmundssons väl dokumenterade intresse för Strindberg har gjort att Swedenborgs universalteori och änglatro avsatt spår i *Universums änglar* vet jag inte, men nu sluts i alla fall cirkeln.

*

Noreens dom över de äldre lexikograferna, som vi här påmint om några gånger, var inte särskilt generös – och heller inte alltid rättvisande. Man kan avslutningsvis fråga sig hur framtidens dom över nutida lexikografer kan komma att te sig. Det vet naturligtvis ingen. Avgörande blir de rådande idealen, tankeklimatet,

vilken kunskap som efterfrågas osv. Man kan ju bara konstatera, att det som vi idag kallar "new age" skulle det ha varit i det närmaste socialt omöjligt att öppet bejaka bara för några decennier sedan. Idémässigt är det mer besläktat med Bureus föreställningar än med det tänkande som modern lexikografi i dess gängse form tar som sin utgångspunkt. Det som vi i ett givet ögonblick betraktar som utveckling kan snabbt komma att ifrågasättas, och gamla avlagda föreställningar kommer understundom åter till heders. Det finns inga garantier för att det som vi tycker är självklart skall anammas av kommande släkten. Det enda vi kan göra är att efter bästa förmåga förse dem med ett rikt och väl genomarbetat material att dra slutsatser av. Själva slutsatserna får bli deras.

Litteratur

A. ORDBÖCKER

- Cavallin, Christian. 1871. *Latinskt lexicon, till läroverkens tjänst, i etymologisk uppställning utarbetadt*. 1–2. Stockholm.
- Cavallin, Christian. 1875–76. *Svensk-latinsk ordbok*. 1–2. Stockholm.
- Codex Upsaliensis C 20. Se *Latinskt-svenskt glossarium*.
- Comenius, Amos Johan. 1640 (och senare uppl.). *Janua Linguarum reserata aurea [...]. Comenii Upläste Gyllene Tungomåls Dör: Eller Alla Språks och Wettskapers Örtgårdh*. Holmiæ.
- Dalin, A.F. 1850–55. *Ordbok öfver svenska språket*. Stockholm: Författarens förlag.
- Dictionarium. Latino-Sveco-Germanicum [...]*. 1640. Lincopiæ.
- Dähnert, J. 1746 (2. uppl. 1775). *Liber memorialis Germanico-Latino-Svecicus*. Holmiæ et Lipsiæ.
- Dähnert, J. 1784. *Kurzgefasstes Deutsches und Schwedisches Hand-Lexicon mit angefügten Französischen Bedeutungen*. Stockholm, Upsala & Åbo.
- Fiellström, Petrus. 1738. *Dictionarium sueco-lapponicum, [...]*. Stockholm.

- Hellquist, Elof. 1922. *Svensk etymologisk ordbok*. Lund.
- Helsingius, Elaus Petri. 1587. *Synonymorum libellus*. Stockholmia.
- Ihre, Johan. 1769. *Glossarium Suiogothicum* [...]. Upsaliae.
- Juslenius, Daniel. 1745. *Suomalaisen sana-lugun coetus* [...] *Fennici lexici tentamen* [...] *Finska Orda-Boks Försök* [...]. Stockholm.
- Latinskt-svenskt glossarium efter Cod. Ups. C 20*. 1918–42. Utg. av Erik Neuman. (*Skrifter utgivna av Svenska forsknings-sällskapet*. 42.)
- [Lind, Olof.] 1738. *Teutsch-Schwedisches Und Schwedisch-Teutsches Dictionarium Oder Wörter-Buch, Thet År, Orda-Bok på Tyska och Swänska*, [...]. Stockholm.
- Lind, Olof. 1749. *Teutsch-Schwedisches und Schwedisch-Teutsches Lexicon oder Wörter-Buch. Orda-Bok på Tyska och Swänska*, [...]. Stockholm.
- Möller, Levin. 1745 (2 uppl. 1755). *Nouveau Dictionnaire François-Svedois et Svedois-François. En ny frantzösk och svensk samt svensk och frantzösk lexicon, eller Orda-Bok*. Stockholm och Upsala.
- NLO = *Nordisk leksikografisk ordbok*. 1997. Utg. av Bergenholtz, Henning et al. (*Skrifter utgitt av Nordisk forening for leksikografi*. Skrift nr. 4.) Oslo 1997.
- Sahlstedt, Abraham M. 1773. *Swensk Ordbok Med Latinsk Ut-tolkning, På Kongl. Maj:ts allernådigste Befallning Författad*. Stockholm.
- Schenberg, Petrus. 1739 (2 uppl. 1747). *Lexicon latino-svecanum*. Norcopiæ [et Lincopiæ]. (Svensk-latinsk del av Schenbergs latinska lexikon: *Swensk Ordabok, ther hwart och ett Swenskt ord och ordalag til thes skrifart, på Latin gifwes och vttydes*.)
- [Schroderus:] *Ericus Johannis Schroderus Upsaliensis. Dictionarium quadrilingue Suedicum Germanicum Latinum Græcum* (Cod. Thott. 2166 4:o). Med en inledning och ett alfabetiskt register över de svenska orden jämte tillägg ur *Lexicon latino-scondicum* 1637 och *Janua Linguarum* 1640 utg. av Bengt Hesselman. 1929. Uppsala.

- Schroderus Upsaliensis, Ericus. 1637. *Lexicon latino-scondicum Quo quatuor celebriores totius Europæ linguæ atque idiomata orbis, scilicet Latinum, Svecicum, Germanicum & Venedicum seu Finnicum [...] proponunter*. Holmiæ Sueonum.
- Serenius, Jacob. 1734. *Dictionarium Anglo-Svethico-Latinum*. Hamburg.
- Serenius, Jacob. 1741. *Dictionarium Suethico-Anglo-Latinum*. Stockholm.
- Serenius, Jacob. 1757. *English and Swedish Dictionary*. Harg & Stenbro.
- Spegel, Haqvin. 1712. *Glossarium – Sveo-Gothicum Eller Svensk-Ordabook [...]*. Lund.
- Stiernhielm, Georg. 1634. *Gambla Swea- och Götha-måles fatebur. Variarum rerum vocabula cum Sueca interpretatione*. 1538. Stockholm.
- Verelius, Olof. 1691. *Index lingvæ veteris scytho-scandicæ sive gothicæ ex vetusti ævi monumentis [...] collectus*. Utg. av Olai Rudbecki. Upsalæ.
- Vocabularium ad usum dacorum [...]*. 1510. Paris.
- Vocabulorum variorum expositio atqve collectio ex lingua sve-tica*. 1576. Kiøbenhavn.

B. ÖVRIG LITTERATUR

- Agrell, Jan. 1955. *Studier i den äldre språkjämförelsens allmänna och svenska historia fram till 1827*. (Uppsala universitets årsskrift. 1955:13.) Uppsala.
- Al-Kasimi, Ali. [1977] 1983. *Linguistics and Bilingual Dictionaries*. Leiden.
- Allén, Sture & Gellerstam, Martin & Malmgren, Sven-Göran. 1989. *Orden speglar samhället*. Stockholm.
- Anna Helga Hannesdóttir. 1987. "Förteckning över äldre svenska ordböcker. Tiden fram till 1850." Rapport inom projektet "Lexikografisk tradition i Sverige". Göteborgs universitet, Institutionen för nordiska språk. [Dupl.]

- Anna Helga Hannesdóttir. 1998. *Lexikografihistorisk spegel. Den enspråkiga svenska lexikografins utveckling ur den tvåspråkiga. (Meijerbergs arkiv för svensk ordforskning. 23.)* Göteborg.
- Anna Helga Hannesdóttir. 2000. "Tvåspråkig lexikografi och språkplanering i 1700-talets Sverige." *Studier i svensk språkhistoria*. 5. Förhandlingar vid Femte sammankomsten för svenska språkets historia. Umeå 20–22 november 1997. Umeå. S. 224–235.
- Einar Már Guðmundsson. 1993. *Englar alheimsins*. Reykjavík. (Svensk övers. 1996. *Universums änglar*. Stockholm.)
- Guðmundsson se Einar Már Guðmundsson.
- Gundersen, Dag. 1992. Om leksikografi som undervisningsämne. *Nordiske studier i leksikografi*. Rapport fra Konferanse om leksikografi i Norden 28.–31. mai 1991. (Skrifter utgitt av Nordisk forening for leksikografi. Skrift nr. 1.) Oslo 1992:76–87.
- Hannesdóttir se Anna Helga Hannesdóttir.
- Harras, Gisela. 1989. "Wörterbücher als Hilfsmittel der linguistischen Forschung." *Wörterbücher/Dictionaries/Dictionnaires. Ein Internationales Handbuch zur Lexikographie* etc. 1. 1989. (Handbücher zur Sprach- und Kommunikationswissenschaft. 5:1.) Berlin & New York. S. 159–163.
- Hast, Sture. 1990. *Samuel Schultze och hans Svenskt Lexicon*. (Skrifter utgivna av Vetenskaps societeten i Lund. 83.) Lund.
- Haugen, Einar. 1986. "Learned Lexicographers of the North: Seventeenth-Century Vignettes." *The History of Lexicography*. Papers from the Dictionary Research Centre Seminar at Exeter, March 1986, ed. by Hartmann, R. R. K. (Amsterdam Studies in the Theory and History of Linguistic Science. Series III. – Studies in the History of the Language Sciences. Vol. 40.) Amsterdam/Philadelphia. S. 99–105.
- Haugen, Einar. 1987. *Blessings of Babel. Bilingualism and Language Planning. Problems and Pleasures. (Contributions to the Sociology of Language. 46.)* Berlin, New York & Amsterdam.

- Haugen, Einar. 1990. *Babels forbrødring. Om tospråklighet og språkplanlegging*. Oslo.
- Hervarar saga*. På Gammal Götska Med Olai Vereli Vttolkning Och Notis. 1672. Upsalæ.
- Hesselman, Bengt. 1929. "Inledning". *Ericus Johannis Schroderus Upsaliensis Dictionarium Quadrilingue Svedicum, Germanicum, Latinum, Græcum [...]*. Utg. av B. Hesselman. Uppsala. S. V–XXVI.
- Holm, Lars. 1986. *Jesper Swedbergs Svensk Ordabok – bakgrund och tillkomsthistoria*. Institutionen för nordiska språk vid Uppsala universitet. Uppsala.
- Holm, Lars & Jonsson, Hans. 1990. "Swedish Lexicography." *Wörterbücher/Dictionaries/Dictionnaires. Ein Internationales Handbuch zur Lexikographie etc.* 3. 1990. (*Handbücher zur Sprach- und Kommunikationswissenschaft*. 5:3.) Berlin & New York. S. 1933–43.
- Hovdhaugen, Even. 1996. "Explanation in linguistic historiography – a case study." *The Nordic Languages and Modern Linguistics*. 9. Proceedings of The Ninth International Conference of Nordic and General Linguistics, University of Oslo, January 11–12, 1995. Oslo. S. 98–105.
- Johannesson, Kurt. 1982. *Gotisk renässans. Johannes och Olaus Magnus som politiker och historiker*. Utg. av Michaelisgillet. Stockholm.
- Johansson, Monica. 1997. *Lexicon lincopense. En studie i lexicografisk tradition och svenskt språk vid 1600-talets mitt*. (*Meijerbergs arkiv för svensk ordforskning*. 21.) Göteborg.
- LexicoNordica*. Tidsskrift om leksikografi i Norden utgitt av Nordisk forening for leksikografi i samarbeid med Nordisk språksekretariat. 7. 2000.
- Lindroth, Sten. 1975. *Svensk lärdomshistoria. Stormaktstiden*. Stockholm.
- Malm, Mats. 1996. *Minervas äpple. Om diktsyn, tolkning och bildspråk inom nordisk göticism*. Stockholm/Stehag.
- Noreen, Adolf. 1903. *Vårt språk. Nysvensk grammatik i utförlig framställning*. Första bandet. Lund.

- Ralph, Bo. 1992. "The Older Dictionaries as Sources for Nordic Language History." *The Nordic Languages and Modern Linguistics*. 7. Proceedings of the Seventh International Conference of Nordic and General Linguistics in Tórshavn, 7–11 August 1989. Tórshavn. S. 493–509.
- Ralph, Bo. 1995. "Ordböcker som språkmodeller." *Nordiske studier i leksikografi*. 3. Rapport fra Konferanse om leksikografi i Norden. Reykjavík 7.–10. juni 1995. (Skrifter utgitt av Nordisk forening for leksikografi. Skrift nr. 3.) Reykjavík. S. 339–350.
- Robins, Robert H. 1979. *A Short History of Linguistics*. 2 ed. London & New York.
- Rogström, Lena. 1998. *Jacob Serenius lexikografiska insats. (Meijerbergs arkiv för svensk ordforskning. 22.)* Göteborg.
- Rudbeck, Olof. 1679–1698[–1702]. *Atland eller Manheim dedan Japhets afkomne, de förnemste keyserlige och kungelige slechter ut til hela werlden, henne at styra utgångne äro [...]. I–III[–IV]*. Upsalæ.
- Rydgqvist, Johan Erik. [1849] 1863. *Den historiska språkforskningen*. Stockholm.
- Spegel, Haqvin. 1685. *Gudz Werk och Hwila / Thet är Hela Werldenes undersamma Skapelse [...]*. Stockholm.
- Stiernhielm, Georg. 1671. "De linguarum origine præfatio." *D. N. Jesu Christi SS. Evangelia Ab Ulfila Gothorum in Moesia Episcopo [...]*. Stockholmia.
- Wollin, Lars. 1989. "Hat Jöns Budde im 'Catholicon' geblättert? Niederdeutsches Lehngut in einem mittelalterlichen schwedischen Wörterbuch." *Niederdeutsch in Skandinavien*. III. (*Zeitschrift für deutsche Philologie*. Beiheft Nr. 6.) S. 77–99.
- Zgusta, Ladislav. 1992. "History and its multiple meaning." *History, Languages, and Lexicographers*, ed. by Ladislav Zgusta. (*Lexicographica. Series Maior*. 41.) Tübingen. S. 1–18.

John Sinclair

Lexical Grammar

Introduction

This paper concerns the relation between the two types of pattern that are mainly recognised as the means whereby language creates meaning. The terms *grammar* and *lexis* will mainly be used for these, but instead of grammar you will sometimes find *syntax* or *structure*, and instead of lexis you may find *semantics* or *vocabulary*. But there is always this basic distinction, of a component which produces patterns of organization and a component which produces items that fill places in the patterns; the items tend to be chosen individually, and with little reference to the surrounding text.

The title of the paper is *Lexical Grammar* and not *lexicogrammar*. Lexicogrammar is now very fashionable, but it does not integrate the two types of pattern as its name might suggest – it is fundamentally grammar with a certain amount of attention to lexical patterns within the grammatical frameworks; it is not in any sense an attempt to build together a grammar and lexis on an equal basis.

When a dichotomy is firmly established in a culture, it is difficult to find a name for it or to talk about it as a unified whole and not two different things; that is the problem here. Recent research into the features of language corpora give us reason to believe that the fundamental distinction between grammar, on the one hand, and lexis, on the other hand, is not as fundamental as it is usually held to be and since it is a distinction that is made at the outset of the formal study of language, then it colours and distorts the whole enterprise. It is worth considering how far, using modern techniques, we can get in describing a language without resorting to such a distinction.

Grammar and Lexis

The distinction between grammar and lexis is a very basic model of language; there would be no motivation to reconsider it unless new evidence gave rise to concern about its accuracy. One reason for such a model becoming so well established could be simply that before the computer age linguists were unable to describe all the complexity of language at once; since it could be represented as a framework and a set of choices to fit the frames, one of those elements could be held steady and the other varied against it. So we could forget, temporarily, about the patterns of semantic choice while we look at the organization of the structures; and then the process could be reversed, and when we come to look at the words and their meanings, then we do not consider at this point whether they are subjects or objects of clauses or objects of prepositions, if they are noun phrases, because that part of the overall organisation is suspended.

In other words, we can put forward for consideration the suggestion that this initial division of language patterning may not be fundamental to the nature of language, but more a consequence of the inadequacy of the means of studying language in the pre-computer age. When the linguist had nothing but his or her five senses, memory and internal awareness, it was difficult to analyse such a complex matter as language; consider phonetics, for example, before the sound wave could be slowed down and divided into its components. Without the ability to manipulate language externally, the observer/analyst has to leave some things steady, or hope they stay steady while other aspects of the whole are examined. And the problem is, in language, that they don't stay steady. So we should at least question the wisdom of dividing the meaningful patterns of language into two at the outset.

Abstractions

There is a related point to be made separately, but also a consequence of the position of the human observer. It is generally recognised that the meaningful patterns of language are of an abstract nature, which is one reason why they are so difficult to explain, and to use in teaching; from the perspective of grammar they are more abstract than they seem to be at first sight. It is possible that the reason for their unexpected level of abstraction is that grammar typically is realised through the common words and morphemes – that is, they seem to be familiar, but in fact many of them are multiply ambiguous and in a complex relationship with the categories that they realise (Sinclair 1999a). So grammar is superficially easy to observe but much more abstract than appears at first sight.

In contrast, the lexical patterns are very difficult to observe, because they are realised by a large vocabulary of infrequent words, and so it is not easy to work out the recurrent patterns that lie beneath the massive variation. The patterns are patterns of combination, and this compounds the problem; whereas in grammar the recurrence of frequent words makes it fairly easy to notice patterns of combination, in lexis the combinations had only been seen in a few hundred idiomatic expressions which were so remarkable that they had to be accounted for separately. With large corpora and powerful computers we are at the frontiers of a new view of language, where we can appreciate its full complexity without getting hampered by the detail.

It is thus no accident that linguists up till now have developed grammars much more than dictionaries and lexicons; we tend to have very elaborate grammars, which contain intricate apparatus with ranks and hierarchies and structures and all sorts of categories, with many different kinds of organization and in contrast we have very, very simple models of lexical structure, which are mainly one-dimensional, based on the word. There is an ad hoc set of terms for multi-word units like idiom and cliché and saying and proverb, but all these are ill-defined terms, and

there is no other network of interconnections between one word and another.

Again this disparity in our descriptions does not necessarily reflect the nature of language, but rather it reflects our collective inability to process language with sufficient power and understanding to see that the complexity of the language as seen from a lexical point of view is just as great as the complexity of the language as seen from the grammatical point of view. So we may expect that simple artefacts like dictionaries will give way to more complex lexical architectures – indeed the development of dictionaries with an influence from corpus research has begun to move in this direction.

Meaning and Structure

There is one consequence of the initial separation of language patterning into two contrasted types that could be very important. To bring it out clearly we will use the terms *meaning* and *structure*. In brief, the point is that if we ignore the meaning while we are describing the structure, then of course we have removed the meaning and will not be able to get it back while we are focused on the structure. That is one way of expressing the problem of grammar, and it has been obscured from careful examination by a kind of meaning substitute. This is the curious terminology that we use, things like positive/negative, singular/plural, active/passive, and so on. If we look at them carefully, these terms are of course quite substantially inaccurate. "Singular" does not always mean "one," and "plural" does not always mean "more than one". "Present" does not always refer to the time of the utterance, and "past" certainly does not always mean some previous time. We have learned as part of our culture to suspend disbelief when we encounter these terms, and apply a rough criterion of *mutatis mutandis* to their interpretation; "singular" means "not more than one, if whatever it is is countable, otherwise general reference". The point is that because these terms are not sensitive to the meaning, then they cannot

actually be used directly to elucidate the meaning of text. The distinctions could have any labels at all, and at best they have a mnemonic function (this argument is well-supported by the retention in current grammars of terms like "finite" or "voice", which bears little relation to meaning in current English).

The only meaning that grammar provides is differentiation. From the *valeur* of Saussure to the systems of systemic linguistics and the choices of transformational grammars, then, the only way in which grammar creates meaning is by setting up mutually exclusive choices, and it exists purely as a record of the choice itself; the significance of the choice – whether a past tense verb relates to past time or present or future time or modality – is determined elsewhere.

If we now view the structure/meaning divide from the other perspective, and look at semantics without structure, then the typical way of presenting the meaning is the dictionary. A dictionary simply lists in an arbitrary order, which we call alphabetical¹ the items that it regards as being meaningful, which are usually the words of the language, and it tries to assign one or more meanings to each of the words. That is the characteristic model of a dictionary. The meanings are denied access to the structural organization that can put them together and show how they work. For example one meaning given in a recently-published dictionary for the word "white" is "counterrevolutionary, very conservative, or royalist"; if this meaning is still current it would take some ingenuity to specify the structural circumstances under which it could occur.

So therefore substitutes, again, are offered, this time standing in for the the linguistic organization that has been discarded. There are in semantics two major types of organization that have been imported; one of these is referential semantics, and the other is logical semantics; let us consider them in turn. The assumption of referential semantics is that meanings are organized with

¹ Alphabetical order is an order whose only virtue is that it is taught to all literate members of societies which use it. The fact that it is the only means of organising the vocabulary of a language merely emphasises the failure of linguists to find a better one.

reference to the world outside; words have meanings which can be understood by indicating objects, events and attributes in the world to which they refer; for abstract entities there is the "figurative" mode which works analogously. This is simple and seems to be broadly usable for a very large range of usable phenomena, and is widely used in education, but from a theoretical point of view it is absurd. Consider the proposal for a moment – on the one hand there is language, which we know is a highly organized phenomenon that operates under major constraints such as linearity, and on the other hand there is the world, which after thousands of years of research we still see as pretty chaotic, exceptionally complex and totally unable to be encompassed in a simple description. We are asked to accept that reference to the world can elucidate the structure of language? We have some reason to believe that language can elucidate some aspects of the world, but hardly the other way round. At best the referential links can help in for example supporting the acquisition of language by a child, before the child can cope with semantic abstractions.

The other type of imported semantic structure that is popular is logical semantics. It seems to have some advantages, being rigorous and much of it being quite close to the patterns of natural language (as well it might be, being derived from them). But it is crucial to the understanding of natural language that the organisation is *not* exact, and is not reliable as an indication of logical relationships. As with the definition of terms in grammar, there is again the problem of the partial fit, the inexact fit.

Here is a brief example to show the problems of relying on logical analogies too closely. Many commentators have noted that the "conditional" *if* does not always have its logical force, for example in the following instances culled from a large number of candidates in just one category of The BNC Sampler (spoken business):

...which is obtainable from Christian Aid if people want....
 I'm just thinking for the meeting if we could photocopy some Yes
 I'll be actually chairing the meeting for him. So so if you'd like to
 kick off...

Mm. Yes. Mm. Erm Mm I could if I could just pick up one other point about you know...

And another one noticed casually in reading:

If you believe me, I swung along that road whistling. (*The Thirty-Nine Steps*, John Buchan)

And one which has already occurred in this paper (one of two):

If we look at them carefully, these terms are of course quite substantially inaccurate.

The Axes of Language Patterning

We now move our perspective to a closely related dichotomy that has long been recognised in language description – the two fundamental axes of language patterning, the paradigmatic and the syntagmatic. They are usually depicted as horizontal and vertical, with the syntagmatic axis on the horizontal, because the languages of modern Europe are written in horizontal lines, and the paradigmatic on the vertical.¹ The paradigmatic axis specifies the possible choices at a particular position on the syntagmatic axis, and the syntagmatic axis controls the structure which is being elaborated. So what we observe in language text is the syntagmatic; the paradigms are the total of what might have been chosen instead.

¹ The most available instance of this mechanism in some societies is the fruit machine, which used to be found in almost any public house in UK. A fruit machine consists of three revolving cylinders, each of which bears a number of drawings of fruit. The player pulls a handle at the side which causes the cylinders to spin independently of each other, and come to rest in a chance combination. A row of three fruit thus appears in a central grille, and if the row corresponds to one of those in a list on the side of the machine, (eg banana banana banana) then the player wins, and receives several times his or her stake. The central row corresponds to the syntagmatic axis, and the cylinders contain the paradigmatic choices; a well-formed structure is one of those that wins a prize.

Now, one of the interesting things about these two axes is that they cannot be simultaneously observed; you must hold one of them steady in order to look at the other. We shall return to this point, but no doubt this is the reason why we have had the division into grammar and lexis from an early stage. It is important to notice that the theoretical development of grammars in recent years has moved across this divide. If we were to map the "grammar" composite (including syntax, structure) and the "lexis" one onto the two axes, then the obvious pairing would be grammar on the horizontal axis and lexis on the vertical – a model of language often called the "slot-and-filler" model, the one presented at the outset of this paper. The syntactic structures form a series of slots, and these are filled with choices from the dictionary. The well-known models of transformational grammar are partly structured in this way, for example at the interface between the phrase structure and the lexicon, where the phrase structure specifies the features that any word must have in order to make a well-formed sentence, and the lexicon associates each word with a bundle of features. However, other influential models insist that they are primarily, if not exclusively, paradigmatic – notably Systemic-Functional Grammar (see Halliday 1995:15).

The syntagmatic patterns of language are not given meaning in a paradigm grammar, nor, of course, are they given meaning in a dictionary type of lexis. The syntagmatic patterns in a grammar are either offered as related through a common node, or they are simply declared. The syntagmatic patterns of lexis only appear in the byway of idiomatic phrase, where they are offered as joint realisations of a single meaningful unit, indicating that they have no meaning in themselves.

Let us consider the grammatical positions a little more. In phrase-structure rules like

$$S \rightarrow NP VP$$

the only relationship between NP and VP is that they are both derived from S in the same operation; their sequence is also

determined in this single step. In the early days of generative grammar there was a plus sign in between NP and VP,

$$S \rightarrow NP+VP$$

but this signalled a quite spurious relationship pertaining on the syntagmatic axis, and became unfashionable.

Where syntagmatic patterns come into being by declaration, there is no explanation of where they come from or how they are to be deployed. The structure of an English clause is said to involve Subject-Predicator-Object-Adjunct, for example, but these categories are mutually defining, and do not have meaning until they are mapped into sets of choices, for example that a transitive clause is one without an object. So, neither in the study of the lexis of the language nor in the study of the grammar of the language are the syntagmatic patterns given meaning. This is to a great extent because there is no framework within which they can be shown to have meaning, because meaning is largely held to reside either in the grammatical choice – on the paradigmatic axis – or in the lexical choice of a word to deliver a meaning.

Syntagmatic Meaning

There is no effort, let us say in summary, to discover or create meaning on the syntagmatic axis; it is the responsibility of a paradigm grammar to build in all possible syntagmatic meaning as constraints on the paradigmatic choices. But such a venture would be remarkably complex, so in practice those grammars fail to describe carefully enough the *combinations* of choices that are just as central and meaningful and rule-governed as the single paradigmatic choices. They give tacit approval to the well-formedness of millions of sentences which range from the odd to the bizarre, and – by claiming as a series of choices phrases which we know to be a single choice – they claim large amounts of meaning which we know those choices do not create.

In corpus linguistics, by contrast, we have to work on the assumption that meaning is created on both axes; for want of more accurate information we may assume that they contain equal meaning potential. There is no reason why one should have a priority in meaning potential over the other. We assume a rough balance between what I have called (Sinclair 1996) the *phraseological tendency*, the tendency of a speaker/writer to choose several words at a time, and the *terminological tendency*, the tendency of language users to protect the meaning of a word or phrase so that every time it is used it guarantees delivery of a known meaning. As we get to know more, these assumptions may well be revised.

Above we have presented a model of language as a balance between opposing forces related to the two axes of language patterning, and above that is an assertion that the two axes cannot be simultaneously observed; these sound like good reasons for keeping them apart, and describing them separately. However, the argument of this paper is that if pattern and meaning are to be aligned, then the two axes have to be inter-related for as long as possible in the description. Consider, for example, the classic model where a choice is made on the paradigmatic axis, which will lead to a particular word appearing in the text. Now it is an axiom of the present approach to corpus linguistics that meaning and context are inter-related in such a way that involves at least partial coselection; so the knock-on effect of a paradigmatic choice will be felt on the syntagmatic axis. If we start from the other axis, then any existing or proposed pattern of choice on the syntagmatic axis provides a framework for the interpretation of any choice to be made on the paradigmatic axis.

Practical Consequences

The remainder of this paper gives some indications of the direction in which this argument is heading and the kind of consequences it is likely to lead to. First we will re-examine the

nature of choice and meaning, then look further into the "meaningful" terminology of grammar, and finally pose a question about an important type of meaning that is largely ignored by both the grammatical and lexical traditions.

Meanings from Nowhere

Let us begin by revisiting the information-theoretic model of paradigm grammar, which says that choice equals meaning, that the number of choices determines the amount of meaning available in each case, and the precise positioning of the choice in the structural framework determines much of the type of meaning that will be created by the choice. A description within this model must take great care that each set of choices is actually relevant and applicable at each point. Because if it is not – if another factor in the environment is affecting the range of choices on offer, then unless the grammar is revised it is creating more meaning than is in fact available.

If this manufacture of illusory meaning is institutionalised throughout a complex grammar, there are two obvious consequences. One is that the grammar (and the grammarians) are misled into thinking that their apparatus is more powerful than it actually is; the other is that there is little meaning left over to be assigned by the lexical structure of the language. Now that it can be demonstrated by corpus evidence that a large proportion of the word occurrence is the result of co-selection – that is to say, more than one word is selected in a single choice – every time that this can be demonstrated there is one less item of meaning to be allocated to the grammar. If you have two words that are selected in the same choice, then they cannot be independently selected. Early estimates were that up to 80% of the occurrence of words could be through co-selections, which would leave, of course, only 20% for the sort of independent paradigmatic choices of the grammar. A recent paper by May Fan (1999) gave hard evidence for this in regard to one of the common verbs in English.

Let us work through a characteristic example. There is a phrase in English, a common recurrent phrase, "out of the corner of my eye," as in "I saw something out of the corner of my eye." There are seven words in the phrase, and they all simultaneously choose one unit of meaning, to do with peripheral vision. Within this primary meaning, there are one or two variants of individual words, and this is where the corpus is essential, because the intuition cannot be relied on: "out of" can sometimes be replaced by "from", and "my" is a possessive adjective that can have other, but probably only singular forms; people do not collectively see things out of the corners of their eyes, so I think "their eyes" is going to be very unusual. This is the full extent of the variation associated with this phrase; the remaining words are fixed, and do not realise any choice beyond the first, overall choice of meaning. So neither of the occurrences of *of* above are the normal occurrence of the preposition, because *of* is fixed in this phrase,¹ and so are "the", "corner" and "eye". The word "my" can be alternated with other possessive adjectives. So here we have a seven-word phrase which realises one overall choice and at most two subsidiary choices. The choice between "out of" and "from" here is a stylistic choice rather than a choice that delivers a totally different type of meaning – there are not two different places, and "out of" and "from" are just different ways of expressing the same basic position.

These single choices can consist of seven words with ease; the phraseology of English quite frequently produces co-selections of five, six and seven words, and there are even some of up to twelve. In this connection Miller (1956) comes to mind. Miller showed that for most people the short-term memory handles seven items with ease.

¹ It would be a digression to argue here that *of* is not a preposition in such structures; for that see Sinclair 1991.

Cross-border categories

Corpus evidence consistently shows that the ways in which a meaning can be realised extend well beyond the definitions of grammatical categories. In pointing out above that grammatical terminology did not correspond to semantically coherent categories, we did not tell the whole story. Consider a term like "negative", which will contrast with "positive" in a two-term system of "polarity". There are a number of realisations of grammatical negatives in English, "no" and "not" and so on. There are also semi-negatives like "hardly" and "scarcely", which share a number of features with true negatives, but not all; these are not normally considered as grammatical negatives.

But there are also morphological negatives like the prefixes "un-" and "in-", which are not recognised in a clause grammar, so that "I am unhappy" is positive and "I am not happy" is negative. We also find that negation as a concept can be lexicalised, so that the verb "refuse" for example has a negative force; "he refused to go" is the same as "he would not go" and yet it is a positive clause in the grammatical sense.

It is easy to understand the grammarian's wish to keep negation pure and simple; to accept lexicalised negation is a slippery slope, and no-one knows what lies at the bottom of it. But if we are intent on elucidating the meaning of running text by analysis, then all these different ways of indicating negation are perfectly acceptable realisations, and if supported by corpus evidence we can take them all together, straddling the borders between grammar and semantics. This straddling is an important feature of lexical structure; lexis is not the residue of a grammatical description, but a different way of describing the same events; it is not bound by the conventions of grammar, and it can recognise a wide variety of realisations of meaningful choices.

The grammarian is left in a dilemma; the more sensitive grammars recognise that categories of meaning like "negative", "modal", "possessive" can readily be lexicalised – or to be more neutral, can occur in grammatical or lexical or morphological realisations – so a complex realisation route is devised for them.

The particular way in which they are realised is then of secondary importance compared with the primary creation of meaning, which is the operative process. The question must arise of the relevance of, for example, the grammatical choice between positive and negative to the study of meaning when negative meaning can be created in so many alternative ways; and, more fundamentally, how valuable is it to be able to point out that there are many clauses which are grammatically negative but in relation to meaning, positive, and vice versa?

Semantic Prosody

Another important point to be made in the study of lexical grammar is the emergence of many latent categories of meaning, which have not been recognised in published grammars, and only occasionally in the very latest dictionaries. The first to be noticed were of the type "something nasty" or "something worrying" or "disturbing"; later others like "something magnificent", "socially appropriate", "positively constructive" etc. These are showing up as repetitive categories that are neither completely grammatical nor completely lexical but are nevertheless very important from a structural point of view. So once again we have to allow for the meaningful categories not to be confined within the grammar as it is normally presented, and if we divide language into these two major categories, then we will never be able to get them satisfactorily together again; also we have to add that the grammar cannot be trusted to set up such essential categories of meaning because it is not sensitive to them.¹

Here it has to be said that the perceptions of native speakers are not to be trusted either; the referential element in meaning is frequently assigned a priority over the attitudinal, for reasons that are not justifiable; clearly an awareness of both aspects of

¹ Here the new "Pattern Grammars" (Hunston and Francis 1999) take the innovative step, guided by corpus evidence, of associating some of these meanings with structural patterning.

meaning is necessary for accurate deployment of the lexical item, and if this is not available it is arguable that more difficulty may arise from a mistake on the pragmatic side than on the referential. To give a real-life example, in the preparation of a dictionary for native speakers of English by the *Cobuild* team, there was a strong feeling among editors and publisher that whereas for learners of English it may be necessary to state the attitudinal meaning, this is already available to native speakers. So the *Cobuild* definition (1987) for *scrawny* is "unpleasantly thin and bony", while in *Today's English Dictionary* (1995) it is "thin and bony"; the two dictionaries define *prattle* identically but *Cobuild* adds "an informal word, often used showing disapproval".¹

Word Class

Professional linguists should not be surprised to experience a rather disturbing effect from the massive surge in the availability of evidence and the growing sophistication of the tools for examining it and testing hypotheses against it that corpus linguistics has brought. Some of the vague but useful categories of traditional language analysis, which have served humans well for centuries, are not easily replicated in computational routines; for example "parts of speech" or "word class" labelling. Human beings have little difficulty assigning words to a dozen or so word classes, but machines have exposed just how untidy a categorisation this is. For English, which has had a lot of attention over many years, there is little or no consensus about how many labels there are – the variation from one analysis to another is very large – or how they are defined. The persistence of researchers has resulted in a significant movement of focus, so that the process is now called "morphosyntactic tagging" – in other words it was found necessary to use some syntactic

¹ This was one of the few arguments that I, as Editor-in-Chief of *Cobuild*, lost; but I am still puzzled at the conviction that native speakers may need to know the referential meaning of a word but not its attitudinal/pragmatic one.

information in order to complete what was originally a morphological analysis.

This movement of focus is well recognised in corpus linguistics – the need to examine the context of an item in order to determine its function or meaning. But nothing seems able to shake belief in the underlying assumption that all the words of a language naturally fall into a small number of classes. The information from a computer examination of a corpus suggests quite otherwise, as I have argued on several occasions.¹ Since few inflections survive into Modern English, and since one of the most productive areas of development in the modern language is the ability of words to move across word classes, it may be preferable to accept what the corpus seems to be signalling, which is the need for a major overhaul of the notion of word class.

In general, we must move toward a theory that reconciles the paradigmatic and the syntagmatic dimensions and allows the description of the language to remain sensitive to both dimensions for as long as the correlation is productive; no doubt there will be some residue of specifically grammatical and specifically lexical information after that stage, but we must wait to see what it is, and what categories and processes are best used to describe it.

¹ (a) in Tickoo (ed.) 1989, reprinted as Chapter 6 of Sinclair 1991, I showed that the second commonest word in English, *of*, had very little in common with other prepositions, and was mainly used in a unique syntactic function.

(b) I followed this up in Sinclair (1999), where I argued that most of the common words in English have individual patterns of occurrence, and do not fit into the general word-classes.

(c) As a contribution to the NERC Report (1996), I pointed out that in English, as well as words which function as nouns, and those which function as verbs, there is a substantial class which function as both (I called them *norbs*). This is a kind of underspecification which postpones a very difficult set of decisions until perhaps the analytical system is better able to deal with them.

Lexical Structure

At present, the lexical structure is presented separately, insensitive to the grammar in the same way as the grammar is, traditionally, insensitive to the lexis. It is probably a valuable exercise to prioritise the lexical patterning and to push a lexical description as far as it is reasonable to do so; the justification is that so little research has been done in this area, especially as compared with the immense attention that the grammar has had over the centuries. But such an effort should not be misunderstood; it must be seen simply as an interim step towards an eventual holistic description, and there is no imperialistic dimension to lexical description.

In the meantime, there are structures of a particularly lexical nature that are worthy of attention, and which are introduced in recent publications, particularly Sinclair (1998). These begin with *collocation*, the co-occurrence of words, and go on to *colligation*, which in this work is defined as the co-occurrence of words with grammatical choices, then *semantic preference*, which is the co-occurrence of words with semantic choices, and *semantic prosody*. The semantic prosodies express attitudinal and pragmatic meaning; they are the junction of form and function. The reason why we choose to express ourselves in one way rather than another is coded in the prosody, which is an obligatory component of a lexical item.

The ways in which the prosody is expressed are extremely varied, and seem to have no limits as to position or shape; we can thus anticipate severe technical problems in retrieving them computationally. This is the central problem in analyzing open text and one of the principal reasons that the performance of devices which depend on some kind of language understanding is so poor. At the present time the goal of the machine understanding of language is far more difficult than it needs to be, because we are not using appropriate theories – once the meaning created by lexical structures becomes available, and integrated with what we already know through grammar, then theories will be articulated that predict the prosodies and the

computer will then know where to look for them. These theories will be developed from the kinds of hypotheses that are taking shape in corpus-driven linguistics.

Example

Let me give as a conclusion an example of the kind of semantic prosody that I'm talking about. Consider the English word *effort*; it is a countable noun and so it has a plural, *efforts*. And one of the most notable collocates of *efforts* is the word *to*, which follows *efforts*, and which is the infinitive marker. So essentially we are focussing on a structure which has a core of *efforts* plus an infinitive. In the Bank of English¹ in Birmingham, which is the reference corpus that I normally use, the one that lies behind the *Cobuild* publications, there are 9,617 instances of *efforts* followed by *to*. For the Figure, the computer has selected 21 by the simple expedient of picking the first one in text sequence, then dividing 9,616 by 20 (= 480 in round figures) and then selecting the 481st, 961st etc instance through the corpus.

If we examine these, it becomes fairly clear that we use this phrasing – we talk of "efforts to" do something – when they appear to be very unlikely to succeed, to be heading for failure, or already unsuccessful. In other words, the prosody that appears in almost every example is the speaker/writer's prejudgement of the efforts, that they are heading for failure. So when we are discussing the machine understanding of language, if we were to talk of the "efforts" of computational linguistics "to" comprehend natural language, we would imply that they are doomed to failure. There are a number of adjectives, for example, like *hysterical*, *frantic*, *futile*, *strenuous*; verbs like *blunder*, *hamper*,

¹ At the time of retrieval the Bank contained almost 350 million words of broad general English text, from native speakers in many parts of the world, their spoken and written expression. The corpus is jointly owned by HarperCollins, publishers, and The University of Birmingham, and access to it can be arranged via the Cobuild Home Page, <www.cobuild.collins.co.uk>

were overwhelmed; people close ranks against efforts, or achieve things despite efforts; efforts exhaust us, and so on. So if this is a representative sample of the behaviour of the word, we can expect to find in the left-hand cotext of *efforts to*, some indication of the likely failure of the efforts.

In a contrast which is almost ironic, we can expect to find in the right-hand cotext a set of verbs which are creative, which talk about creative action, like *please, revive, work together, protect, support, gain, raise, activate, kindle (a debate), help, give (the city something good), save*, etc. So before a reader/listener discovers that the efforts are to do something constructive and beneficial, they are already sabotaged. Our first draft of the lexical item that has as its core *efforts to* will thus contain three elements of structure – the core, the semantic preference for a verb of constructive action, and the semantic prosody of anticipated failure. The selection of the item is controlled by the prosody, because the whole point of expressing oneself in this way is to pre-evaluate the actions, which would otherwise be positively evaluated by the reader/listener.

It is likely that other expressions with structural similarity to this tentative item will be found; *attempts to* may be similar,¹ etc. The singular forms, *effort, attempt* may show some tendency in the same direction. A set of forms sharing similar meanings could be a further step in the mapping of an organisational framework for lexis. This work is only beginning; a few probes have been made into the lexical structure of the language, and some tentative hypotheses have been formulated.

¹ The appearance of *despite, failed, unsuccessful, desperate, repeated, several*, as very significant collocates of *attempts to* suggests considerable similarity, but measures have not yet been devised to compare collocational profiles.

Conclusion

Despite the recent rush to welcome corpora into the resource collections of many students of language, we must note that the vast majority of work with corpora still takes place under the assumptions of pre-corpus linguistics, and is thus insensitive to the possibilities put forward here. It is clear that the first step towards a new view of language has now been taken by the linguistics profession, in recognising that corpora are relevant and useful; this has been effectively completed approximately thirty years from the advent of electronic corpora. It is only natural that to begin with scholars will appreciate the security of familiar concepts in engaging with such a total revolution in the availability of evidence of usage, and only gradually will they accept that some of those concepts are sorely in need of being revised and updated.

The initial separation of grammar and lexis in language description, and the subsequent prioritisation of grammar at the expense of lexis, is one of the most firmly-held positions among theoretical and descriptive linguists, and it will take some time before it is held up to scrutiny and approached with an open mind.

References

- Fan, M. 1999. "An investigation into the pervasiveness of delexical chunks in authentic language use and the problems they present to L2 language learners." In: Berry, R. et al. (eds.): *Language Analysis, Description and Pedagogy*. Language Centre, HKUST, Hong Kong.
- Miller, G. 1956. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." *The Psychological Review*, vol. 63:81-97.
- NERC Report 1996; *Network of European Reference Corpora Project Report*. Pisa, Giardini.

- Sinclair, J. 1991. *Corpus Concordance Collocation*. Oxford, OUP.
- Sinclair, J. 1996. "The Search for Units of Meaning." In: *TEXTUS* vol. IX no. 1.
- Sinclair, J. 1999a. "A Way with Common Words." In: Hasselgård, H. and S. Oksefjell: *Out of Corpora*. Rodopi, Amsterdam & Atlanta.
- Sinclair, J. 1999b. "The Lexical Item." In: Weigand, E. (ed.): *Contrastive Lexical Semantics*. Amsterdam/Philadelphia: John Benjamins. (Volume 17 of series *Current Issues in Linguistic Theory*), pages 1–24.

Reference Works

- Collins English Dictionary*, Hanks et al. 1998.
- Collins Cobuild English Language Dictionary*, Sinclair et al. 1987.
- Today's English Dictionary*, Sinclair et al. 1995.

Emma Sköldberg

Lever idiomerna i svenska ordböcker sitt eget liv?*

This article presents a study of 40 Swedish idioms. Firstly it focusses on to what degree the expressions are listed in five contemporary Swedish dictionaries, and secondly to what extent the stated form of the expressions in the dictionaries corresponds to the actual usage of the expressions as they are reflected in the text corpus Parole. Among other things, the study indicates that there is good correspondence in approximately half of the cases. However in other cases the differences are more apparent. For example one often sees a distinct variation in the usage of an idiom in Parole, but this variation does not appear in the dictionaries. In other cases the opposite is true, i.e. the usage in Parole is relatively homogeneous but, in spite of this, alternative forms are shown in the dictionaries. The principal aim of the study is to initiate a discussion on the principles that are applied when listing idioms in dictionaries and on whether text corpora are utilised to their full potential.

1. Inledning

I samband med mitt avhandlingsarbete – som framför allt rör formell och semantisk variation hos svenska idiom – har jag ibland förvånats över ordboksuppgifter vad gäller vissa uttryck. Som exempel kan jag nämna idiomerna *lägga lök på laxen* som just har formen *lägga lök på laxen* i samtliga ordböcker jag har konsulterat. Granskar man däremot bruket i Parole¹ är bilden en annan. Här följer några förtydligande exempel ur textkorpusen; markeringarna med fet stil är mina:

- (1) – Och som **lök på laxen** kom alla budgetunderskott.

* Artikeln har skrivits inom ramen för min doktorandtjänst som finansieras av Bokförlaget Natur och Kultur.

¹ Parole ingår i Språkbanken vid Institutionen för svenska språket, Göteborgs universitet. Korpusen består av närmare 25 miljoner ord och utgörs främst av tidningstexter och skönlitterära texter från 1970–1990-talet.

- (2) Nu är Stenungsund känt i hela landet som Sveriges mest skattetyngda kommun och därtill kommer som *lök på laxen* en hel flora av kommunala skandaler.
- (3) Som *lök på laxen* ser vi i dag också en tendens till bakåtsträvande fackföreningar.

I Parole föregås samtliga användningar av *lök på laxen* av ett *som*. Uttrycket kombineras också ofta med verbet *komma*. Men i vilken mån är detta exempel representativt? Det har jag för avsikt att diskutera i den här artikeln. Jag kommer att redogöra för en undersökning av 40 vanligt förekommande svenska idiom. Samtliga av dessa har ett tydligt semantiskt innehåll samt innehåller konstituenten som i princip skulle kunna varieras med en synonym eller vad gäller morfologiska kategorier som exempelvis numerus och species.¹

För det första har jag granskat i vilken utsträckning de 40 idiomerna upptas i fem samtida svenska ordböcker. Ordböckerna är *Illustrerad svensk ordbok* (ISO) (1964), *Svensk handordbok* (SHO) (1966), *Svenska ordboken* (SOB-en) (1995), *Nationalencyklopedins ordbok* (NEO) (1995–96) samt *Bonniers svenska ordbok* (BSO) (1998). Ordböckerna skiljer sig åt på ett flertal punkter; de har bl.a. givits ut på skilda förlag och utarbetats av olika ordboksredaktörer. Det finns även tydliga skillnader vad gäller böckernas tillkomstår, omfång samt avsikter med respektive verk. Samtidigt utgör de tillsammans en stor och viktig del av de synkrona svenska ordböcker som finns att tillgå idag.

För det andra har jag excerperat samtliga användningar av de 40 idiomerna ur Parole. Utifrån dessa belägg kan jag dra vissa slutsatser om hur det faktiska bruket av uttrycken ser ut, vilka formella variationer som förekommer etc. Därefter har jag jämfört i vilken mån bruket i Parole överensstämmer med de uppgifter som finns i de ovan nämnda verken.

¹ Dimitrios Kokkinakis, Institutionen för svenska språket vid Göteborgs universitet, har varit till mycket stor hjälp då han gjort en grov sammanräkning av frekvensen av olika idiom som ingår i Press 97 (som också ingår i Språkbanken). Samtliga uttryck som granskats i denna undersökning återfinns bland de vanligast förekommande i det pressmaterialet.

2. Kortfattad forskningsöversikt

I en uppsats som denna är det givetvis relevant med en diskussion kring *idiom*. Med begreppet avses som regel en kategori som inte är möjligt att klart avgränsa – gränsen mellan idiomatiska uttryck och bland annat syntaktiska nykonstruktioner och kollokationer är i högsta grad flytande (jfr Kjellmer 1996). Traditionellt betraktas idiomerna som relativt fasta uttryck som uppvisar klara restriktioner vad gäller exempelvis böjningar och konstruktionstyp (se t.ex. Anward & Linell 1976; Fleischer 1997). Man brukar även säga att idiomerna inte är kompositionella, d.v.s. idiomets konventionaliserade betydelse går inte att härleda utifrån de ingående konstituenternas betydelse. Men det senare har kommit att ifrågasättas allt mer. Man har framför allt betonat de skillnader som finns mellan olika idiom (jfr t.ex. Cacciari 1993; Nunberg et al. 1994).¹

Av litteraturen framgår det att idiomerna granskats ur såväl allmänlingvistiska som psykolingvistiska perspektiv. Uttrycken har också intresserat lexikologer och lexikografer (jfr Corpas Pastor 1996; Everaert 1995). Forskare har bland annat undersökt om idiom och andra fasta fraser är upptagna i ordböcker (se exv. Alexander 1992; Kjellmer 1996). Man har även diskuterat under vilket eller vilka lemman uttrycket skall anges (se t.ex. Benson 1985; Lorentzen 1996) eller om hela idiom skall utgöra egna lemman (jfr Botha 1992). Ett flertal forskare har även granskat idiom och formell variation samt huruvida denna fångas i ordböcker (jfr exv. Krohn 1994; Moon 1996; Sköldberg 1999). Men det har inte, vad jag vet, genomförts någon studie som i någon större utsträckning liknar min egen. Ett undantag är dock Ulla Clauséns undersökning från 1996 av 23 idiom – en förekomst av respektive uttryck – på *Dagens Nyheter*s ledarsidor. Författaren menar att skribenterna gärna använder idiom, men att uttrycken sällan har den form som vi är vana att finna i ordböckerna (Clausén 1996:38). Det är emellertid svårt att dra mer långtgående

¹ Resonemanget kan och borde utvecklas mer, men tyvärr finns det inte utrymme för detta här. För mer utförliga diskussioner kring idiom och idiomdefinitioner, se därför bl.a. Wood 1986 och Corpas Pastor 1996.

ende slutsatser utifrån ett sådant fåtal användningar av varje uttryck, men detta problem kan jag till stor del undvika genom att undersöka ett flertal användningar av ett och samma idiom.

3. Idiomerna och deras förekomst i ordböckerna

Vilka idiom har jag då undersökt och i vilken mån förekommer de i Parole respektive i ordböckerna? I spalten längst till vänster i tabell 1 nedan anges vilket idiom det gäller.¹ Därefter återges de söksträngar som jag använt vid excerperingsarbetet samt antalet belägg för respektive idiom i Parole. Ett kryss "X" innebär att någon idiomform av idiomerna är angiven under ett eller flera lemmen i ordböckerna. Till exempel har angivelserna *sanningen kommer nog i dagen* (SHO) och *lägga i dagen* (NEO) betraktats som idiomformer av idiomerna *komma i dagen*. En divis "–" innebär att uttrycket inte presenteras i form av en fast fras i ordboken.²

Av tabell 1 framgår det bland annat att antalet förekomster av respektive idiom i Parole ligger mellan 21 (*falla ngn på läppen*) och 112 (*få bukt med ngt*). Det är också tydligt att det råder en klar skillnad mellan huruvida idiomerna upptas i ordböckerna eller inte, men detta är föga förvånande med tanke på böckernas omfattning och karaktär. Ett flertal idiom nämns emellertid i samtliga verk, exempelvis *gå upp i rök*, *hamna på efterkälken* och *när det kommer till kritan*. Men det finns också ett uttryck som inte nämns i form av ett idiom i någon av ordböckerna, nämligen *gå i ngn's fotspår*. I vissa fall ser man även att idiomerna inte nämns i de tidigare utgivna böckerna men däremot i de senare. Detta gäller till exempel *ha is i magen*, *sopa ngt under mattan* och *få kalla fötter*. En orsak till detta kan vara att ordboksförfattarna tagit intryck av varandras arbeten. Ett annat möjligt skäl – som exem-

¹ Av utrymmesskäl anges endast en eller några former som är vanlig(a) i Parole.

² Ibland nämns konstruktionen i ordboken, men inte som idiom eller annan fast konstruktion. Detta är framför allt tydligt i NEO där de saknade idiomerna betydelser anges som en betydelsenyans hos huvudordet.

pelvis i fallet med *få kalla fötter* – är att uttrycket med tiden blivit allt mer etablerat i det svenska språksamfundet.

TABELL 1. *Antal belägg för respektive idiom i Parole samt idiomets eventuella förekomst i respektive ordbok.*

	PAR- OLE	ISO	SHO	SOB -en	NEO	BSO
1. dra (stora) växlar på ngt ("växellväxlar" "på")	21	X	X	-	X	X
2. en röd tråd/den röda tråden ("rödlröda" "tråd.*")	87	X	X	X	X	X
3. falla ngn på läppen ("på" "läppen")	21	X	X	-	X	X
4. få bukt med ngt ("buktt" "medpå")	112	X	X	X	X	X
5. få fotfäste ("fotfäste")	47	X	X	X	-	X
6. få kalla fötter ("kalla" "fötter")	22	-	-	X	X	X
7. få/vara/bli kalla handen ("kalla" "handen")	22	-	X	-	X	X
8. få stryka på foten ("stryka" "på" "foten")	40	X	X	X	X	X
9. gå i nngs fotspår ("i" [] {0,6} "fotspår")	78	-	-	-	-	-
10. gå stick i stäv med ngt ("stick" "i" "stäv")	34	X	X	-	X	X
11. gå upp i rök ("i" "rök")	50	X	X	X	X	X
12. gå ur tiden ("ur" "tiden")	27	X	X	-	X	-
13. ha blick för ngt ("blick" "för")	35	X	X	X	-	X
14. ha is i magen ("is" "i" "magen")	23	-	-	-	X	X
15. (ha ngt/finnas) i bakhuvudet ("i" "bakhuvudet")	39	-	-	-	-	X
16. ha ngt i sin hand ("i" "sin" "hand")	30	X	X	-	-	-
17. hamna på efterkälken ("efterkälken")	46	X	X	X	X	X
18. hålla huvudet kallt ("huvudet" "kallt")	25	-	X	-	X	X
19. hålla masken ("h.*" "masken")	36	-	X	-	-	X
20. hålla ngt/ngn i schack ("i" "schack")	59	X	X	X	X	X
21. komma i dagen ("kom.*" "i" "dagen")	31	X	X	-	X	X
22. ligga i luften ("i.*" "i" "luften")	31	X	X	-	X	X
23. ligga ngn i fatet ("i" "fatet")	32	X	X	X	X	X
24. ligga ngn varmt om hjärtat ("om" "hjärtat")	48	-	-	-	X	X
25. lägga band på sig ("band" "på")	23	X	X	-	X	X
26. lägga ngt på hyllan ("på" "hyllan")	46	X	X	-	X	-
27. ngt står för dörren ("st.*" "för" "dörren")	46	-	X	-	X	-
28. när allt kommer omkring ("kom.*" "omkring")	87	X	X	-	X	X
29. när det kommer till kritan ("till" "kritan")	42	X	X	X	X	X
30. e dagens ljus ("dagens" "ljus")	48	X	X	-	X	X
31. se mellan fingrarna med ngt ("mellanlgenom" "fingrama")	33	X	X	-	X	-
32. slå näven i bordet ("näv.*" "i" "bord.*")	23	X	X	-	X	-
33. sopa ngt under mattan ("under" "mattan")	38	-	-	-	X	X
34. stå på egna ben ("egna" "ben")	47	X	X	-	X	X
35. sätta käppar i hjulet för ngn ("käpp.*" "i" "hjul.*")	38	X	X	-	X	X
36. sätta ngt ur spel ("ur" "spel")	80	X	-	-	X	-
37. ta ngt med en nypa salt ("nypa" "salt")	35	X	X	-	X	X
38. ta saken i egna händer ("sak.*" "i" "eg.*")	26	X	-	-	X	X
39. toppen av/på ett isberg/isberget ("pålav" [] {0,1} "isberg.*")	29	-	-	-	X	X
40. vara med på noterna ("på" "noterna")	40	X	X	X	X	X

Hittills har diskussionen främst gällt huruvida de 40 idiomerna nämns i de undersökta ordböckerna eller inte. Men hur är det när de faktiskt nämns?

4. Förhållanden mellan former i Parole och ordböcker

Det råder delvis olika förhållanden mellan idiomformerna som förekommer i Parole och de former som tas upp i ordböckerna. Jag skiljer här på tre huvudtyper, nämligen a) stor överensstämmelsen, b) mindre överensstämmelse och c) ingen överensstämmelse mellan Parole och ordböckerna.

4.1. Stor överensstämmelse mellan Parole och ordböcker

Det råder hög grad av överensstämmelse mellan Parole-former och ordboksformer i ungefär hälften av de undersökta fallen. Viss variation förekommer inom båda materialen, men den är förhållandevis marginell. Ett tydligt exempel är *gå ur tiden*. I Parole finns det 27 belägg för uttrycket och i samtliga fall har skribenterna valt just formen *gå ur tiden*. Variationen är således obefintlig. Man ser också att angivelserna i de tre ordböcker där uttrycket nämns är fullständigt enhetliga; i samtliga fall anges formen *gå ur tiden*.

Andra exempel är *när allt kommer omkring* och *när det kommer till kritan*. I båda fallen finns det en klar preferens för presensformen i Parole, men uttrycken förekommer även i andra tempus. Imperfektformen är exempelvis mycket vanlig. I ordböckerna har man i nästan samtliga fall angivit presensformerna av uttrycken. Det står således *när allt kommer omkring* och *när det kommer till kritan*. Ordboksförfattarna har därmed frångått den traditionella infinitivformen, men detta är knappast förvånande då formerna *när allt komma omkring* och *när det komma till kritan* inte skulle upplevas som särskilt naturliga.

Ännu ett intressant exempel är *ha is i magen*. Formen *ha is i magen* är utan tvekan den mest frekventa i Parole, men också

med is i magen är vanligt förekommande (jfr även *ha blick för ngt* och *med blick för ngt*). I detta fall nämns formen *ha is i magen* i ordböckerna. Troligtvis betraktar ordboks författarna de två formerna som semantiskt överensstämmande och menar att det därför är tillräckligt att ange en av formerna. Jag är dock inte säker på att ordboks användarna har detta helt klart för sig.

4.2. Mindre överensstämmelse mellan Parole och ordböcker

Hur ser det ut då överensstämmelsen inte är fullt så stor? Här utkristalliserar sig fyra undergrupper.

4.2.1. Homogent (1) i Parole och homogent (2) i ordböckerna

Den första typen innebär att bruket i Parole är förhållandevis homogent. Viss variation förekommer, men denna är marginell. Detsamma gäller angivelserna i de olika ordböckerna. De allra flesta, om inte alla, anger samma form eller former. Vid en jämförelse framgår det dock att formerna inte är desamma. Detta är tydligt i exempelvis *lägga lök på laxen* som jag nämnde tidigare. Ett annat exempel är *se dagens ljus*. I Parole är bruket mycket konsekvent; i samtliga fall har formen *se dagens ljus* använts (och uttrycket har som regel betytt 'födas' eller 'bli till'). Formen *se dagens ljus* överensstämmer med en av de undersökta ordböckerna, nämligen *BSO*. Men man får lätt ett annat intryck av bruket om man konsulterar andra ordböcker. I *SHO* anges endast formen *skåda dagens ljus* och i *NEO* och *ISO* ges *se* och *skåda dagens ljus* som alternativa former. Var kommer då alternativformen *skåda dagens ljus* ifrån? Formen kan exempelvis ha varit mer frekvent förr i tiden. Det kan även röra sig om en etablerad form som råkar vara underrepresenterad i Parole. En genomgång av samtliga texter i Språkbanken visar emellertid att det endast finns två belägg av *skåda dagens ljus* och det är i ungdomsdramer av Strindberg. Det står: "Du föraktligaste Bland

dödliga, som ville stjäla bort Från mig min makes kärlek, vet, jag blott Behöver ge en vink och du ej mer Skall skåda dagens ljus” (ur *Det sjunkande Hellas* (1869) och *Hermione* (1870)). Här kan man således se att *skåda dagens ljus* snarast betyder ’leva’.

4.2.2. Homogent i Parole och heterogent i ordböckerna

Den andra typen innebär att användningen är förhållandevis homogen i Parole – variationen är marginell. Angivelserna i ordböckerna är däremot mer heterogena. Variationen kan bestå i att alternativa former anges i respektive ordbok, att uppgifterna i respektive ordbok skiljer sig åt eller en kombination av detta. Ett illustrativt exempel är användningen av *hålla huvudet kallt*. Bruket i Parole är mycket enhetligt; i samtliga fall har verbet *hålla* använts. I två av de tre ordböcker där uttrycket nämns anges dock alternativa former, nämligen *hålla huvudet kallt* och *bevara huvudet kallt*. Som jag tidigare nämnt har ett flertal forskare diskuterat det faktum att det ibland är befogat med alternativformer hos idiom i ordböckerna. I det här fallet är det tvärtom – det faktiska bruket är enhetligt men alternativformer ges.

4.2.3. Heterogent i Parole och homogent i ordböckerna

Den tredje typen innebär att bruket i Parole är heterogent. Det finns många alternativa former varav några är mer frekventa än andra. Däremot är angivelserna i de olika ordböckerna relativt samstämmiga; samma form eller former anges i de flesta av verken. Denna typ är förhållandevis vanlig. Ett tydligt exempel är det faktum att man exempelvis både kan *ge* och *få kalla handen*. I Parole finns det viss variation vad gäller verbet, men vanligast är *är/blir kalla handen*. Formerna *ge/få* är också vanligt förekommande och i ett enda fall används formen *vinka med kalla handen*. Trots detta är det den sista formen som nämns i de tre ordböcker där uttrycket förekommer. Det står således konsekvent *vinka med kalla handen*, vilket är något förvånande.

Ett annat tydligt exempel är *sätta käppar i hjulet för ngn*. Uttrycket har använts 38 gånger i Parole och beläggen är fördelade på fyra alternativa former. Dessa är *käppar i hjulet* (20 ex.), *käppar i hjulen* (11 ex.), *en käpp i hjulet* (6 ex.) och *käpp i hjulet* (1 ex.). Det råder således en klar variation. En granskning av de fyra ordböcker där uttrycket nämns ger dock ett annat intryck. I tre av dessa anges endast *sätta en käpp i hjulet*, d.v.s. den form som är minst frekvent i materialet. I den fjärde ordboken står det *sätta käpp i hjulet*, vilket i princip inte förekommer i Parole. Men hur ska ordboksförfattaren göra i ett fall som detta? Ska han endast ange den mest frekventa formen? En möjlig angivelse skulle vara *sätta en käpp/käppar i hjulet/hjulen (för ngn)*. Men detta medför ett problem, nämligen den totala avsaknaden av formen *sätta en käpp i hjulen (för ngn)*. Den finns för övrigt inte med i hela Språkbanken. En annan fråga som naturligen inffinner sig är om det är lämpligt med sådana angivelser. Risken är stor att de snarare förvirrar än förklarar.

4.2.4. Heterogent (1) i Parole och heterogent (2) i ordböckerna

Den fjärde och sista typen innebär att det finns en stor formvariation både i textkorpus och i ordböcker. Som exempel kan jag nämna *hamna på efterkälken*. I Parole finns det klara skillnader vad gäller bruket av verb, men i mer än hälften av fallen har *hamna* använts. *Komma* utgör cirka en fjärdedel av beläggen. Utöver dessa verb förekommer även *vara* och *lämnas*. I åtminstone tre av ordböckerna anges alternativa verb, som regel *komma*, *bli* och *vara*. I de två ordböcker där endast ett verb anges är detta *komma*. Däremot nämns inte *hamna* i något fall, vilket kan tyckas vara en smula märkligt. I detta fall är det således tydligt att de alternativa idiomformer som anges i ordböckerna kanske inte alltid är de mest frekventa.

4.3. Ingen överensstämmelse mellan Parole och ordböcker

Avslutningsvis skall nämnas att det även finns exempel där det inte råder någon form av överensstämmelse mellan Parole och ordböckerna. Detta är främst om ett idiom förekommer i skriftspråkskorpusen, men inte är upptaget i ordböckerna. Exemplet som är aktuellt här är *gå i ngns fotspår*. Detta förekommer så många gånger som 78 i Parole. Verbanvändningen är ytterst varierande, men vanligast är just *gå i ngns fotspår* (32 ex.) Men även *följa i ngns fotspår* är rikligt representerat. Uttrycket nämns dock inte i form av ett idiom i någon av de undersökta ordböckerna.

5. Sammanfattning och slutdiskussion

I artikeln har jag presenterat en undersökning av ett fyrtiotal svenska idiom. För det första har jag redogjort för i vilken utsträckning dessa upptas i fem samtida svenska ordböcker. För det andra har jag diskuterat i vilken mån det finns en överensstämmelse mellan de former som idiomerna har i ordböckerna med de former av idiomerna som är vanligast förekommande i textkorpusen Parole. Undersökningen visar bland annat att det råder god överensstämmelse mellan former i Parole och ordböcker i ungefär hälften av fallen. I resten av exemplen är skillnaderna däremot mer påtagliga. Exempelvis finns det många gånger en klar variation i Parole utan att någon sådan variation avspeglas i ordböckerna. Men skillnaden kan även bero på det rakt motsatta förhållandet. I vissa fall är användningen i Parole förhållandevis homogen, men trots detta anges det alternativa former i ordböckerna. Det finns också fall där det råder en klar samstämmighet beträffande formen eller formerna i ordböckerna men det är knappast denna form/former som används i de faktiska texterna.

I den här studien har jag således koncentrerat mig på idiomens form, men det vore givetvis intressant att också granska idiomens betydelsesida. För i vilken mån överensstämmer betydelsean-

givelserna i ordböcker med den betydelse som man kan se hos uttrycken i den faktiska användningen?

Slutligen vill jag även besvara den fråga som ställs i titeln. Jag anser inte att man kan säga att idiomerna i svenska ordböcker lever sitt eget liv. Men skillnaderna mellan i det här fallet Parole och ordboksformerna är ändå så pass tydliga att de är värda att uppmärksamma. Med artikeln vill jag främst – och det tål att betonas – inleda en diskussion kring frågor som: Vad är ett idiom? Vilka idiom ska man ha med i en ordbok? Och kan inte textkorpusar utnyttjas i större utsträckning än vad som sker idag?¹ Något jag även undrar över är vilka principer man har använt för att ange ett idiom i ordböckerna och hur konsekvent dessa har följts. I artikeln har jag också till viss del berört hur idiomerna i ordböckerna kan tolkas av de faktiska användarna. Vilka former ingår i vad? Och hur tolkar de faktiska ordboksanvändarna angivelserna? Även denna aspekt är mycket relevant och utan tvekan värd en egen studie.

Litteratur

A. ORDBÖCKER

BSO = *Bonniers svenska ordbok*. 1998. Red. Sten Malmström et al. 7 uppl. Stockholm: Bonniers.

ISO = *Illustrerad svensk ordbok*. 1964. Red. Bertil Molde. 3 rev. uppl. Stockholm: Natur och Kultur.

NEO = *Nationalencyklopedins ordbok*. 1995–96. Utarbetad vid Språkdata, Göteborgs universitet. Höganäs: Bra böcker.

SHO = *Svensk handordbok*. 1966. Red. Ture Johannisson och K.G. Ljunggren. Stockholm: Norstedts.

SOB-en = *Svenska ordboken*. 1995. Red. Bengt Oreström. Lund: Gustava.

¹ För mer ingående diskussioner kring användning av textkorpusar vid ordboksarbete, se bl.a. Gellerstam 1996 och McEnery & Wilson 1996.

B. ANNAN LITTERATUR

- Alexander, Richard J. 1992. "Fixed expressions, idioms and phraseology in recent English learner's dictionaries." Tommola, Hannu et al. (red.): *Euralex '92 proceedings I. Papers submitted to the 5th EURALEX International Congress on Lexicography in Tampere, Finland*. S. 35–42.
- Anward, Jan & Linell, Per. 1976. "Om lexikaliserade fraser i svenskan." *Nysvenska studier* 55/56:77–119.
- Benson, Morton. 1985. "Collocations and idioms." Ilson, Robert (red.): *Dictionaries, lexicography and language learning*. Oxford: Pergamon Press. S. 61–68.
- Botha, Willem. 1992. "The lemmatization of expressions in descriptive dictionaries." I: Tommola, Hannu et al. (red.): *Euralex '92 proceedings II. Papers submitted to the 5th EURALEX International Congress on Lexicography in Tampere, Finland*. S. 465–472.
- Cacciari, Cristina. 1993. "The place of idioms in a literal and metaphorical world." Cacciari, Cristina & Tabossi, Patrizia (red.): *Idioms: Processing, structure, and interpretation*. Hillsdale: Lawrence Erlbaum. S. 27–52.
- Clausén, Ulla. 1996. "Idiom i bruk." *Språket lever! Festskrift till Margareta Westman den 27 mars 1996*. Utgiven av Svenska språknämnden. S. 36–42.
- Corpas Pastor, Gloria. 1996. *Manual de la fraseología española*. Madrid: Gredos.
- Everaert, Martin, van der Linden, Erik-Jan, Schenk, André & Schreuder, Rob (red.). 1995. "Introduction." *Idioms: Structural and psychological perspectives*. Hillsdale: Lawrence Erlbaum. S. 1–13.
- Fleischer, Wolfgang. 1997. *Phraseologie der deutschen Gegenwartssprache*. 2., durchgesehene und ergänzte Auflage. Tübingen: Max Niemeyer.
- Gellerstam, Martin. 1996. "Rätt och fel i korpusen." *LexicoNordica* 3:35–47.
- Kjellmer, Göran. 1996. "Idiomen, kollokationerna och lexikonet." *LexicoNordica* 3:79–90.

- Krohn, Karin. 1994. *Hand und Fuß. Eine kontrastive Analyse von Phraseologismen im Deutschen und Schwedischen.* (Göteborgger germanistische Forschungen. 36.) Göteborg.
- Lorentzen, Henrik. 1996. "Lemmatization of Multi-word Lexical Units: in which Entry?" Gellerstam, Martin et al. (red.): *Euralex '96 Proceedings I. Papers submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden.* S. 415–421.
- McEnery, Tony & Wilson, Andrew. 1996. *Corpus linguistics.* Edinburgh: Edinburgh University Press.
- Moon, Rosamund. 1996. "Data, description, and idioms in corpus lexicography." Gellerstam, Martin et al. (red.): *Euralex '96 Proceedings I. Papers submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden.* S. 245–256.
- Nunberg, Geoffrey, Sag, Ivan A. & Wasow, Thomas. 1994. "Idioms." *Language. Journal of the Linguistic Society of America* 70:491–538.
- Sköldberg, Emma. 1999. "Varianter av idiom." *Förhandlingar vid Tjugotredje sammankomsten för svenskans beskrivning, Göteborg 15–16 maj 1998.* Lund: Lund University Press. S. 384–392.
- Wood, Mary McGee. 1986. *A definition of idiom.* Bloomington, Indiana: Indiana University Linguistic Club.

Bo Svensén

Markeringsangivelser i ordböcker: några strukturella aspekter

This contribution deals with some structural aspects of diasystematic marking in dictionaries. The functions and structural properties of the markers themselves are discussed as well as their relationship to other textual elements, applying the conceptual and terminological framework used in *Nordisk leksikografisk ordbok* [Nordic Dictionary of Lexicography]. In the concluding section, some problems pertaining to the user aspect are briefly touched upon.

1. Inledning

1.1. Diasystematisk markering

Markering, eller rättare sagt *diasystematisk markering*, innebär i regel att det i en ordbok anges att en viss lexikalisk enhet i något avseende avviker från huvudmassan av de enheter som förekommer i ordboken och att dess användning är underkastad restriktioner av något slag. Markeringsystemet i en ordbok består av ett antal delsystem, som vart och ett tar sikte på en viss typ av egenskaper hos de lexikaliska enheterna, t.ex. *diakronisk markering* (som gäller den tidsmässiga aspekten), *diatopisk markering* (som gäller den rumsliga aspekten), *diastratisk markering* (som gäller den stilmässiga aspekten) och *diateknisk markering* (som gäller tillhörighet till ett visst fackområde). Genom att välja ett visst slags *markeringsangivelse* specificerar ordboken närmare vari avvikelsen eller restriktionen består:

Markeringstyp	Dimension	Exempel på markeringsangivelse
diakronisk	tid	åld., hist.
diatopisk	plats	dial., finlandssv.
diastratisk	stilmnivå	litt., vard., sl., vulg.
diateknisk	fackområde	biol., jur., fil., mus.

Markering är ett intressant område från strukturell synpunkt, och syftet med mitt bidrag är att försöka beskriva såväl markeringsangivelsernas egen struktur som deras roll i ordboksstrukturen i stort med hjälp av den begreppsapparat och den terminologi som förekommer i *Nordisk leksikografisk ordbok* (NLO 1997). Men eftersom NLO inte är fullt tillräcklig för det här ändamålet, skall jag komplettera med ytterligare några begrepp och termer, de flesta hämtade från det stora samlingsverket *Wörterbücher, Dictionaries, Dictionnaires* (1989–91). Jag skall också ta upp några problem som har att göra med användaraspekten.

1.2. Funktion

1.2.1. Uttryckssidan eller innehållssidan?

Markeringsangivelserna kan avse restriktioner på uttryckssidan eller på innehållssidan. Vissa av dem kan avse båda typerna (Svensén 1987:174ff.).

Inom det diatekniska delsystemet kan en angivelse som ”med.” avse innehållssidan och betyda ’detta uttryck avser en företeelse som hör hemma inom medicinen’ (som i det första exemplet nedan), men den kan också avse uttryckssidan och betyda ’detta uttryck tillhör medicinskt fackspråk’ (som i det andra exemplet):

cataplasma [...] s med. grötomslag

(Norstedts eng–sv 1993)

pancreatic [...] *adj.*, ~ *juice* bukspott; med. pankreassaft

(Norstedts eng–sv 1993)

På liknande sätt måste man inom det diakroniska delsystemet skilja mellan ”hist.” och ”åld.”, där den förra angivelsen gäller innehållssidan och betyder ’detta uttryck betecknar en företeelse som inte längre förekommer i sinnevärlden’, medan den senare tar sikte på uttryckssidan och betyder ’detta uttryck tillhör ålderdomligt språk’:

varlet [...] s **1** hist. page, väpnare **2** äld. el. skämts. lymmel, skurk
(Norstedts eng–sv 1993)

1.2.2. Information eller differentiering?

En markeringsangivelse kan ha två olika uppgifter. Antingen skall den bara ge information om en viss lexikalisk enhet, eller också är dess uppgift att genom att ge denna information differentiera mellan flera enheter av samma slag (Werner 1991: 2796ff.). I följande exempel kan det knappast vara fråga om någon differentiering, eftersom alla angivelser i artikeln är av olika slag:

corrélat [...] m gram. korrelat

(Norstedts fra–sv 1998)

Differentiering mellan lexikaliska enheter med hjälp av markeringsangivelser är vanligast i tvåspråkiga ordböcker, där det som regel handlar om *ekvivalentdifferentiering*:

membran s anat. el. biol. membrane; tele. el. radio o.d. el. i pump
diaphragm

(Norstedts sv–eng 1993)

Det är också vanligt att markeringsangivelser används vid differentiering mellan homografer:

1 basilic [...] m bot. el. kok. basilika

2 basilic [...] m mytol. basilisk

(Norstedts fra–sv 1998)

Exemplen är typiska också på så sätt att det för det mesta är just diatekniska markeringsangivelser som används i differentierande syfte.

2. Angivelsestruktur

Markeringsangivelsernas inre struktur kan vara utformad på flera olika sätt. Man kan exempelvis skilja mellan angivelser som har

undergått någon form av *kondensering* eller *textförtätning* (jfr Wolski 1989) och sådana som inte har gjort det. Kondensering innebär att ett textelement förkortas med hjälp av ett visst förfarande. Det ursprungliga textelementet kallas i NLO *fulltext*, och resultatet av textförtätningen benämns *kondensat*.

Det är få ordböcker som genomgående har icke-kondenserade markeringsangivelser, helt naturligt eftersom detta är mycket utrymmeskrävande. En av dem som har gått in för fulltextangivelser är COBUILD:

calorie [...] 2 the amount of heat needed to raise the temperature of one gram of water by 1C; a technical term in physics.
(COBUILD 1987)

bullshit [...] a very rude or swear word. 1 If you say 'bullshit!', you are telling someone that you think what they have just said is complete nonsense or absolutely untrue. [...]
(COBUILD 1987)

Det normala är i stället att alla eller nästan alla typer av markeringsangivelser i en ordbok är kondenserade på något sätt. Kondensering kan ta sig olika former, och när det gäller markeringsangivelser förekommer åtminstone fem olika typer. I två av fallen utgörs angivelsen av en s.k. *markör*. Det vanligaste är att markören består av en standardiserad förkortning:

sleeper [...] 2 järnv. sovvnagn; sovplats 3 järnv. sliper, syll 4 vard. plötslig [och] oväntad succé [...] 6 vard. sömntablett, sömnpiller
(Norstedts eng-sv 1993)

Markören kan också bestå av en *angivelsesymbol*, dvs. en symbol som själv utgör en angivelse. Den typen av markörer var vanligare förr. Vi vet ju exempelvis hur ämnesområden ofta brukade anges i äldre ordböcker: ankaren för sjöväsen, kugghjul för teknik, korslagda värjor för militärväsen osv.:

W (Buch)	wissenschaftlich.
♀ (Blume)	Pflanzenkunde, Pflanze.
⊕ (Zahnrad)	technisch, Handwerks-Ausdruck.
⚔ (Hämmer)	den Bergbau betreffend.
✖ (Degen)	militärisch.
⚓ (Anker)	{Marine, Schifffahrt; {Schiffersprache.
⊕ (Welschädel)	Handel, Kaufmännisch.
✉ (Posthorn)	Post.
🚂 (Locomotive)	Eisenbahn.
🚲 (Fahrrad)	Radfahr-sport.
🎵 (Note)	Musik.

(Sachs-Villatte 1911)

Bruket av standardiserade förkortningar och angivelsesymboler kombineras i regel med en tredje typ av kondensering, nämligen utelämnning av textelement. Förkortningen "fam." i följande exempel betyder inte bara 'familjär' utan skall uttydas 'uttrycks-sättet tillhör familjärt språk'.

trang II -t, adj. [...] 3 (fam.) dum; tung i begripelsen [...] (Aschehoug og Gyldendal 1992)

3. Åtkomststruktur

Den fjärde typen av kondensering har inte med angivelsens egen struktur att göra utan gäller *åtkomststrukturen*. Det handlar alltså om de strukturella egenskaper hos ordbokstexten som leder användaren till sökt information, dvs. i vårt fall visar att ett visst textelement är en markeringsangivelse.

Åtkomststrukturen åskådliggörs genom *strukturindikatorer*. Man skiljer mellan *typografiska strukturindikatorer* och *icke-typografiska strukturindikatorer*.

De typografiska strukturindikatorerna består i att vissa teckensnitt (t.ex. Times, Helvetica), vissa snittvarianter (t.ex. kursivstil,

fetstil, kapitäl), vissa teckengrader etc. används för att ange olika funktioner hos textelement. Till de vanligaste typografiska strukturindikatorerna i samband med markeringsangivelser hör mindre teckengrad och kursiv snittvariant, den senare särskilt populär i anglosaxisk lexikografi:

sleep^r [...] 2 järnv. sovvagn; sovplats 3 järnv. sliper, syll 4 vard. plötslig [och] oväntad succé [...] 6 vard. sömntablett, sömnpiller
(Norstedts eng-sv 1993)

bugger^r [...] *n sl, esp. BrE* 1 taboo an offensive or disagreeable person 2 taboo a SODOMITE [...] (LDOCE 1987)

De icke-typografiska strukturindikatorerna består av tecken, i regel skiljetecken, parentestecken och symboler. De vanligaste typerna som förekommer i samband med markeringsangivelser är parenteser av olika slag:

trang II -t, adj. [...] 3 (fam.) dum; tung i begripelsen [...] (Aschehoug og Gyldendal 1992)

Glas^r [...] er hat ein bißchen zu tief ins ~ geguckt <umg.> er ist leicht beschwipst [...] (Wahrig 1980)

4. Räckvidd

En annan viktig faktor när det gäller markeringsangivelser är deras *räckvidd*. Med en angivelses räckvidd menar jag här summan av de textelement i ordboksartikeln utöver lemmat som den på rent strukturella grunder kan tänkas avse.

I en markeringsangivelses "närområde" finns det vissa faktorer som begränsar räckvidden, i huvudsak icke-typografiska strukturindikatorer såsom parenteser, skiljetecken och momentsymboler. I de följande exemplen har hakparenteser insatts för att visa vilka delar av resp. artikel som till följd av strukturindikatorer

ernas placering har kommit att ligga utanför markeringsangivel-
sens räckvidd:

fotogenlampa *s* paraffin] (amer. kerosene) [lamp
(Norstedts sv-eng 1993)

tensile /'tensəl] ; US 'tensl/ *adj* [...]
(OALDCE 1989)

pojke *s* [petit] garçon *m*] , vard. gosse *m* [...]
(Norstedts sv-fra 1998)

came² [...] *nf* [...] snow (arg), [junk (arg), [[...]
(Robert-Collins 1978)

kat [...]] 5 sjö. katt, piska; *den nihalede* ~ den niosvansade
katten
(Molde 1980)

Men vi skall inte uppehålla oss vid detta utan i stället se efter hur stor en markeringsangivelses räckvidd maximalt kan bli. Vi kan exempelvis tänka oss att vi har ett antal lexikaliska enheter som tillhör samma lemma och som är markerade på samma sätt. På ett visst stadium i redigeringsprocessen kan de ha sett ut som i följande exempel. Varje markeringsangivelse har en räckvidd som sträcker sig till slutet av resp. textstycke:

spat [...] *s* amer. [små]gräl, gruff
spat [...] *s* amer. dask, smäll
spat [...] *s* amer. stänk
spat [...] *vb itr* amer. smågräla, gruffa
spat [...] *vb itr* amer. stänka
spat [...] *vb tr* amer. daska till, smälla till

På ett något senare stadium i redigeringsprocessen kan de sex enheterna ha sammanförts i grupper efter ordklassstillhörighet, och det hela skulle kunna se ut så här:

spat [...]

I s I amer. [små]gräl, gruff 2 amer. dask, smäll 3 amer. stänk

II vb itr I amer. smågräla, gruffa 2 amer. stänka

III vb tr amer. daska till, smälla till

Men eftersom alla delbetydelser inom vart och ett av ordklassmomenten visar sig ha samma markeringsangivelse, flyttar lexikografen ut angivelsen till närmast högre nivå, dvs. till ordklassnivån, och stryker alla markeringar på delbetydelsenivå. Varje markeringsangivelse har nu inom sin räckvidd samtliga betydelsemoment inom resp. ordklassmoment:

spat [...]

I s amer. 1 [små]gräl, gruff 2 dask, smäll 3 stänk

II vb itr amer. 1 smågräla, gruffa 2 stänka

III vb tr amer. daska till, smälla till

Men eftersom nu alla ordklassmomenten också visar sig ha samma markeringsangivelse, flyttar lexikografen ut markeringen ytterligare en gång till närmast högre nivå, dvs. till lemmanivån, och stryker alla markeringar på ordklassnivå. Den enda kvarstående markeringsangivelsen har nu inom sin räckvidd samtliga ordklassmoment, dvs. hela artikeln:

spat [...] amer. I s 1 [små]gräl, gruff 2 dask, smäll 3 stänk II vb itr 1 smågräla, gruffa 2 stänka III vb tr daska till, smälla till

Det som här har beskrivits är i själva verket en femte typ av kondensering, som på tyska kallas *Auslagerung* (Wiegand 1989: 447ff.), på svenska förslagsvis *utflyttning*. Utflyttning är en kondenseringsform som är mycket typisk för markeringsangivelser. Det just visade exemplet representerar den vanligaste typen av utflyttning, s.k. *vänsterutflyttning*, där markeringsangivelsen flyttas till momentets eller artikelns början, varigenom dess räckvidd kommer att innefatta alla lexikaliska enheter som följer därefter, ända fram till momentets eller artikelns slut.

Vänsterutflyttning av markeringsangivelser kan också ske i betydligt mindre skala. Det följande exemplet måste rimligen uppfattas så att den diatopiska angivelse "isht amer." som man

måste förutsätta även vid "penitentiary" har flyttats ut till den första av de två amerikanska ekvivalenterna och därigenom har båda inom sin räckvidd:

fängelse *s* prison, gaol; isht amer. jail, penitentiary [...] (Norstedts sv-eng 1993)

Det förekommer också *högerutflyttning*, som är en typisk egen-skap för etymologiangivelser. De flesta ordböcker som ger etymologier placerar dem ju sist i artiklarna, och när detta sker i en artikel med ett polysemt lemma blir det alltså fråga om höger-utflyttning. När det gäller markeringsangivelser blir det bara undantagsvis fråga om högerutflyttning och då i mycket liten skala. Följande exempel skulle möjligen kunna anses vara ett fall av högerutflyttning, där den diatekniska angivelse "äv. data." som man måste förutsätta även vid "flowchart" har flyttats ut till den senare av de två ekvivalenterna och därigenom har båda inom sin räckvidd:

flödesschema *s* flowchart, flowsheet båda äv. data. (Norstedts sv-eng 1993)

5. Adresseringsstruktur

Som i de flesta fall när det gäller ordboksstruktur kan det vara fruktbart att betrakta markeringsangivelserna som komponenter i s.k. *bearbetningsenheter*. En bearbetningsenhet består i regel av två komponenter, varav den ena utgör en utsaga om den andra. Deras inbördes relation kan betraktas som en relation mellan ett *tema* och ett *rema*, där utsagan är rema och det som utsagan gäller är tema. Vi kan betrakta följande enkla exempel ur en tänkt engelsk-svensk ordbok:

car bil

I exemplet är "car" tema och "bil" rema. Det som i en sådan relation utgör rema kallas i ordboken *angivelse*, i detta fall alltså

betydelseangivelsen "bil", och det som utgör tema kallas angivelsens *adress*. Man säger att det föreligger en *adressering* från "bil" till "car":

car ← bil
TEMA ← REMA
ADRESS ← ANGIVELSE

Ordningsföljden och relationerna mellan adresser och angivelser utgör ordbokens *adresseringsstruktur*.

Den vanligaste adressen för en angivelse är lemmat, och man talar då om *lemmatisk adressering*:

firre s vard. fish

(Norstedts sv-eng 1993)

Här kan var och en av de tre angivelserna betraktas som adresserad till lemmat, och exemplet innehåller alltså följande tre bearbetningsenheter:

firre ← s
firre ← vard.
firre ← fish

Men det är också mycket vanligt med adressering inom mikrostrukturens ram, s.k. *infralemmatisk adressering*, där angivelsen är adresserad till ett annat textelement än lemmat. I själva verket är markeringsangivelserna ett ganska typiskt exempel på en informationstyp som ofta är infralemmatiskt adresserad. En markeringsangivelse kan exempelvis avse en enstaka uttalsvariant eller böjningsvariant, en målspråklig ekvivalent, ett språkprov osv.:

tensile /'tensəl/; US 'tensl/ adj [...]

(OALDCE 1989)

beget [...] (imperf.: *begot*, äld. el. bibl. *begat* ; [...])

(Norstedts eng-sv 1993)

fängelse s prison, gaol; isht amer. jail [...]

(Norstedts sv-eng 1993)

Glas¹ [...] er hat ein bißchen zu tief ins ~ geguckt <umg.> er ist
 leicht beschwipst [...] ↑ ↓
 (Wahrig 1980)

Angivelsen kan i ordbokstexten stå före eller efter adressen. Står angivelsen efter adressen sägs den vara *vänsteradresserad*. Detta är naturligtvis det normala när adressen är lemmat, men andra fall förekommer också. Ett typiskt sådant fall är när adressen är ett språkprov:

fièvre [...] ; ~ *de cheval* vard. hög feber [...] ↑
 (Norstedts fra-sv 1998)

Men även det motsatta fallet är vanligt. En artikelintern angivelse som är placerad före lemmat är naturligtvis alltid *högeradresserad*. Här betyder asterisken att ordet tillhör amerikansk engelska:

*pard², *par¹dner, *ns.* (sl.) partner, comrade [...] ↑ ↑
 (COD 1982)

Vi har nyss sett åtskilliga exempel också på infralemmatisk högeradressering:

tensile /'tensail; US 'tensl/ *adj* [...] ↑
 (OALDCE 1989)

beget [...] (imperf.: *begot*, ätd. et. bibl. *begat* ; [...] ↑
 (Norstedts eng-sv 1993)

fängelse *s* prison, gaol; isht amer. jail [...] ↑
 (Norstedts sv-eng 1993)

Man kan också skilja mellan *enspråkiga bearbetningsenheter* och *tvåspråkiga bearbetningsenheter*.

I enspråkiga ordböcker, där det ju bara finns ett objektspråk och detta är identiskt med beskrivningsspråket, finns förstås bara enspråkiga bearbetningsenheter.

I de tvåspråkiga ordböckerna finns två objektspråk, varav det ena också fungerar som beskrivningsspråk. Därför finns det i sådana ordböcker både enspråkiga och tvåspråkiga bearbet-

ningsenheter. Markeringsangivelserna, i den mån de uttrycks i ord, hör till ordbokens beskrivningsspråk, och detta är i regel identiskt med de tilltänkta användarnas modersmål.

De tvåspråkiga bearbetningsenheter där markeringsangivelser ingår blir därför en typisk företeelse i passiva ordböcker, eftersom det i regel är det främmande språket som användaren behöver upplysningar om:

fièvre [...]; ~ *de cheval* vard. hög feber [...]

(Norstedts fra–sv 1998)

Bland de enspråkiga bearbetningsenheterna i tvåspråkiga ordböcker förekommer både *källspråkliga bearbetningsenheter* och *målspåkliga bearbetningsenheter*. I det första av följande två exempel ingår markeringsangivelsen i en källspråklig bearbetningsenhet, i det andra i en målspåklig:

firre *s* vard. fish

(Norstedts sv–eng 1993)

pancreatic [...] *adj.*, ~ *juice* bukspott; med. pankreassaft

(Norstedts eng–sv 1993)

De här två exemplen är typiska när det gäller just markeringsangivelser. Enspråkiga bearbetningsenheter på källspråket förekommer i aktiva ordböcker, där källspråk och beskrivningsspråk är identiska, och enspråkiga bearbetningsenheter på målspåket uppträder i passiva ordböcker, där målspåk och beskrivningsspråk sammanfaller.

Adresseringsförhållandena kan också beskrivas på olika sätt beroende på om markeringsangivelsen enbart skall ge information om en lexikalisk enhet eller om den också har en differentierande funktion. I följande två exempel handlar det enbart om att ge information:

firre [...] vard. fish

(Norstedts sv–eng 1993)

kosing [...] vard. fric *m* [...]

(Norstedts sv–fra 1998)

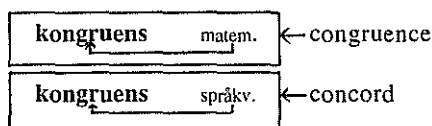
I exemplet ”**firre**” förekommer två angivelser, som var och en är adresserad direkt till lemmat. I exemplet ”**kosing**” skulle man kunna tänka sig att markeringsangivelsen är adresserad även till ekvivalenten.

I följande exempel har markeringsangivelsen däremot en differentierande funktion:

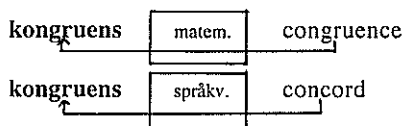
kongruens [...] matem. congruence; språkv. concord
(Norstedts sv-eng 1993)

Här måste adresseringen beskrivas på annat sätt, och flera olika tolkningar har föreslagits (Hausmann & Werner 1991:2732ff.), bland annat följande:

a) Det är fråga om två angivelser, ”congruence” och ”concord”, vilka är adresserade till varsin tvåledad adress: ”congruence” till ”**kongruens** matem.” och ”concord” till ”**kongruens** språkv.”. Var och en av adresserna utgör dessutom själv en bearbetningsenhet, där markeringsangivelsen är adresserad till lemmat. Det innebär att den matematiska termen ”**kongruens**” har ekvivalenten ”congruence”, och den språkvetenskapliga termen ”**kongruens**” har ekvivalenten ”concord”.



b) Det är fråga om två angivelser, ”congruence” och ”concord”, vilka båda är adresserade till lemmat ”**kongruens**”. Textelementen ”matem.” och ”språkv.”, som står mellan adress och angivelse, är själva varken adresser eller angivelser utan kan betraktas som ett slags ”angivelsemodifierare”. Innebörden är att termen ”**kongruens**”, när den används i matematiska sammanhang, skall översättas med ”congruence”, och att den när den används i språkvetenskapliga sammanhang skall översättas med ”concord”.



6. Markeringsystemet och användaren

Markeringsangivelserna kan vara problematiska från användarsynpunkt, eftersom det ibland kan vara svårt att avgöra till vilka andra textelement en viss angivelse kan vara adresserad. Dessa problem är av olika slag och av olika svårighetsgrad beroende dels på ordbokstypen, dels på typen av markering.

I enspråkiga ordböcker är läget förhållandevis enkelt. Där finns bara ett objektspråk, och man kan därför räkna med att markeringsangivelserna i enspråkiga ordböcker enbart avser ingångsspråkliga enheter, dvs. i stort sett bara lemmat och språkprov av olika slag. Det betyder att det knappast behöver råda några tvivel om en markeringsangivelses adressering, även om den är placerad på det här sättet, långt ifrån det som i första hand måste betraktas som dess adress, nämligen lemmat:

mju'kis subst. ~en ~ar

↑ ORDLED: mjuk-is-en

• (ofta i sms.) person som är mjuk dvs. (alltför) vänlig, snäll, förstående, mänsklig etc.; särsk. om man {MOTS. hårding, tuffing}

<vard>: mjukispappa [...]

(NEO 1995–96)

I tvåspråkiga ordböcker är läget ett annat. Där finns två objektspråk, och användaren kan därför bli osäker om en markeringsangivelse gäller enbart källspråkssidan, enbart målspråkssidan eller kanske både källspråkssidan och målspråkssidan.

De tolkningssvårigheter som uppstår i tvåspråkiga ordböcker gäller dock inte i lika hög grad för alla typer av markeringsangivelser. Rent allmänt borde man exempelvis kunna utgå från att diakroniska angivelser (t.ex. "åld.") och diatopiska angivelser (t.ex. "dial.", "prov.") som regel avser källspråkliga enheter och framför allt förekommer i passiva ordböcker: man brukar ju inte ange ålderdomliga eller dialektala ekvivalenter till ommarkerade

källspråkliga uttryck, inte heller brukar ålderdomliga eller dialektala uttryck förekomma på källspråkssidan i aktiva ordböcker.

Men det finns en typ av markeringsangivelser som inte erbjuder sådana generella hållpunkter, och det är de diastratiska (jfr Werner 1980). Det är nämligen ganska vanligt att ordböckerna även anger vardagliga ekvivalenter till omarkerade källspråkliga enheter (som i det första av de följande exemplen) och att vardagliga uttryck förekommer på källspråkssidan även i aktiva ordböcker (som i de tre andra exemplen):

pojke *s* [petit] garçon *m*, vard. gosse *m* [...]

(Norstedts sv-fra 1998)

kosing *s* vard. fric *m* [...]

(Norstedts sv-fra 1998)

firre *s* vard. fish

(Norstedts sv-eng 1993)

gängla [...] III *vb rfl*, ~ *sig* vard., gifta sig se mettre la corde au cou, se marier

(Norstedts sv-fra 1998)

I det första exemplet vet den svenske användaren redan från början att det källspråkliga uttrycket är omarkerat och kan sedan på strukturella grunder dra slutsatsen att angivelsen "vard." är adresserad enbart till den målspråkliga enheten "gosse". I de tre andra exemplen vet den svenske användaren redan att det källspråkliga uttrycket är vardagligt, men hur skall han rent principiellt kunna avgöra om angivelsen är adresserad enbart dit eller om den också avser ekvivalenterna?

Normalt förväntas väl användaren anta att stilnivån hos den målspråkliga ekvivalenten skall vara densamma som hos det källspråkliga uttrycket om inget annat anges, i varje fall i en aktiv ordbok. Det antagandet verkar ju vara riktigt när det gäller exemplet "**kosing**", men i de två sista exemplen riskerar användaren att bli vilseförd: i det tredje tillhör ekvivalenten inte samma stilnivå som lemmat, och i det fjärde tillhör visserligen

den första ekvivalenten samma stilnivå som det källspråkliga uttrycket men inte den andra.

Ett sätt att förebygga osäkerhet på den här punkten är förstås att förse både det källspråkliga uttrycket och de målspråkliga ekvivalenterna med markeringsangivelser. Detta sker normalt inte, men det finns en typ av ordböcker där det i regel blir nödvändigt att göra det. Jag syftar då på de *bidirektionella ordböckerna*, dvs. tvåspråkiga ordböcker avsedda för både användare som har källspråket som modersmål och användare som har målspråket som modersmål. En sådan ordbok är Robert-Collins' fransk-engelska och engelsk-franska ordbok (Robert-Collins 1978):

came² [...] *nf* (arg *Drogue*) snow (arg), junk (arg), stuff (arg) [...] (Robert-Collins 1978)

rappeler [...]; (*frm*) rappelez-moi à son bon souvenir
kindly remember me to him (*frm*). [...]

(Robert-Collins 1978)

Att ha markeringsangivelser på både källspråkssidan och målspråkssidan är utrymmeskrävande, och detta är sannolikt förklaringen till att man i Robert-Collins har valt att använda angivelsesymboler för de vanligaste stilsikterna (en asterisk för tämligen ledigt vardagsspråk, två för ännu ledigare språk osv.). Angivelsesymboler av detta slag har också den fördelen att de rent typografiskt kan placeras dikt an mot adressen så att man inte behöver tveka om adresseringen:

glamour [...] 2 *cpd*: glamour boy* beau gars*,
beau mec**, glamour girl* pin-up** *f inv*, beauté *f*.
(Robert-Collins 1978)

De två objektspråken i Robert-Collins liknar varandra ganska mycket när det gäller de termer som brukar användas i markeringsangivelser, i synnerhet de diatekniska. I ordbokens markeringsangivelser, i den mån de består av standardiserade förkortningar, har man också företagit en ytterligare sammanjämkning av de två språken, och resultatet har blivit att markeringsystemet

i Robert-Collins kan sägas vara mer eller mindre enspråkigt. (Man kan exempelvis lägga märke till hur ordboken i exemplet ”**rappeler**” ovan använder förkortningen ”*frm*”, som står för både fra. *formel* och eng. *formal*.)

Om objektspråken skiljer sig mer från varandra måste man ta till andra medel. Ett sådant är att använda sig av ett tvåspråkigt markeringsystem. En bidirektionell ordbok som har infört ett sådant system är Tom Hustads *Stor norsk-tysk ordbok*:

skjel'm² -en (1) (veralt.) Betrüger – m, Schlingel – m, (foreld.)
Schelm m [...]

(Hustad 1979)

De tyskspråkiga markeringsangivelserna skall ge de tyska användarna information om källspråkliga enheter och är vänsteradresserade. De norskspråkiga markeringsangivelserna skall ge de norska användarna information om de målspråkliga enheterna och är högeradresserade. Till yttermera visso särskiljs de två typerna av markeringsangivelser med hjälp av typografiska strukturindikatorer: de tyskspråkiga kännetecknas av rak snittvariant och de norskspråkiga av kursiv.

7. Slutord

Allt som här har sagts avser givetvis enbart pappersordböcker och har ingenting med ”the floating dictionary” (Sinclair 2001) att göra. I takt med att formerna för digital lagring och distribution av lexikalisk information utvecklas och den elektroniska ordboken i allt större utsträckning blir var mans egendom, kommer de problem jag har berört sannolikt att förlora allt mer i aktualitet och slutligen övergå till att enbart utgöra en del av lexikografins historia.

Litteratur

A. ORDBÖCKER

- Aschehoug og Gyldendal 1992 = *Aschehoug og Gyldendals Store norske ordbok*. Red. T. Guttu. 3. opplag. Oslo: Kunnskapsforlaget.
- COBUILD 1987 = *Collins COBUILD English Language Dictionary*. Ed. in chief J. Sinclair. London & Glasgow: Collins.
- COD 1982 = *The Concise Oxford Dictionary of Current English*. 7th ed. Ed. J. B. Sykes. Oxford: Clarendon Press.
- Hustad 1979 = T. Hustad: *Stor norsk-tysk ordbok*. Oslo – Bergen – Tromsø: Universitetsforlaget.
- LDOCE 1987 = *Longman Dictionary of Contemporary English*. New Edition. Harlow, Essex: Longman.
- Molde 1980 = *Dansk-svensk ordbok*. Utarbetad [...] av B. Molde. 3:e upplagan. Stockholm: Esselte Studium.
- NEO 1995–96 = *Nationalencyklopedins ordbok 1–3*. Höganäs: Bokförlaget Bra Böcker.
- NLO 1997 = H. Bergenholtz/I. Cantell/R. V. Fjeld/D. Gunder- sen/J. H. Jónsson/B. Svensén: *Nordisk leksikografisk ordbok*. (Skrifter utgitt av Nordisk forening for leksikografi. Skrift nr. 4.) Oslo: Universitetsforlaget.
- Norstedts eng–sv 1993 = *Norstedts stora engelsk-svenska ordbok*. 2:a upplagan. Stockholm: Norstedts.
- Norstedts fra–sv 1998 = *Norstedts stora fransk-svenska ordbok*. Stockholm: Norstedts Ordbok.
- Norstedts sv–eng 1993 = *Norstedts stora svensk-engelska ordbok*. 2:a upplagan. Stockholm: Norstedts.
- Norstedts sv–fra 1998 = *Norstedts stora svensk-franska ordbok*. Stockholm: Norstedts Ordbok.
- OALDCE 1989 = *Oxford Advanced Learner's Dictionary of Current English*. Fourth Edition. Chief Editor: A.P. Cowie. Oxford: Oxford University Press.
- Robert-Collins 1978 = *Robert-Collins French-English English-French Dictionary*. Glasgow: Collins.

- Sachs-Villatte 1911 = *Sachs-Villatte enzyklopädisches französisch-deutsches und deutsch-französisches Wörterbuch. Hand- und Schulausgabe. Erster Teil: französisch-deutsch.* Revidierte Ausgabe. Berlin-Schöneberg: Langenscheidtsche Verlagsbuchhandlung.
- Wahrig 1980 = G. Wahrig: *Deutsches Wörterbuch.* Völlig überarbeitete Neuausgabe. Mosaik Verlag.

B. ÖVRIG LITTERATUR

- Hausmann, F.J. & Werner, R.O. 1991. "Spezifische Bauteile und Strukturen zweisprachiger Wörterbücher: eine Übersicht." *Wörterbücher, Dictionaries, Dictionnaires* 1989-91:2729-2769.
- Sinclair, J. 2001. "Lexical Grammar" (denna volym s. 323-343).
- Svensén, B. 1987. *Handbok i lexikografi. Principer och metoder i ordboksarbetet.* Stockholm: Esselte Studium och Tekniska nomenklaturcentralen.
- Werner, R. 1980. "Ein technisches Detail der Stilebenenzeichnung im zweisprachigen Wörterbuch." *Lebende Sprachen* 25:154-158.
- Werner, R. 1991. "Die Markierungen im zweisprachigen Wörterbuch." *Wörterbücher, Dictionaries, Dictionnaires* 1989-91:2796-2803.
- Wiegand, H.E. 1989. "Der Begriff der Mikrostruktur: Geschichte, Probleme, Perspektiven." *Wörterbücher, Dictionaries, Dictionnaires* 1989-91:409-462.
- Wolski, W. 1989. "Formen der Textverdichtung im allgemeinen einsprachigen Wörterbuch." *Wörterbücher, Dictionaries, Dictionnaires* 1989-91:956-967.
- Wörterbücher, Dictionaries, Dictionnaires* 1989-91. Ed. F.J. Hausmann/O. Reichmann/H.E. Wiegand/L. Zgusta. Berlin/New York: Walter de Gruyter.

Lars Törnqvist

Från engelsk–svensk namnordbok till svensk–engelsk

Engelsk-svensk namnordbok, published in 1996, is a bilingual dictionary comprising 3600 proper names which are spelled differently in English and Swedish. The articles contain short explanations, usually consisting of the genus proximum only. The aim of the author was to create a tool for translators, but many people are actually using the dictionary for other purposes, e.g. as a passive reception dictionary or as a language learning tool. The author is currently engaged in compiling a Swedish-English proper-name dictionary, which isn't just a reversed version of the English-Swedish counterpart, but a dictionary with a wider scope. As the English pronunciation of the names will be given in the new dictionary, it will comprise not only names with different spelling, but also names which are pronounced differently in Swedish and English.

Bakgrund

Egennamn förekommer inte så ofta i vanliga ordböcker. Det finns åtminstone två skäl för detta: dels är den totala mängden av namn ohanterlig och dels medför egennamnen sällan några språkliga problem. Nya egennamn lånas in från andra språk i fullkomligt ohämmad omfattning utan olägenheter. Stavningen behålls oförändrad (åtminstone om det är fråga om västerländska språk) och uttalet anpassas bara nödtorftigt – och detta fungerar tillfredsställande. Däremot uppstår problem i de sällsynta fall när namnen anpassas till de lånande språken och man står inför uppgiften att översätta mellan två språk där namnformerna skiljer sig åt. Ett försök att skapa ett hjälpmedel för att hantera det problemet var *Engelsk-svensk namnordbok*.

Under min tid som handläggare vid Tekniska nomenklaturcentralen medverkade jag i flera ordboksprojekt, och i arbetsmaterialet till dessa ordböcker förekom ofta egennamn i olika språkliga varianter. Dessa var svåra att hantera i det normala beskrivningsformatet för facktermer – framförallt var det omöjligt att

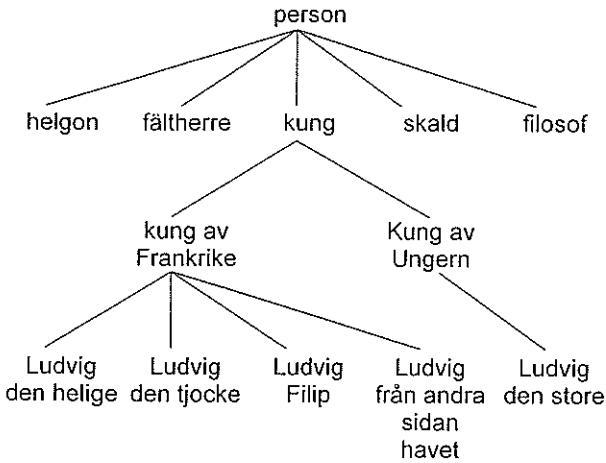
skriva korrekta terminologiska definitioner – och därför blev de i allmänhet bortsorterade i tidigt stadium. Jag kände mig mer och mer otillfredsställd med detta förfaringssätt, och jag började därför fundera på hur egennamn skulle kunna hanteras i översättningsordböcker.

Detta var problemet; sedan upptäckte jag behovet. I början av 1990-talet kom de kommersiella TV-kanalerna med ett stort utbud av filmer, översatta från främst engelska. Många av översättarna var uppenbarligen orutinerade, och de hade problem inte minst med namn på kungar, helgon, bibliska personer, grekiska gudar och filosofer och andra personer som det ofta anspelas på i dialoger. Och det var inte bara i de nya TV-kanalerna som folk hade problem; även i seriösa dagstidningar kunde man se översatta telegram om exempelvis ”påven John Paul II” (i stället för Johannes Paulus II). Här skulle det behövas ett hjälpmedel – en engelsk-svensk ordbok över egennamn.

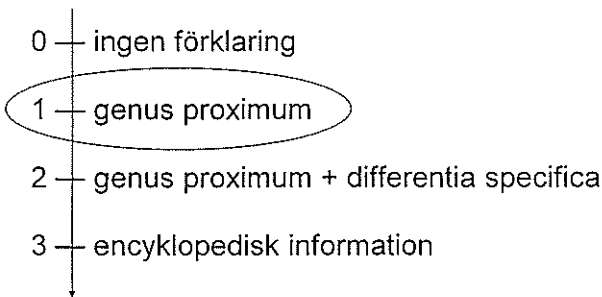
Modell

Det fanns inga lämpliga förebilder för den tänkta ordboken. Därför var det nödvändigt att först skapa en modell för ordbokens struktur. En av inspirationskällorna till modellen var den traditionella terminologiska ordboken. Varje person eller ort eller konstnärligt verk i namnordboken motsvarar ju ett begrepp i en terminologisk ordbok, och ett traditionellt terminologiskt angreppssätt enligt figur 1 kan vara ganska funktionellt. Här var det dock bara fråga om att behandla den allra nedersta nivån i begreppssystemet.

Ett av problemen var att bestämma en lämplig nivå för betydelsebeskrivningen. En skala för förklaringar med olika intentionsdjup visas i figur 2. Vanliga tvåspråkiga översättningsordböcker håller sig vanligen till nivå 0, men tillämpar ibland nivå 1 för att särskilja olika betydelser. Terminologiska ordböcker håller sig normalt till nivå 2, men når inte sällan nivå 3.



FIGUR 1. *Exempel på begreppssystem för historiska personer.*

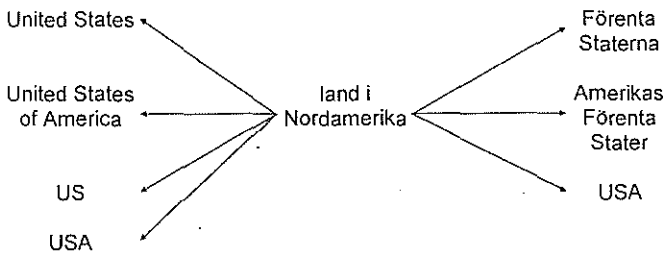


FIGUR 2. *Grad av intentionsdjup hos förklaringar i ordböcker.*

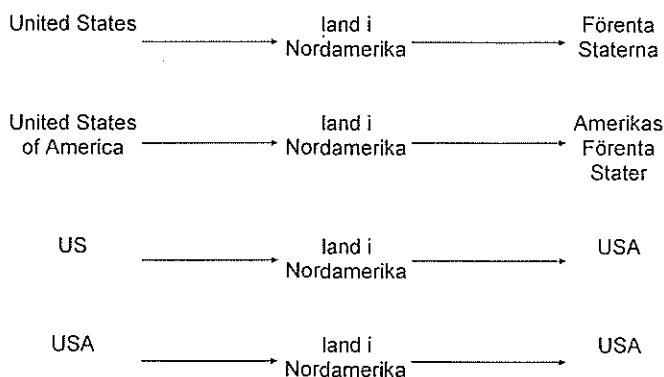
För den tänkta ordboken var nivå 0 otillräcklig, eftersom många av dess översättningsekvivalenter har ganska begränsad giltighet – *James* motsvaras ju av *Jakob* bara om det är fråga om en kung eller ett helgon, inte om personen i fråga är exempelvis en nutida politiker eller idrottsman. Nivå 2 är svår att tillämpa på egenamn, eftersom det är nästan omöjligt att ange särskiljande kännetecken på ett enkelt och konsekvent sätt. Jag valde därför

att hålla mig till nivå 1, vilken motsvarar närmast högre nivå i begreppssystemet (figur 1).

Principerna för hantering av synonymer och kortformer behövde också fastställas. Här har terminologiska ordböcker en helt annan struktur än vanliga översättningsordböcker. Utmärkande för terminologiska ordböcker är att begreppet står i centrum och att alla språkliga uttryck knyts direkt till begreppsrepresentationen. Detta innebär att det inte finns någon direkt relation mellan term motsvarigheter på olika språk (figur 3). Översättningsordböckerna har däremot direkta relationer mellan betydelsemässiga och stilistiska motsvarigheter på de olika språken, men de har å andra sidan inga relationer mellan olika stilistiska varianter av uttryck för samma begrepp (figur 4). Den senare modellen ger en enklare ordboksstruktur eftersom man slipper hänvisningar från synonymer. Eftersom den tänkta ordboken i första hand var avsedd som översättningshjälpmedel, inte som nyckel till begrepps samband, valde jag att använda översättningsordboken som modell.



FIGUR 3. *Logisk struktur hos terminologiska ordböcker. Begreppet är utgångspunkten och alla språkliga uttryck på de olika språken är sekundära i förhållande till begreppet.*



FIGUR 4. *Logisk struktur hos översättningsordböcker. Stilistiska motsvarigheter i källspråk och målspråk har direkt koppling, men begreppsmässiga samband saknas.*

För att få en viss kvalitetsgaranti tog jag bara med namnformer som var belagda i någorlunda pålitliga källor. I huvudsak använde jag allmänna encyklopedier samt geografiska, historiska och biografiska uppslagsverk. Arbetsmetoden beskrivs närmare i Törnqvist (1995).

Resultatet av detta arbete blev en ordbok med cirka 3600 artiklar, omfattande namn på historiska personer och personer ur mytologi, sagor och tecknade serier, namn på länder, floder och städer, bok- och filmtitlar samt vissa namn på internationella organisationer. Ordboken tog två år att sammanställa och utgavs 1996.

Reaktioner och idéer

Efter utgivningen av *Engelsk-svensk namnordbok* kom reaktionerna från recensenter och läsare. Recensenterna ägnade – som väntat – stort utrymme åt förteckningar över namn som de ansåg borde ha varit med, och de påpekade också vissa stavfel och

missvisande betydelsebeskrivningar. Från annat håll kom synpunkter som var mer oväntade. Den allra första reaktionen kom från en engelsklärare som tyckte att boken var fullkomligt värdelös eftersom den inte angav uttalet för de engelska namnen. Det visade sig att även de flesta andra som hade synpunkter på ordboken använde den för ändamål som den inte var tänkt för. Många såg den som ett hjälpmedel vid läsning av engelsk text och reagerade på att den innehöll alltför många självklara översättningar, medan andra använde boken som skrivhjälpmedel för engelska och var missnöjda med att så många namn saknades. Några efterlyste också utförligare upplysningar om de nämnda personerna och orterna.

En annan synpunkt, som har kommit från flera håll, är att det finns större behov av en ordbok i andra riktningen: "Men för en svensk, som utan vidare tyder Chekhov som Tjechov, är den motsatta vägen, den från Tjechov till Chekhov, problematisk. Det är alltså vid överföringen av namn från svenska till engelska som svensken främst behöver hjälp. Och eftersom boken visar vissa tecken till att ha tillkommit med datorhjälp, skulle framställandet av en svensk-engelsk version kosta relativt liten möda." (Kjellmer 1997).

Det är lätt att avfärda vissa av ovanstående synpunkter med att användarna inte har förstått ordbokens syfte. En mera konstruktiv inställning är att fundera över vilka syften användarna anser att den borde ha, och att anpassa ordboken till dessa. Eftersom många har antytt att *Engelsk-svensk ordbok* har allför begränsad användbarhet, så borde en eventuell reviderad utgåva utformas så att den inte bara fungerar som aktivt hjälpmedel för översättare, utan även fyller funktionerna som passiv ordbok och språkläromedel. Antalet namn borde därför ökas väsentligt och uttalsangivelser tillfogas, och högst på prioriteringslistan borde stå en motsvarande svensk-engelsk namnordbok.

Vändning och bearbetning

Att vända en ordbok helt mekaniskt kostar relativt liten möda. Resultatet blir dock en ganska dålig ordbok, främst beroende på att urvalet av uppslagsord inte blir naturligt i förhållande till ordens förekomst i källspråket. En nödvändig åtgärd när man gör om den engelsk-svenska ordboken till en svensk-engelsk är alltså att man kompletterar med namn som är vanliga i svenska texter men sällan förekommer i engelska, framförallt namn på svenska personer, institutioner och litterära verk.

En viktig sak som måste beaktas vid vändning är föråldrade eller på annat sätt olämpliga namnformer. Föråldrade namn bör finnas med som uppslagsord på källspråket med tanke på att användarna kan möta sådana namn i äldre litteratur; däremot bör man aldrig ange icke rekommenderade former på målspråket. I den engelsk-svenska ordboken finns därför en viss rysk författare på tre ställen, därav två försedda med ett kors som markering av olämplig form:

Chekhov → Tjechov
 Chehov† → Tjechov
 Tchekhov† → Tjechov

I den svensk-engelska ordboken finns nämnde författare bara på ett ställe, eftersom hans namn bara har en svensk stavning, och bara den rekommenderade engelska stavningen anges:

Tjechov → Chekhov

I den engelsk-svenska ordboken står förnamn och efternamn i omvänd ordning på källspråket, men i rättvänd ordning på målspråket. Detta underlättar letandet i ordboken och ger en tydligare upplysning om hur namnet skrivs på målspråket. När man vänder ordboken måste man beakta även detta. Den enklaste lösningen på det problemet är att man har separata fält för förnamn och efternamn på de båda språken i arbetsdatabasen. För-

namnsfältet kan också användas för bestämd artikel i förekommande fall.

Tillfogandet av uttalsangivelser är den uppgift som innebär mest arbete. Dels medför det att namnurvalet måste utökas till att omfatta även namn med samma stavning men olika uttal, dels är det mycket tidsödande att leta efter pålitliga uttalsuppgifter. Jag har i möjligaste mån utnyttjat de stora engelska uttalsordböckerna Jones (1997) och Wells (1990). Många lågfrekventa namn saknas dock i dessa ordböcker, så jag har fått använda många andra källor med varierande grad av pålitlighet. Som modell för den fonetiska notationen i ordboken har jag valt Jones (1997) med vissa förenklingar.

För att kunna hantera materialet på ett praktiskt sätt har jag lagt upp det i en Access-databas med följande fält:

- Klassifikation
- Engelsk sortering
- Engelskt prefix (förnamn/artikel)
- Engelskt huvudnamn
- Källa för engelskt namn
- Engelskt uttal
- Källa för engelskt uttal
- Förklaring på svenska
- Svensk sortering
- Svenskt prefix (förnamn/artikel)
- Svenskt huvudnamn
- Källa för svenskt namn
- Selektion för Engelsk-svensk ordbok
- Selektion för Svensk-engelsk ordbok

- Dafne** person i grekisk mytologi Daphne ['dæf.ni]
Dafnis och Chloe roman av Longos Daphnis and
 Chloe ['dæf.nɪs.ənd.'kləʊ.i]
Dagestan land i Europa Daghestan
 [ˌdɑːgr.'stɑːn]
Daidalos person i grekisk mytologi Daedalus
 ['diː.də.ləs]
Dakien historiskt rike i Europa Dacia ['deɪ.si.ə]
Dalarna landskap i Sverige Dalecarlia
 [ˌdæ.lɪ.'kɑːl.jə]
Dalmatien landskap på Balkanhalvön Dalmatia
 [dæl.'meɪ.fə]
Damaskus stad i Syrien Damascus
 [də.'mæs.kəs]
Damianus helgon Damian ['deɪ.mi.ən]
Damokles hovman i Syrakusa Damocles
 ['dæm.ə.kliːz]
Danelagen del av England under vikingatiden
 (the) Danelaw ['deɪ.ləʊ]
Dardanellerna sund mellan Europa och Asien
 (the) Dardanelles [ˌdɑːdə.'neɪlz]

FIGUR 5. *Exempel ur Svensk-engelsk namnordbok.*

Tidsplan

- * Utarbetandet av *Engelsk–svensk namnordbok* drevs som fritidsprojekt utan extern finansiering, och hela arbetet tog två kalenderår i anspråk. Författningen drivs under samma betingelser. Detta innebär att arbetet kan ligga stilla månader i sträck på grund av brist på fritid. Med de ambitioner jag har för revidering och vändning av ordboken är projektet mer omfattande än sammanställningen av den ursprungliga ordboken, och det är sannolikt att arbetet kommer att ta ytterligare några år i anspråk.

Litteratur

- Jones, D. 1997. *English Pronouncing Dictionary*. 15th ed. Cambridge: Cambridge University Press.
- Kjellmer, G. 1997. "Lars Törnqvist, *Engelsk-svensk namnordbok*." I: *LexicoNordica* 4.
- Törnqvist, L. 1995. "Engelsk-svensk namnordbok". *Nordiske studier i leksikografi*. 3. Reykjavík: Nordisk forening for leksikografi.
- Törnqvist, L. 1996. *Engelsk-svensk namnordbok*. Lund: Studentlitteratur.
- Wells, J. C. 1990. *Longman Pronunciation Dictionary*. Harlow: Longman.

Ruth Vatvedt Fjeld

Om leksikalsk ubestemthet

Indefiniteness is a precondition for efficient linguistic communication. Modern linguistic theories emphasize the context relative meaning of lexical units. This article gives an overview of different types of indeterminacy and indicates how lexical indeterminacy as polysemy and vagueness have been treated in traditional dictionaries, and point out some alternative approaches.

1. Innledning

Et språk består av ord og av regler for hvordan ordene kan kombineres. Ordene er merkelapper på deler av "verden", ved hjelp av dem lager vi kategorier for konkrete, abstrakte, forestillinger og erfaringer og alt vi kan tenke oss. Det er imidlertid ikke alltid godt å vite hva som hører til hvilken kategori, og betegnelsene gjøres da mer eller mindre ubestemte. Ubestemthet er dermed egenskaper som hefter ved svært mange leksikalske enheter i naturlig språk. Opp til ca. 1970 var det blant språkforskere likevel enighet om tre semantisk-teoretiske forutsetninger for leksikalsk beskrivelse:

- (1) Betydning kan analyseres i semantiske trekk eller kjennetegn.
- (2) Betydning kan analyseres uttømmende ved hjelp av kjennetegnene.
- (3) Medlemmene av en kategori har sine kjennetegn med nødvendighet.

De teoriutviklende lingvistene så helt bort fra at innholdsmangfold og betydningsvariasjon er en språkøkonomisk grunnbetingelse for alt naturlig språk. For leksikografien, der man må finne praktiske måter å beskrive naturlig språk på, har disse teoretiske forutsetningene gitt store konsekvenser.

Men i dag er de fleste språkforskere enige om at ubestemthet er et karakteristisk trekk ved de leksikalske enhetene i naturlig språkbruk. Men hvordan kan leksikografer takle det at den noe naive forestillingen om at ord har en fast og avgrenset betydning, ikke lenger godtas? Og hvordan kan man lage ordboksdefinisjoner med en forståelse av at ords betydning er relativ?

Ubestemthet kan spesifiseres i vaghet og flertydighet. I ordbøkene er flertydighet og vaghet vanligvis behandlet på en av følgende fire forskjellige måter:

- flertydighet blir behandlet som homonymi med klart atskilte betydninger;
- flertydighet blir behandlet som polysemi med beslektede betydninger;
- vaghet blir behandlet som polysemi med beslektede betydninger;
- vaghet blir underslått.

Det jeg her vil sette søkelys på, er leksikografenes ofte svært arbeidsintensive betydningsbeskrivelse av polyseme leksemer. Det man kan lure på, er hvor verdifull en komplisert betydningsinndeling i polysemer er, og videre hvor mye lingvistisk realitet det er i en slik inndeling.

En systematisk strukturering av et språks leksikon for presentasjon som løsrevne leksemer i en ordbok er imidlertid ikke bare en lingvistisk utfordring, leksikografene må også ta hensyn til lesbarhet og tilgjengelighet. Når vurderingene ikke alltid tas ut fra lingvistiske hensyn, men også fra pedagogiske eller estetiske, kan det skje at lingvistiske og andre hensyn kommer i konflikt med hverandre. Vi må dessverre konkludere med at lingvistikken ennå ikke har klart å utvikle et teoretisk-metodisk verktøy som gir leksikografene mulighet til å beskrive ubestemte leksemer tilfredsstillende. Men Pinkal (1985), som bl. a. bygger på Lakoff (1970) kommer svært langt med tre kriterier eller semantiske instrumenter: tvetydighetstesten, presiseringspåbudet og presiseringsforbudet.

Vaghet og flertydighet følger av forskjellige typer kontekst-avhengighet. Flertydighet kan presiseres i flere delbetydninger, mens ekte vaghet er og blir vag. Enkel flertydighet krever ofte bare litt kontekstspesifisering for å bli entydiggjort, mens vaghet bare kan klargjøres ved *en valgt tolkning* ut fra noen av alle de kontekstvariabler som kan virke inn. I teorier om det mentale leksikon antar man at vage uttrykk har uendelig mange betydninger med samme prototypiske kjerne som er lagret i hjernen som en helhet, mens flertydige uttrykk har et endelig antall betydninger som er lagret på forskjellige steder. Ved vaghet er det et kontinuum av presiseringsmuligheter, mens det ved flertydighet er et lite og endelig antall tolkningsvarianter.

2. Flertydighet

I eldre leksikografi gikk leksikografene ut fra et monosemt tegn-begrep, og hver enkelt delbetydning ble følgelig oppført som eget lemma (jf. Zöfgen 1989:780). Men ordbøkernes makrostruktur ble som følge av den praksisen svært uoversiktlig. Først ut-over på 1700-tallet tok man til med en hierarkisk gruppering av innholdsenheter under ett og samme lemma (jf. Landau 1984: 44). Flertydigheten ble dermed beskrevet innenfor ordbøkernes mikrostruktur, dvs. under det enkelte lemma. Slik blir det stort sett fortsatt gjort, men i nyere korpusbasert dataleksikografi kan man se en tendens til å vende tilbake til den gamle metoden igjen. Etter hvert blir selve datapresentasjonen underordnet, det å isolere og tydeliggjøre enkeltbetydningene er viktigere, og med en enklere mikrostruktur er det lettere å gjøre rede for hvilke syntaktiske og leksikalske forbindelsesmuligheter de enkelte betydningene har, og la det være grunnlag for å separere betydningene i homonymer. Ved korpusstudier viser slike forbindelsesegenskaper seg tydelig.

Som illustrasjon på tradisjonell leksikografisk behandling av flertydighet har jeg valgt beskrivelsen av ordet **messe**. Betydningsinndelingen er forskjellig i *Svensk ordbok*, *Bokmålsord-*

boka, Nudansk ordbog og den noe større *Norsk riksmålsordbok* (jeg har her bare tatt med hovedbetydningene):

Svensk ordbok:

1mässa subst.

- 1 den kristna kyrkans huvudgudstjänst med nattvard
- 2 typ av flerstämmig körkomposition som utvecklats ur kyrkomusiken
- 3 större utställning med viss typ av varor från flera olika försäljare

De tre delbetydningene utgjør ett lemma ut fra felles etymologi. Derimot skilles det mot homonymet **2mässa** etter morfosyntaktisk forskjellighet:

2mässa verb

- tala entonigt og utdraget
 äv. förrätta mässa <åld.>
 äv. framföra sångpartier i gudstjänst <åld.>

Inndelingen i *Bokmålsordboka* er imidlertid annerledes:

I messe fl el. m1 (etymol. par.)

- 1 katolsk (nattverds)gudstjeneste
- 2 (vekselsang mellom prest og menighet i) protestantisk gudstjeneste
- 3 musikkverk med utgangspunkt i den faste teksten i en messe (I,1)

Lemmaet holdes sammen ut fra felles etymologi her også, men samtidig skiller man ut **II messe** som eget lemma pga. semantisk forskjellighet:

II messe fl el. m1 (eg om marked i forbindelse med messe (I)) kjøp-
 stevne, særlig med tanke på salg til grossist; vareutstilling.

Dessuten er et tredje homonym skilt ut, muligens ut fra en ganske diskutabel etymologi (jf. Falk & Torp 1903–6):

III messe fl [m1] (eng. *mess* 'matrett'...) matsal og oppholdsrom for militært personell, skipsmannskap el l

Forskjellig morfosyntaks gjør at det fjerde homonymet skilles ut:

IV messe v1 (av *I messe*)

- 1 om prest: halvt tale, halvt synge i altertjenesten
2 si fram noe på en høytidelig, halvt syngende måte

I *Nudansk ordbog* er inndelingen gjort på en tredje måte:

messe I subst. – 1. jf. *messe III*) i den protestantiske kirke om altertjeneste, (spec.) om recitativlignende sang, hvori præsten fremfører kollekt, epistel og evangelium; jf *højmesse* – 2. i den katolske kirke: gudstjeneste; *stille m.* gudstjeneste uden sang, og hvor evangeliet læses sagte; *holde, læse m.:* i musik: komposition for kor og solostemmer i afsnit, der svarer til den katolske gudstjeneste – 3. forkort. af *kermesse* (s.d.) om købestævne – [middelda.; gennem angelsaksisk fra middellat. *missa* egl. perf.partc. fem. af latin *mittere* sende, jf. *mission*; antagelig efter de ord, hvormed den katolske gudstjeneste slutte: *ite* (gå), *missa est* (*contio*) (forsamlingen) er sendt bort]

messe II subst. spisesal (og opholdsrum) *officersmesse, mand-skabsmesse* i skibe – [fra eng. *mess* messe, ret mad; sml. fransk *messe* retter; af latin *mittere* sende, om de fra køkkenet sendte retter; jf *messe I*]

messe III verb: forrette altertjeneste; synge evangeliet; (overf.): tale høytideligt (og ensformigt); – [af. af *messe I I*]

I den største ordboka for norsk bokmål, *Norsk Riksmålsordbok* (NRO), har betydningsinndelingen gitt to lemmaer med hver sine polysemer som er organisert på to nivåer. Det øverste, som markerer ”grunnforskjellige hovedbetydninger” ifølge forordet, er igjen delt inn i ”de nødvendige tallnummererte underavdelinger” (NRO:1937:XIV). Ordboka oppgir 2 homonymer og 7 polysemer, noe som er en ganske fin betydningsinndeling, til tross for at redaksjonen ifølge forordet (ibid.) har lagt vekt på å lage vide betydningsinndelinger:

I. messe ... A. (ags. *mæsse* over vulgærlat. *messa* fra kirkelat. *missa*, til lat. *mittere* ’sende’; betydningsutviklingen tvilsom)

1) a) romersk-katolsk gudstjeneste, hvorunder Kristi forsonings-offer fornyes derved at alterbrødet og altervinen tenkes å forvandle seg til Kristi legeme og blod, som nydes av presten (på menighetens vegne ell. til gagn for enkeltpersoner, for avdødes sjeler):

b) protestantisk gudstjeneste

2) a) mus., musikk som knytter sig til den katolske messe (solo, kor, orgel og/ell. orkester)

b) (i den protestantiske gudstjeneste) prestens sang og menighetens svar til den

3) (som siste ledd i smnsetninger, ofte i formen *-messe*) kirkelig festdag (fra romersk-katolsk tid); særl., helgendag, f. eks. *kyndelsmesse*

B. (ty. *messe*, sm. ord som i A) **1)** særl. om eldre forh., større, fast, årlig kjøpstevne, ofte av internasjonale dimensjoner (tradisjonsmessig og opr. av praktiske grunner henlagt til tiden for en stor kirkefest) **2)** (i moderne tid) større prøveutstilling (og demonstrasjoner) av varer, med anledning for produsenter og detaljhandlere til å treffe salgsavtaler, og oftest utstyrt med restauranter og forlystelsesetablissemeter, jvf. *varemesse*

II. messe (eng. *mess* 'bordselskap...') felleshusholdning for offiserer og underoffiserer i tjeneste (ombord i skib ell. i leir, jvf. *oficersmesse*...) ell. for annet overordnet personell som på grunn av sitt arbeide lever mere ell. mindre på feltfot (f. eks. ingeniører ved et anlegg)

messe, -et, -et, -ing (om prest under gudstjeneste) halvt lese, halvt synge (bønner ell. stykker av Bibelen) foran alteret

De fire ordbøkernes betydningsinndeling kan skjematisk framstilles slik:

	NRO	SO	BOB	NUD
romersk-katolsk gudstjeneste	1.A.1.a	1.1	1.1	1.2
protestantisk gudstjeneste	1.A.1.b	1.1	÷	1.1
musikk v/katolsk messe	1.A.2.a	1.2	1.3	÷
vekselsang i protestantisk gudstjeneste	1.A.2.b	1.2	1.2	÷
helgendag	1.A.3	÷	÷	1.3 (henv.)
kjøpstevne	1.B.1	÷	2	1.3
vareutstilling	1.B.2	1.3	2	÷
spisesal	2	-	3	2
å messe v/gudstjeneste	(3)	2	4.1	3
høytidelig tale	÷	2	4.2	3

I en artikkel om definisjonen av lemmaet *fri* har jeg sammenliknet betydningsinndelingen i åtte forskjellige ordbøker (Fjeld 1997). Inndelingen varierte fra 32 polysemer i *Oxford English Dictionary* til 5 i *Norsk Riksmålsordbok*, uten at det var mulig å se at antallet hadde sammenheng med typen ordbok eller størrelsen på ordbøkene. De forskjellige redaksjonene fulgte forskjellig, men implisitt, praksis, og det var tydelig å se at inndelingen ofte var styrt av det eksempelmaterialet som tilfeldigvis var tilgjengelig.

I praktisk leksikografi legges det stort arbeid i å skille ut delbetydningene eller polysemene og beskrive dem i forhold til hverandre. En undersøkelse av hva leksikografer synes er vanskeligst i ordboksarbeidet, viste at "Splitting, identifying the senses for a word" fikk nest høyest vanskelighetsskår, etter det mer uklare "Finding right wording" (Kilgarriff 1998:54). Men så lite konsistent betydningsinndeling har liten informasjonsverdi for brukerne og vitner mer om forskjellige leksikografiske arbeidsmåte enn om dokumentasjon av systematiske leksikalske relasjoner.

Med utgangspunkt i Bo Svenséns *Handbok i leksikografi* (1987) viser Malmgren (1992) hvordan syntagmatiske kriterier kan være utilstrekkelige eller gi unødig komplisert betydningsinndeling, og også hvordan den tradisjonelle strukturalistiske kommutasjonsprøven med synonymer er utilstrekkelig som arbeidsredskap for å beskrive polysemien ved et lemma. Malmgren viser til **tvetydighetstesten** slik den er presentert i Mel'čuck 1988, en test som opprinnelig ble lansert av Lakoff (1973).

Tvetydighetstesten forbyr koordinering av strøkne ledd:

Bjørn gikk til en messe, og det gjorde Erik også

Denne koordineringen kan vi ikke akseptere hvis tolkningen skal være at Bjørn går i St. Olavskirken mens Erik drar på Skøyen varemesse. Tvetydige uttrykk er nemlig underlagt et **presiseringspåbud** slik at bare én av betydningene kan aktiveres. Ved tvetydighet vet jo den som snakker alltid hvilken betydning som er ment. Slik koordinering kan kun gjøres for å gi en morsom

effekt, et zeugma. Der tvetydighetstesten gir positivt utslag, virker altså presiseringspåbudet.

3. Vaghet

Systematisering og beskrivelse av flertydighet er en gjenganger i leksikografiske metodediskusjoner. Derimot er verken innsamling, systematisering eller beskrivelse av vage uttrykk gjort for norsk, og en metodediskusjon om dem har meg bekjent ikke vært ført i leksikografimiljøene. Verken det store trebindsverket om leksikografi på over 3000 sider fra 1989 eller Bo Svenséns *Handbok i leksikografi* tar opp vaghet som eget tema. Et lite unntak utgjør Malmgren (1994), i den kortfattede læreboka *Svensk lexikologi* vies problemet en halv side.

Når et uttrykk har svært mange betydningsnyanser som gir forskjellige tolkningsmuligheter, er det sannsynligvis ikke lenger snakk om polysemi, men om vaghet. Polysemiinndeling som er svært lite konsistent fra ordbok til ordbok, kan egentlig være forsøk på å presisere vage leksikalske enheter ut fra de kontekstvariasjonene som forekommer i de mer eller mindre tilfeldig innsamlede beleggene. Det er svært vanskelig å avgjøre hva som er kontekstavhengige bruksvalører av samme betydning og dermed ikke et presiserbart polysem som kan inndeles klart i atskilte delbetydninger.

En fullstendig beskrivelse av et vagt tegns betydning omfatter beskrivelse både av et kjerneområde, vaghetsområde og utelukkelsesområde, også kalt *kjerne*, *randzone* og *utelukkelse* (jf Wichter 1988:42). Typisk vage uttrykk har et ubestemtetsområde der det er en kontinuerlig overgang fra uttrykkets positivområde til negativområdet, og i denne overgangssonen kan det trekkes uendelig mange forskjellige grenser. Dette er vist i undersøkelser av hva vi godtar som *stol* (Black 1937) eller *kopp* (Labov 1974). Det er svært vanskelig å si nøyaktig når en kopp er for vid til at vi vil regne den nettopp som *kopp* og ikke som *bolle* eller *skål*, men likevel fins det et sted en grense. Også når en enhet består av diskrete elementer, kan det være problemer

med grensesetting. Dette kalles i filosofien for sorites- eller sandhaugsparadokset – det er ikke godt å si nøyaktig hvor mange sandkorn som skal til før vi har en sandhaug.

Det er nødvendig med en *tolkningstoleranse* ved vage begreper. Der tvetydighetstesten gir negativt utslag, foreligger det vaghet, og da gjelder **presiseringsforbudet**. En vag leksikalsk enhet i naturlig språk er slik at den verken *trenger* presisjon eller *tåler* presisjon uten at den konvensjonaliserte betydningen endres. Vage uttrykk har ingen naturlige presiseringer, for de har et ikke-eliminertbart vaghetsområde. Presiseringspåbudet og presiseringsforbudet har komplementær funksjon. Denne innsikten er svært relevant i leksikografien, der man forsøker å kartlegge begrepenes yttergrenser.

Tvetydighetstesten er utilstrekkelig for å avgjøre vaghet ved naturlige begreper: f. eks. for å trekke skillet mellom *elv/bekk*, *fjell/li/dal*, *tre/bus*k (jf. bl. a. Quine 1960). I *Bokmålsordboka* er *bekk/elv* og *busk/tre* definert slik:

bekk	1 lite, naturlig vassfar 2 strøm av vann eller væske
elv	1 naturlig strøm av vann som renner i et far til havet eller til en innsjø el som renner sammen med en annen elv
busk	1 en forholdsvis lav flerårig vedplante som grener seg ut fra bunnen 2 lite tre som likner en busk
tre	1 flerårig plante som er høyere enn busk, med stamme, kvister, blad, blomster og frukter

Det er ikke mulig ut fra disse definisjonene å trekke grenser mellom *bekk* og *elv* eller *busk* og *tre*, og det er heller ikke gjort forsøk på å beskrive vagheten i grenseovergangen, bortsett fra modifikatorene *lite* og *forholdsvis lav*.

Usikker kategoritilhørighet kan komme av randsoneuskarphet. I andre tilfeller dreier det seg om kjerneuskarphet, som innebærer at selve kategorien er vag fordi prototypen er ubestemt, f. eks. *greie*, *sak*, *grøn*sj.

3.1. Randsoneuskarphet

Mange ord for naturlige arter og sosiale konvensjoner er i det naturlige språket alltid randsoneuskarpe og dermed vage. Ord som *demokrati* og *frihet* og følelsesord som *hat*, *kjærlighet*, *ømhet* har ingen stabil og klart avgrenset betydning som alle i et språksamfunn er enige om. Ofte har slike ord mange konkurrerende betydninger, og uansett hvordan leksikografen bærer seg ad, kan den leksikografiske definisjonen ikke dekke dem tilfredsstillende. Som regel blir randsoneuskarpheten ikke beskrevet i ordboksdefinisjonene.

Den vanlige leksikografiske løsningen er å oppgi noen synonymmer, som etter hvert viser tilbake til de samme leksemene, dvs. det lages større eller mindre definatoriske sirkler. Her følger en kort oppsummering av hva jeg fant ved å forfølge synonymdefinisjonene med utgangspunkt i lemmaet **sorg** i *Bokmålsordboka*:

sorg dyp, varig smerte i sinnet, tung bekymring
smerte noe som gjør vondt, pine, lidelse
bekymring angst, sorg, nedslåttethet

Og om man følger tråden videre, finner man:

angst sterk frykt, otte, uro
frykt uro, engstelse, redsel
otte uro, engstelse
uro 1 bevegelse, liv, fart 2 bråk, larm, leven 3 gjæring, misnøye, nervøsitet 4 mild verk 5 svingende del i urverk 6 lett dekorasjonsgjenstand som beveger seg med luftdraget
engstelse angst, uro, bekymring
nedslåttethet ikke definert, derimot **nedslått** 1 som er slått ned 2 motløs, oppgitt (-> kunne forfølges via **motløshet**, **oppgittethet** ---)
pine lidelse, smerte, plage
lidelse pine, smerte, sykdom
plage lidelse, fortred
fortred skade
sykdom det å være syk -> **syk** ikke frisk -> **frisk** 1 uskjemt, god, uvisnet, ufalmet, ren 2 usvekket, fornyet, ny 3 ny, fersk 4 ikke syk, bra ---

Det ligger nær å tenke på katta som går rundt den varme grøten når man skal prøve å finne ut av slike definisjoner. Konklusjonen må bli at det *er* vanskelig å lage analytiske definisjoner til randsoneskarpe begreper.

Denne uskarpheten kan være ubevisst eller uønsket, slik den er i uttrykk for naturlige arter og konvensjonelt bestemte kategorier, men den kan også være bevisst og villet, det jeg kaller "klart uklar tale". Når klart uklar tale er markert ved randsoneskarpe kategorier ved uttrykk som *en slags*, kalles de ordgardering eller språklige hekker.

3.2. Kjerneuskarphet

En del begreper har verken fast kjerne eller klar randsoner, de er vage tvers igjennom, og er ofte laget spesielt nettopp for å markere det. Det er klart uklar tale. Av og til har vi nemlig behov for å finne et uttrykk for fenomener vi ikke kan eller ønsker eller gidder å definere, f.eks. når noe er nytt og vi ikke vet hva vi skal kalle det. Det vanligste kjerneuskarpe uttrykket er vel ting, med variantene tingest eller **dings**, som brukes mest om konkrete:

Jeg så noen spille curling, voksne damer og herrer som skjøv noen **dingser** med feiekoster så det ut for.

Arbbl. 1968 248 22/1 Inns b

Mens **sak**, **greie** brukes mest om abstrakter:

Fotballspilleren Jan Åge Fjørtoft: Fire millioner nordmenn vet at jeg er kristen. Flere behøver ikke vite det. Nå er det min personlige **greie**.

Arbbl. 1989 132 18/1 Sport b

Ofte regnes slike leksemer til slang eller uordentlig tale, og slik markeres de ofte i ordbøkene, dersom de i det hele tatt er tatt med. Som regel defineres de bare ved synonymer, den spesielle kjernevage betydningen forklares ikke. Felles for dem er at det ikke er helt sikkert hva som er i kjernen, bare en viss antydning, f. eks.:

Det virker for øvrig litt tilfeldig hva som er med i ordbøkene av slike ord. Det er vanskelig å finne belegg på *tingeltangel* i moderne korpora, men det er med i *Bokmålsordboka*:

tingeltangel 1 støyende, enkel musikk, varietéunderholdning, 2 uekte pynt, jugl

Derimot fins ikke det mer vanlige *dingeldangel*, som det er lett å finne belegg på:

FET ENGLEBOA på plastre og glitrende engel med **dingeldangel**

BT981211

Det samme gjelder *duppedit* som er noe lite og litt teknisk:

Det ble mye **duppeditter** og skjulte finesser på komponentplan.
TU 1987 12 12/1-2 Art b

Ordbøkene er altså ganske usystematiske og ufullstendige med hensyn til å lemmatisere og beskrive vagheten i kjerneuskarpe uttrykk. I beste fall må man som bruker avlede kategoriusikre betydninger selv ut fra eksemplene. Korpusbasert leksikografi vil sannsynligvis vise tydeligere hvor viktige slike funksjoner er rent leksikalsk, i ordbøker utarbeidet på basis av manuelt ekserpert materiale er beleggene til dels oversett, til dels unngått av normative grunner.

3.3. Ordgardering

En del uttrykk brukes for å markere at vi er usikre på om det begrepet vi tenker på, passer helt inn i den betegnelsen vi velger: *Han er en slags dikter*. Denne kategorimodifiserende betydningen eksisterer ved siden av grunnbetydningen og markerer en diskutabel kategoritilhørighet som gjør randsonen eksplisitt. Den kategorimodifiserende funksjonen er sjelden tatt med i ordboksdefinisjonene av slike leksemer. Som regel er det adjektiv som brukes slik:

ordentlig : 1 ordinær, sedvanlig 2 skikkelig 3 til gagns, dyktig
(*Bokmålsordboka*)

I utsagn som "nå synes jeg du oppfører deg som en ordentlig unge!" passer ingen av de tre oppgitte betydningene, gardereringen sier noe om kategoritilhørighet og utsagnet har irettesettende virkning bare overfor noen som egentlig ikke hører til i den nevnte kategorien. Lakoff & Johnson (1980) sier at garderering ("hedges") situerer objekter relativt til en forutsatt prototyp. Utsagnet *Han er en typisk ungar!* kan man ikke forstå ut fra en definisjon av ungar som 'en mann som er ugift'. Adjektivet retter oppmerksomheten mot varierende egenskaper som 'egenrådig, uordentlig, råflott ...'. Språklig garderering gjør taleren i stand til å påvirke forholdet mellom et uttrykk og dets innhold ved å lede mottakerens oppmerksomhet mot uttrykkets kjernebetydning eller mot en mer eller mindre frynsete randzone. Man kan ved hjelp av dem til og med tvinge noe som er utenfor definisjonsområdet inn i et uttrykk.

Pinkal (1985) skiller mellom presiserende og depresiserende garderinger, de fleste leksemene som brukes slik, kan være begge deler. Kolde (1989:857) deler gardereringene inn i **sentralitetsindikatorer**, til forskjell fra **periferiindikatorer**. Slike indikatorer opptrer både i form av egne leksemer, som *en slags, virkelig* og som avledningsmorfemer: *-aktig* og det mer sjeldne *-haftig* og er en periferiindikator, mens *-ete* som i *guttete, jentete*, og prefikset *erke-* som i *erkeengel* er sentralitetsindikerende. Som sentralitetsindikerende virker også redupliserte former som *mannemann* mens sammensetninger med forskjellige ledd virker periferiindikerende, som i *guttemann*. Ordboksdefinisjonene har sjelden eksplisitt med den kategorimodifiserende betydningen, bare implisitt ved å oppgi synonymer som har den samme funksjonen. Nedenfor følger noen eksempler på forskjellige kategori-modifikatorer.

3.3.1. Sentralitetsindikerende

Sentralitetsindikerende markører er svært vanlige. Likevel kan man ikke spore det i *Bokmålsordboka* i følgende tilfeldig valgte definisjoner:

virkelig 1 som eksisterer, fins 2 ekte, sann 3 adv, brukt først: sannelig, faktisk
(*Bokmålsordboka*)

sann 1 som stemmer med virkeligheten, riktig, korrekt, 2 ekte, virkelig, 3 ærlig, oppriktig
(*Bokmålsordboka*)

Da *Bokmålsordboka* ble redigert, hadde man bl. a. tilgang til dette belegget:

Jeg er en mann over de 50 som søker en **virkelig** dame som ikke bare ser på statussymboler.

Dagbl. 1971 65 13/6 Inns b

Det var altså ikke mangel på belegg som er årsaken til at den sentralitetsindikerende funksjonen er oversett. Det er et mønster i mange ordbøker at den kategorimodifiserende funksjonen er utelatt, i stedet føres man inn i definisjonssirkler uten at den kategorimodifiserende funksjonen blir oppgitt.

Sentralitetsindikerende redupliserte former som *mannemann* er heller ikke med i *Bokmålsordboka*:

Sverre Anker Ousdal dukker opp overalt for tiden. Fra barsk **manne-mann** i ishavet i filmen "Orions belte", til den lystige unnavikende grev Danilo i "Den glade enke".

KK 1987 23 3/1 Art b

3.3.2. Periferiindikerende

De periferiindikerende kategoriindikatorerne er også svært vanlige i allmennspråk, men som regel dårlig redegjort for i leksikografiske beskrivelser. Det nærmeste ordbøkene kommer, er å stilmarkere dem: *ironisk, humoristisk, spöttisch* eller liknende.

Kolde har i de tyske ordbøkene bare funnet forklaringen 'etwas Ähnliches wie' til *eine Art (von), kind/sort of (a)* Men i *Bokmålsordboka* er funksjonen faktisk eksplisitt markert:

slags slag, sort det var sju s- kaker til kaffen--- den s- mennesker like, han er en s- skuespiller kan til nød regnes som skuespiller.

Det er altså ingen praktiske eller pedagogiske hinder for at ordbøkene tar med denne funksjonen i definisjonene, men det er vanskelig innenfor en analytisk definisjon. Det er bedre om man velger en metaspråklig kommentar, slik som i *Bokmålsordboka* over.

Collins Cobuild English Language Dictionary har også klar markering av denne funksjonen, men legger for sikkerhets skyld til at den er "very informal", noe som vel kan diskuteres:

3 You use **kind of** 3.1 to describe or refer to something that you are uncertain about but that is roughly like the things mentioned; used in very informal English. *I'm a kind of anarchist, I suppose*

...

3.3.3. Tangeringsindikerende

Kolde regner bare med sentralitets- og periferiindikatorer, men det er lett å finne uttrykk som markerer at man beveger seg ut på begrepenes utelukkelsesområde. De kan kalles **tangeringsindikerende**, nemlig *liksom, nesten* o. fl.:

Vanligvis har verden vært et sted for voksne. xx Nå er det omsnudd. Moten blir mer og mer for ungdom, og **nesten** barn.
Berg Arbbl 1968 290 5/1 Underh b

Også her har *Bokmålsordboka* løst problemet med en eksplisitt markering:

nesten adv 1 på det nærmeste, omtrent, så godt som 2 brukt for å gjøre utsagnet litt mindre bestemt el. bindende

Men det er ikke bestemtheten som er det avgjørende, men at overleddet ikke hører med til kategorien i det hele tatt, den som *nesten barn*, er ikke *barn*! Men man er på rett spor definisjonsteknisk ved å forklare funksjonen direkte ved såkalt narrativ klargjøring.

3.3.4. Pseudoindikerende

Det fins også andre typer kategorimodifikatorer, f. eks. **pseudoindikerende**, som prefiksene *pseudo-*, *kvasi-* og den norske ekvivalenten *skinn-*, og med "skaps-avledninger" (Jarvad 1995) som *salong-* eller *skap-*:

I "Klister kafe" går han til felts mot selvopptatte **salong-radikallere** og **kvasi-kunstnere**, men også mot "verten" som doper oss "så vi ikke tør gå ut og slåss".

Berg Tid. 1977 277 4/6,7 Musikk b

Når funksjonen defineres direkte feil, er det tegn på at kunnskapen om språklige garderinger er for dårlig:

liksom conj 1 på samme måte som; som om 2 adv: på en måte, nesten, tilsynelatende 3 subst i uttr *gjøre noe på liksom* late som om en gjør, gjøre på skøy, liksom vi gjorde det bare på liksom
(Bokmålsordboka)

Betydningsnummer 2 sier to helt forskjellige ting, de gir en typisk periferiindikator som synonym, samt en tangeringsindikator (ev. pluss en pseudoindikator.)

3.3.5. Etterstilt kategorimodifikator

I stedet for å være føre var hender det at man i etterhånd føler behov for å modifisere et utsagn. Jeg vil kalle fenomenet for halemodifikator i samsvar med den norske termen "halespørsmål" for engelsk "tag question". Den vanligste halemodifikatoren er *eller* i utsagn som: *Er du dum, eller?* Her åpnes det for at

kategorien *dum* er uheldig valgt. Denne funksjonen er ikke kommet med i ordbøkene:

eller 1 brukt for å sideordne to ledd som gjensidig utelukker hverandre (eksklusiv disjunksjon), 2 brukt for å sideordne to ledd som ikke gjensidig utelukker hverandre (inkludativ disjunksjon), 3 brukt for å binde sammenledd som betyr det samme.

Ingen av betydningene der dekker den kategorimodifiserende funksjonen. I grammatikkene er den heller ikke godt beskrevet:

Eller får her funksjon som ein slags spørjeartikkel. I *Kan du åpne døra, eller?* blir ytringa anten eit vanleg spørsmål (...), eller ho kan oppfattast som ein reprimande.
(Norsk referansegrammatikk 1997: 1131)

Et annet mye brukt haleheng er *og greier*:

I Sovjet synes det å svikte når det gjelder å integrere systemer for å få effektive produksjonslinjer. De kom jo til Kongsberg med fagfolk med doktorgrader og greier, med store teoretiske kunnskaper.

TeknUkebl 1987 39 3/3 Art. b

Nedenfor følger en kort oversikt over noen uttrykk med leksikalske kategorimodifikatorer. Det ville være en morsom leksikografisk oppgave å gjøre listen mer fullstendig.

Sentralitetsindikatorer

en virkelig dame
ei typisk kjerringrekke
en sann svir
et sant rabalder
en ordentlig mannemann
selve grunnlaget

Periferiindikatorer

såkalt
en slags
for så vidt
omtrent

om lag
 rundt
 rundt regnet

Tangeringsindikatorer

nesten
 nær sagt
 4 måneder eller så
 tett oppunder 1000 kroner

Pseudoindikatorer

liksom samling i bønn
 på-liksom-alvor
 kvasifilosofere
 salongradikaler
 skinndød
 skinnargumenter
 pseudonyheter

Halemodifikatorer

eller
 kanskje
 og greier

4. Konklusjon med forslag til endret definisjonspraksis

Ved vage uttrykk er det ikke oppklarende å dele opp beskrivelsen i opp mot femti forskjellige delbetydninger, slik man kan finne eksempel på i en del av de store ordbøkene. Det er sannsynligvis bedre å markere i klartekst eller med leksikografisk metaspråk at overgangen mellom leksikalske enheter er vag. Tolkning av vaghet forutsetter ofte mye kontekstvitene, og en fyllestgjørende leksikografisk betydningsbeskrivelse krever at denne kontekstavhengigheten på en eller annen måte tydeliggjøres. Det kan gjøres med en narrativ utlegning av de spesifikke bruksbetingelsene, slik som i "kan til nød regnes som

skuespiller". En slik definisjonspraksis er utstrakt i *Collins Cobuild*.

Leksikografene tar fortsatt ofte utgangspunkt i en "essensialistisk" klassisk-objektivistisk kategoriseringsmodell, en modell de fleste semantikere i dag mener er utilstrekkelig. Det fører til at vaghet i ordbøkene ofte presenteres som flertydighet. Den vagheten som blir igjen når et ofte tilfeldig antall delbetydninger er analysert ut, blir ikke behandlet eller kommentert. Man kan få det inntrykk at leksikografer prøver å styre eller "rense" språket ved at de stilmarkerer kategoriusikre eller kategorimodifiserte betydninger som *slang* el. l. De regnes som språklige utflukter eller at man ikke tar seg bryet med å være presis, og det bør lukes ut av språket.

Jeg har her stort sett beskrevet flertydighet og vaghet i et fylogenetisk perspektiv. Men ut fra et ontogenetisk perspektiv er ubestemtheten også interessant. Barn oppdager tidlig at det kan være vanskelig å finne rett uttrykk for den kategorien de tenker på. Da lager de ofte lange nølemarkører: *sånn derre derre dings*. I ungdomsspråk og slang er markering av vag kategori ganske vanlig: *greie, dings, dingsboms* og det er som regel lavt presisjonsnivå på kvantorer og kvalifikatorer: *ørten, kult* er for tiden gjengangere, men inventaret skifter fort. Men det er ikke bare unger og ungdom som strør om seg med vaghet, helt siden Jespersens *Language* fra 1933 er det blitt påstått at kvinners språk også har flere vaghetsmarkeringer. Årsaken skulle være at de ikke hadde evne til å tenke klart. Channell (1994) fant imidlertid ikke noe belegg for det i sitt ganske store materiale.

Normerende tendenser skulle ikke hindre leksikografene i å gi en omsorgsfull behandling av kategorimodifikatorer på linje med andre leksikalske egenskaper. Det utskjelte i språket skal også være interessant for leksikografene! Hvis ordbøkene skal avspeile språkbrukernes leksikalske kompetanse, må man også ta med vaghetsuttrykkene.

Litteratur

A. ORDBØKER

- Collins Cobuild English Language Dictionary*. 1987. London & Glasgow.
- Svensk ordbok*. 1986. Utarbetad vid Språkdata, Göteborgs universitet. Solna.
- Falk, H.S & A. Torp. 1903–06. *Etymologisk ordbog over det norske og det danske sprog*. Oslo.
- Knudsen, T. og A. Sommerfelt (red.). 1937–1997. *Norsk Riksmålsordbok*. Oslo.
- Landrø, I. og B. Wangensteen (red.). 1993. *Bokmålsordboka. Definisjons- og rettskrivningsordbok*. Oslo.

B. ANNEN LITTERATUR

- Black, M. 1937. "Vagueness. An exercise in logical analysis." I: *Philosophy of Science* 4.
- Channell, Joanna. 1994. *Vague Language*. Oxford.
- Fjeld, Ruth Vatvedt. 1997. "Definisjoner og betydningsinndeling i Norsk Ordbok belyst ved artikkelen *fri*." I: Lars S. Vikør (red.): *Norsk Ordbok – nynorskens leksikografiske kanon? Rapport frå eit seminar på Blindern 31. mars 1996*.
- HSK 5.1–3 = F.J. Hausmann/O. Reichmann/H.E. Wiegand/L. Zgusta (red.): *Wörterbücher/Dictionaries/Dictionnaires. Ein internationales Handbuch zur Lexikographie. Handbücher zur Sprach- und Kommunikationswissenschaft*. Band 5.1. Berlin-New York.
- Jäger, Ludwig & Sabine Plum. 1989. "Probleme der Beschreibung von Gefühlswörtern." I: HSK 5.1–3.
- Jarvad, Pia. 1995. "Kryptosammensætninger og skabsafledninger i virkeligheden og i ordbøgerne." I: Svavarsdóttir, Á, G. Kvaran & J.H. Jónsson: *Nordiske studier i leksikografi*. 3. Reykjavík.
- Jespersen, Otto. 1922. *Language*, Chapter XIII "The Woman." London.

- Kilgarriff, Adam. 1998. "The hard parts of Lexicography". I: *International Journal of Lexicography* 11. Oxford.
- Klappenbach, Ruth. 1971. "Homonyme oder polyseme Wort?" I: *Deutsch als Fremdsprache*. 7. Jg. Leipzig.
- Kolde, Gottfried. 1989. "Probleme der Beschreibung von sog. Heckenausdrücken im allgemeinen einsprachigen Wörterbuch." I: HSK 5.1-3.
- Labov, William. 1974. "The Boundaries of Words and their Meanings." I: C.-J.N. Bay & R. W. Shuy (eds.): *New Ways of Analyzing Variation in English*. Washington.
- Lakoff, George. 1973. "Hedges: A study in meaning criteria and the logic of fuzzy concepts." I: *Journal of Philosophic Logic* 2.
- Landau, Sidney I. 1984. *Dictionaries*. New York.
- Lyons, John. 1977. *Semantics*. Cambridge.
- Malmgren, S-G. 1992. "Om behandling av polysema ord i enspråkiga ordböcker." I: Fjeld, R.V. (red.): *Nordiske studier i leksikografi*. 1:137-143.
- Mel'čuck, Igor. 1988: "Semantic Description of Lexical Units in an Explanatory Combinatorial Dictionary: Basic Principles and Heuristic Criteria." I: *International Journal of Lexicography* 1:165-188.
- Pinkal, Manfred. 1985. "Kontextabhängigkeit, Vagheit, Mehrdeutigkeit." I: C. Schwarze/D. Wunderlich (Hrsg.): *Handbuch der Lexikologie*. Königstein.
- Wichter, Sigurd. 1988. *Signifikantgleiche Zeichen*. Tübingen.
- Wiegand, Herbert Ernst. 1989. "Der gegenwärtige Status der Lexikographie und ihr Verhältnis zu anderen Disziplinen." I: HSK 5.1-3.
- Zöfgen, Ekkehardt. 1989. "Homonymie und Polysemie im allgemeinen einsprachigen Wörterbuch." I: HSK 5.1-3.

Fredrik Westerland
 Sjur Nørstebø Moshagen
 Eva Orava
 Juhani Birn

Morfologisk analys och disambiguering som stöd i skapandet av frekvensordlistor

When applying morphological analysis (TWOL) and disambiguation (Constraint Grammar) to a text material, the linguist gets an opportunity to search for contextually disambiguated words on the basis of their base forms. A search on a specific base form gives a result that also includes all the inflected forms and some common derivations of the word. This way of assembling inflected words can offer a lot of help in text lemmatization and also in various types of frequency listings. This paper contains and discusses some frequency lists, which have been generated out of a morphologically analysed and disambiguated annual volume of the newspaper Göteborgs-Posten.

De frekvensordböcker som har kommit ut i Norden de senaste decennierna skiljer sig märkbart från varandra såväl till omfånget som till sin inre uppbyggnad. Materialet i de nordiska frekvensordböckerna är till stora delar baserat på tidningstext som har bearbetats och sammanställts i frekvenslistningar av olika slag: i form av listor med de mest frekventa orden i initial- och finalalfabetisk ordning, de mest frekventa orden fördelade på olika ordklasser, eller i form av listor på de ord som bara förekommer en gång i materialet, för att nämna några sorteringsätt i ordböckerna.

De kvantitativa uppgifterna som ges för varje ord varierar lika så allt efter på vilket sätt författarna presenterar sitt material och hur djupgående analysen av orden är. Vissa frekvensordböcker listar orden enbart enligt deras grafiska form, dvs. i den form de uppträder i textmaterialet. Till dessa ordböcker hör *Norsk frekvensordbok – De 10 000 vanligste ord fra norske aviser* (Heggstad 1982), *Dansk Frekvensordbog* (Bergenholtz 1992)

och *Hyppige Ord i Danske Aviser, Ugeblade og Fagblade* (Maegaard & Ruus 1986).

Nusvensk frekvensordbok baserad på tidningstext (del 1–3 Allén 1970–75, del 4 Allén et al. 1980) är den största frekvensordboken på den nordiska marknaden. I den finns materialet även lemmatiserat, dvs. alla böjningsformer av ett visst ord har förts upp under ifrågavarande ords grundform, såsom även textmaterialet i *Nynorsk frekvensordbok* (Vestbøstad 1989) och *Íslensk orðtíðnibók* (Pind et al. 1991). Också i frekvensordboken *Tiotusen i topp* (Allén 1972), som huvudsakligen bygger på de två första volymerna av *Nusvensk frekvensordbok*, har böjningsformerna av ett ord sammanförts under något som i boken benämns "lexikaliska ord".

I denna artikel redogör vi för hur morfologisk analys bland annat vid lemmatisering kan vara till stor hjälp i framställandet av material till frekvensordböcker och vid frekvenslistningar överlag.

1. Morfologisk analys och disambiguering

TWOL (Koskenniemi 1983) är en morfologisk analysmodell som Lingsoft Ab har tillämpat på flera språk. En TWOL-analysator förser varje ord i en text med en eller flera morfologiska analyser, beroende på om ordet är lexikalt entydigt eller flertydigt. En morfologisk analys består av en grundform och ett antal koder som upplyser om ordets ordklass samt, beroende på ordklass, bl.a. om genus, species, numerus, kasus, diates och tempus. Analysatorn markerar också sammansättningsgränserna i ordet. Modellens tillämpning på svenska, SWETWOL (Karlsson 1992), har vidareutvecklats på Lingsoft.

Vid tillämpning av restriktionsgrammatiken Constraint Grammar, CG, disambigueras de olika morfologiska analysmöjligheterna av ett ord så att i idealfallet endast en analys av ordet kvarstår. (För en fullständig presentation av CG-formalismen se Karlsson et al. 1995.) I detta sammanhang är det SWECG, disambigueraren för svensk text, som är aktuell, se Birn 1998.

Om man har tillämpat den morfologiska analysen på ett textmaterial är det möjligt att söka ord på deras grundformer, och med en enda sökning samtidigt få fram belägg på alla böjningsformer av ett ord. Lemmatiseringen av ett textmaterial förlöper på detta sätt betydligt mer smidigt än om man vore tvungen att söka varje böjningsform separat, och man slipper också den redundanta information som sökning med trunkerade ord ger.

För att med exempel kunna belysa hur morfologisk analys och disambiguering av en text kan vara till stöd vid frekvenslistningar har vi tillämpat morfologianalysatorn och restriktionsgrammatiken på årgång 1997 av Göteborgs-Posten, totalt drygt 20 miljoner ord fördelade på knappt 94 000 artiklar. Tidningsmaterialet finns följaktligen i en form som tillåter en mångsidig frekvensanalys med möjligheter till olika slag av frekvenslistningar. De små frekvenslistor som ingår i denna artikel är av utrymmesskäl begränsade till de tio ord som befinner sig på toppen, dvs. de tio mest frekventa orden i respektive exempelgrupp. Det kan i vissa fall anses vara i minsta laget, men med tanke på materialets omfattning i förhållande till det tillgängliga utrymmet är det motiverat att begränsa listan. Om flera ord i slutet av listan har samma frekvens tas alla ord med ifrågavarande frekvenstal med. Varje exempelord anförs i den form som den morfologiska analysen ger det, dvs. som en grundform, vilken åtföljs av koder som berättar i vilken form ordet uppträder i texten.

De ord som den morfologiska analysatorn inte känt igen har uteslutits ur frekvensberäkningarna. Det rör sig till en stor del om felstavningar och ord på främmande språk. Siffror ingår inte heller i undersökningen. En automatisk analys av en text innebär alltid att det i resultatet finns en liten felmarginal som endast en manuell genomgång kunde förhindra. Förskjutningar i frekvensvärden beror dels på att disambigueraren i vissa fall lämnar kvar flera än en analys, dels på att den inte helt kan undvika att den korrekta analysen slopas. I en undersökning av en text på 10 600 ord klarade disambigueraren av att till 96,59 % ge orden en entydig analys, medan en andel på 0,43 % inte hade kvar den korrekta analysen (Birn 1998).

2. Frekvensen för ett visst lemma

Som ett första exempel kan vi ta fasta på frekvensen för ett visst ord och alla böjningsformer av ordet. Den morfologiska analysen är speciellt välkommen när det exempelvis gäller sökning av verb som genomgår vokalväxling vid konjugation, eftersom det vanligen skulle krävas flera sökningar för att man skall kunna komma åt alla böjningsformer av ordet. Valet av exempelord är i detta fall slumpartat såtillvida att vi gjort två uppslag i tolfte upplagan av *Svenska Akademiens ordlista* och på respektive uppslag letat oss fram till det första verbet med vokalväxling i böjningen. Metoden gav verben *snyta* och *gråta*. Med en sökning på vardera ordet fås frekvensen på alla böjningsformer samt vissa avledningar av de båda verben i Göteborgs-Posten 1997.

TABELL 1. *Lemmat snyta.*

9	"snyta" V ACT PRES
7	"snyta" V ACT PAST
5	"snyta" V ACT INF
1	"snyta" V PASS PRES
1	"snyta" V ACT SUPINE
1	"snyta" V ACT IMP
1	"snyta" DER/-ning N UTR INDEF SG NOM
1	"snyta" DER/-ning N UTR INDEF PL NOM

TABELL 2. *Lemmat gråta.*

196	"gråta" V ACT PAST
174	"gråta" V ACT INF
147	"gråta" V ACT PRES
55	"gråta" <V/DER> <PCP1> A UTR/NEU DEF/INDEF SG/PL NOM
29	"gråta" DER/-nde N NEU INDEF SG NOM
26	"gråta" V ACT SUPINE
6	"gråta" <V/DER> <PCP2> A UTR INDEF SG NOM
4	"gråta" V ACT IMP
2	"gråta" DER/-nde N NEU DEF SG NOM
1	"gråta" V PASS PAST

Eftersom analysatorn i hög grad kan skilja på homografer är det också möjligt att med en sökning få fram t.ex. alla former av ord som har samma grafiska grundform, men som hänför sig till

skilda kategorier. Till denna grupp räknas t.ex. ordet *resa*, som ju både är ett verb och ett substantiv.

TABELL 3. *Substantivet respektive verbet resa.*

1174	"resa" N UTR INDEF SG NOM
889	"resa" V ACT INF

Det kan även vara svårt, eller åtminstone tidsödande, att i en text få tag i samtliga former av ord där efterleden utgörs av ett visst lemma men där förleden varierar. En morfologiskt analyserad text är betydligt mer lättbearbetad. En sökning ger vid handen att t.ex. ord med former av verbet *bryta* som sista led oftast uppträder i form av ordet *utbröt* (271 ggr), följt av olika former av verbet *avbryta*. Först på femtonde plats i frekvenslistan kommer följande nya lemma i form av perfekt particip-formen *uppbruten*. Participen (PCP) klassificeras som (verbavledda) adjektiv (A) på basis av böjning och syntaktisk funktion.

TABELL 4. *Lemman på -bryta.*

271	"utbryta" V ACT PAST
201	"avbryta" V ACT INF
146	"avbryta" V PASS PAST
123	"avbryta" V ACT PAST
83	"avbryta" V ACT PRES
65	"avbryta" V PASS INF
59	"avbryta" <V/DER> <PCP2> A UTR INDEF SG NOM
49	"utbryta" V ACT PRES
49	"avbryta" V PASS PRES
42	"avbryta" <V/DER> <PCP2> A UTR/NEU DEF/INDEF PL NOM
	/---/
20	"uppbruta" <V/DER> <PCP2> A UTR INDEF SG NOM

3. Ord i sammansättningar

Frekvensen för ett visst enskilt ord i en text kan intressera en språkvetare, men även frekvensen för det aktuella ordet i eventuella sammansättningar, samt vilken position det har i det sammansatta ordet, kan vara av intresse. Med hjälp av den morfolo-

giska analysen kan man söka efter ett ord och få veta vilken ställning det har, dvs. hur ofta det förekommer initialt, medialt respektive finalt.

Ordet *medium* är i detta sammanhang ett intressant exempelord, med tanke på att det ofta uppträder i varierande former. Som förled i substantivsammansättningar förekommer ordet 467 ggr i den språkriktiga formen *medie-* och 250 ggr som *media-*. I medial ställning förekommer ordet 84 ggr i formen *-medie-*, och i final ställning flest gånger i form av *-media* (542 ggr) och endast hälften så många gånger i form av efterleden *-medium* eller någon böjd form av ordet.

Då man t.ex. tar en titt på de frekventa orden *medie-* och *mediaföretag* märker man att den form som rekommenderas av språkvården, *medieföretag*, förekommer mer än dubbelt fler gånger jämfört med den andra formen. De språkriktiga formerna (den första kategorin i figur 5) har allmänt taget också ett stort försprång, trots att variationen mellan de båda formerna i materialet är bred.

TABELL 5. *Ordet medium som förled i sammansatta ord.*

22	"medie_värld" N UTR DEF SG NOM
18	"medie_företag" N NEU INDEF SG/PL NOM
13	"medie_koncern" N UTR DEF SG NOM
12	"medie_våld" N NEU DEF SG NOM
12	"medie_program" N NEU DEF SG NOM
11	"medie_våld" N NEU INDEF SG NOM
11	"medie_marknad" N UTR DEF SG NOM
11	"medie_företag" N NEU DEF SG NOM
9	"medie_område" N NEU DEF SG NOM
9	"medie_bransch" N UTR DEF SG NOM
	/--/
9	"media_program" N NEU DEF SG NOM
8	"media_företag" N NEU INDEF SG/PL NOM
8	"media_dag" <TIME> <DAY> N UTR DEF PL NOM
7	"media_bild" N UTR DEF SG NOM
6	"media_våld" N NEU INDEF SG NOM
6	"media_värld" N UTR DEF SG NOM
6	"media_magasin" N NEU DEF SG NOM
5	"media_uppbåd" N NEU DEF SG NOM
5	"media_teknik" N UTR DEF SG NOM
5	"media_program" N NEU INDEF SG/PL NOM

- 5 "media_område" N NEU DEF SG NOM
 5 "media_mogul" N UTR INDEF SG NOM
 5 "media_bransch" N UTR DEF SG NOM
 5 "media_bolag" N NEU INDEF SG/PL NOM
 /--/

TABELL 6. *Ordet medium som mellersta led i sammansatta ord.*

- 6 "mass_medie_koncentration" N UTR INDEF SG NOM
 6 "mass_medie_forskning" <RETAIN> N UTR INDEF SG NOM
 4 "mass_medie_uppbåd" N NEU DEF SG NOM
 4 "mass_medie_ekonomi" N UTR INDEF SG NOM
 3 "mass_medie_uppbåd" N NEU INDEF SG/PL NOM
 3 "mass_medie_forskare" N UTR DEF SG NOM
 2 "multi_medie_teknik" N UTR INDEF SG NOM
 2 "multi_medie_projekt" N NEU INDEF SG/PL NOM
 2 "multi_medie_företag" N NEU INDEF SG/PL NOM
 2 "mass_medie_intresse" N NEU DEF SG NOM
 2 "eter_medie_område" N NEU DEF SG NOM
 /--/

TABELL 7. *Ordet medium som efterled i sammansatta ord.*

- 351 "massmedia" N NEU DEF/INDEF PL NOM
 102 "multi_media" N UTR/NEU DEF/INDEF PL NOM
 100 "multi_media" N UTR INDEF SG NOM
 95 "multi_media" N NEU DEF/INDEF PL NOM
 29 "massmedia" N NEU DEF/INDEF PL GEN
 18 "info_media" N UTR/NEU DEF/INDEF PL NOM
 18 "info_media" N UTR INDEF SG NOM
 18 "info_media" N NEU DEF/INDEF PL NOM
 16 "eter_media" N UTR/NEU DEF/INDEF PL NOM
 16 "eter_media" N UTR INDEF SG NOM
 16 "eter_media" N NEU DEF/INDEF PL NOM
 /--/

- 84 "mass_medium" N NEU DEF PL NOM
 62 "mass_medium" N NEU INDEF PL NOM
 13 "mass_medium" N NEU DEF PL GEN
 12 "nyhets_medium" N NEU INDEF PL NOM
 7 "nyhets_medium" N NEU DEF PL NOM
 7 "film_medium" N NEU DEF SG GEN
 6 "mass_medium" N NEU INDEF SG NOM
 5 "eter_medium" N NEU INDEF PL NOM
 4 "orts_medium" N NEU INDEF PL NOM
 4 "köld_medium" N NEU INDEF PL NOM
 4 "film_medium" N NEU DEF SG NOM
 /--/

Ordet *medium* förekommer som väntat som efterled flest gånger i former av *massmedia* och *multimedia*. Analysatorn anger grundformen *-media* för de ord som i texten står i former som slutar på *-media* (den första kategorin i tabell 7 nedan). De ord som däremot står i en böjd form, t.ex. *massmediernas* och *filmmediet*, får slutleden *-medium* som grundform (den andra kategorin i tabell 7). Som efterled tycks *-media* vinna överlägset över *-medium* vad frekvensen beträffar.

Tack vare att ordens sammansättningsgränser markeras kan man också undersöka hur ofta ord som har samma form, sånär som på en bokstav, förekommer t.ex. som mellersta led i ett sammansatt ord. Nedan listas de sammansatta ord vars mellersta led utgörs av en konsonant plus bokstavskombinationen *and*.

TABELL 8. *Ord på konsonant + and som mellersta led i sammansatta ord.*

57	"folk_tand_vård"	N	UTR	DEF	SG	NOM
23	"privat_tand_läkare"	N	UTR	INDEF	SG/PL	NOM
19	"dam_hand_boll"	N	UTR	INDEF	SG	NOM
14	"video_band_spelare"	N	UTR	INDEF	SG/PL	NOM
8	"studie_hand_ledning"	N	UTR	INDEF	SG	NOM
8	"klubb_hand_boll"	N	UTR	INDEF	SG	NOM
7	"över_tand_läkare"	N	UTR	INDEF	SG/PL	NOM
7	"privat_tand_läkare"	N	UTR	DEF	PL	NOM
7	"kassett_band_spelare"	N	UTR	INDEF	SG/PL	NOM
7	"folk_tand_vård"	N	UTR	INDEF	SG	NOM
7	"dörr_hand_tag"	N	NEU	DEF	SG	NOM
	/--/					

4. Frekvensen för en viss böjningsform

Eftersom orden i det analyserade materialet är kodade enligt den form de står i är det även möjligt att ta fram frekvensen för en viss böjningsform. Nedan följer en titt på verben, substantiven och adjektiven i tidningstexten.

4.1. Verb

För verbens del får man med hjälp av den morfologiska analysatorn uppgifter om att Göteborgs-Posten år 1997 innehöll flest verbformer i presens, hela 1 576 546 st. (46,6 %), därefter i infinitiv, 625 840 st. (18,5 %). Preteritumformerna kommer på tredje plats, 558 453 st. (16,5 %), och supinumformerna på fjärde plats, 248 091 (7,3 %). Ovanstående uppgifter gäller verb i aktiv form. Verbet *vara* toppar inte oväntat alla böjningskategorier. Olika hjälpverb figurerar också flitigt bland de mest frekventa verben.

TABELL 9. *Verb: aktiv presens.*

315546	"vara" <COP> V ACT PRES
201493	"ha" <AUX> <SUPINE> V ACT PRES
70389	"kunna" <AUX> V ACT PRES
58889	"säga" <VCOM> V ACT PRES
43353	"ska" <AUX> V ACT PRES
40452	"komma" <ATT-INF> V ACT PRES
37279	"få" V ACT PRES
34335	"skulle" <AUX> V ACT PRES
33264	"bli" <COP> V ACT PRES
29895	"vilja" <AUX> V ACT PRES
	/--/

TABELL 10. *Verb: aktiv infinitiv.*

33386	"vara" <COP> V ACT INF
28080	"få" V ACT INF
27983	"ha" <AUX> <SUPINE> V ACT INF
21153	"bli" <COP> V ACT INF
18873	"ta" V ACT INF
16142	"göra" V ACT INF
12599	"kunna" <AUX> V ACT INF
12363	"gå" V ACT INF
11289	"se" <VCOG> V ACT INF
8832	"komma" <ATT-INF> V ACT INF
	/--/

TABELL 11. *Verb: aktiv preteritum.*

94097	"vara" <COP> V ACT PAST
35778	"ha" <AUX> <SUPINE> V ACT PAST
25439	"få" <AUX> V ACT PAST
23528	"bli" <COP> V ACT PAST
15702	"komma" <ATT-INF> V ACT PAST
11946	"kunna" <AUX> V ACT PAST
11251	"gå" V ACT PAST
10075	"göra" V ACT PAST
9575	"ta" V ACT PAST
9201	"säga" <VCOM> V ACT PAST

/---/

TABELL 12. *Verb: aktiv supinum.*

19108	"vara" <COP> V ACT SUPINE
14793	"få" V ACT SUPINE
9275	"bliva" <COP> V ACT SUPINE
8361	"göra" V ACT SUPINE
6899	"ha" <AUX> <SUPINE> V ACT SUPINE
5879	"gå" V ACT SUPINE
5050	"ta" V ACT SUPINE
4515	"komma" <ATT-INF> V ACT SUPINE
3618	"se" <VCOG> V ACT SUPINE
2697	"kunna" <AUX> V ACT SUPINE

/---/

Även bland verben i passiv är det presensformerna som är allmännast, 111 769 st. (3,3 %). Materialets preteritumformer i passiv är 66 344 till antalet (2 %), dvs. något fler än passiv infinitivformerna, 64 355 st. (1,9 %). Antalet verb böjda i passiv supinum är 36 712 till antalet (1,1 %).

TABELL 13. *Verb: passiv presens.*

2603	"behöva" <INF> V PASS PRES
2445	"kräva" V PASS PRES
1724	"vänta" V PASS PRES
1627	"använda" V PASS PRES
1578	"tvinga" V PASS PRES
1496	"ta" V PASS PRES
1496	"kalla" V PASS PRES
1366	"beräkna" V PASS PRES
1345	"anse" V PASS PRES
1306	"göra" V PASS PRES

/---/

TABELL 14. *Verb: passiv preteritum.*

1366	"tvinga" V PASS PAST
1194	"gripa" V PASS PAST
1172	"skada" V PASS PAST
1152	"göra" V PASS PAST
1114	"ta" V PASS PAST
985	"döma" V PASS PAST
924	"föra" V PASS PAST
822	"hitta" V PASS PAST
705	"bilda" V PASS PAST
670	"drabba" V PASS PAST
	/---/

TABELL 15. *Verb: passiv infinitiv.*

1852	"använda" V PASS INF
1421	"göra" V PASS INF
1325	"ta" V PASS INF
911	"bygga" V PASS INF
869	"se" <VCOG> V PASS INF
731	"tvinga" V PASS INF
658	"utveckla" V PASS INF
657	"lägga" V PASS INF
625	"hålla" V PASS INF
575	"genom_föra" V PASS INF
	/---/

TABELL 16. *Verb: passiv supinum.*

1375	"göra" V PASS SUPINE
882	"drabba" V PASS SUPINE
695	"ta" V PASS SUPINE
659	"tvinga" V PASS SUPINE
476	"tilldela" V PASS SUPINE
471	"använda" V PASS SUPINE
441	"utveckla" V PASS SUPINE
421	"döma" V PASS SUPINE
403	"utsätta" V PASS SUPINE
396	"utse" V PASS SUPINE
	/---/

Förutom verben i aktiv och passiv förekommer det 70 141 deprensverb i materialet. I spetsen står presensformen *finns* med frekvenstalet 37 136.

4.2. Substantiv

Med hjälp av analysatorn får man för substantivens del veta att 1 545 363 substantiv (25,8 %) står i grundform, dvs. som indefinita nominativer i singular, medan motsvarande siffra för substantiven i plural är 653 417 (10,9 %). SWETWOL ger den underspecificerade analysen SG/PL för de ord vilka kan kombineras med såväl singulara som plurala framförställda bestämmningar (t.ex. *hus* som i *ett/några hus*). Frekvensen för de orden presenteras i en separat lista (se tabell 19).

Av den första gruppen substantiv nedan kan det nämnas att 8 002 st. (av drygt 1,5 miljoner) står som första led i uttryck av typen *lager- och kontorsfastighet*.

TABELL 17. *Substantiv: indefinit singular nominativ.*

15051	"del" N UTR INDEF SG NOM
12072	"plats" N UTR INDEF SG NOM
11895	"tid" N UTR INDEF SG NOM
9189	"gång" N UTR INDEF SG NOM
7719	"man" N UTR INDEF SG NOM
6796	"dag" <TIME> <DAY> N UTR INDEF SG NOM
6431	"publik" N UTR INDEF SG NOM
5978	"kommun" N UTR INDEF SG NOM
5817	"väg" N UTR INDEF SG NOM
5633	"hånd" N UTR INDEF SG NOM
	/-:/

TABELL 18. *Substantiv: indefinit plural nominativ.*

32591	"krona" N UTR INDEF PL NOM
19782	"miljon" <NUM> N UTR INDEF PL NOM
9917	"människa" N UTR INDEF PL NOM
9410	"peng" N UTR INDEF PL NOM
8272	"person" N UTR INDEF PL NOM
5718	"kvinna" <FEM> N UTR INDEF PL NOM
5405	"miljard" <NUM> N UTR INDEF PL NOM
4991	"dag" <TIME> <DAY> N UTR INDEF PL NOM
4858	"gång" N UTR INDEF PL NOM
4356	"man" <MASC> N UTR INDEF PL NOM
	/-:/

TABELL 19. *Substantiv: indefinit singular/plural nominativ.*

59045	"år" <TIME> N NEU INDEF SG/PL NOM
20635	"procent" <RETAIN> N UTR INDEF SG/PL NOM
11295	"barn" <FamRel> N NEU INDEF SG/PL NOM
7829	"exempel" N NEU INDEF SG/PL NOM
6670	"sätt" N NEU INDEF SG/PL NOM
6499	"problem" N NEU INDEF SG/PL NOM
6491	"par" N NEU INDEF SG/PL NOM
6228	"ordförande" N UTR INDEF SG/PL NOM
6172	"mål" N NEU INDEF SG/PL NOM
5997	"jobb" N NEU INDEF SG/PL NOM
	/---/

Då man undersöker samma böjningskategorier bland de substantiv som är böjda i genitiv erhåller man följande frekvensuppgifter: 45 458 substantiv (0,76 %) står i indefinit form genitiv singular (2 241 är första led i uttryck av typen *tävlings- och äventyrsprogram*) och 8 865 substantiv (0,15 %) står i indefinit form genitiv plural.

Som jämförelsematerial till de nyss presenterade substantiven i obestämd form kan man ställa upp tidningstextens substantiv i bestämd form. I singular är de definitiva substantiven i nominativ 1 265 375 (21,1 %) till antalet, i plural 324 604 (5,4 %).

TABELL 20. *Substantiv: definit singular nominativ.*

10363	"år" <TIME> N NEU DEF SG NOM
9070	"polis" N UTR DEF SG NOM
8521	"regering" N UTR DEF SG NOM
7472	"kommun" N UTR DEF SG NOM
7299	"tid" N UTR DEF SG NOM
7094	"fråga" N UTR DEF SG NOM
6583	"man" <MASC> N UTR DEF SG NOM
5675	"land" N NEU DEF SG NOM
5443	"ställe" N NEU DEF SG NOM
5308	"sida" N UTR DEF SG NOM
	/---/

TABELL 21. *Substantiv: definit plural nominativ.*

5634	"år" <TIME> N NEU DEF PL NOM
4164	"barn" <FamRel> N NEU DEF PL NOM
3331	"peng" N UTR DEF PL NOM
2487	"politiker" N UTR DEF PL NOM
2345	"elev" N UTR DEF PL NOM
2275	"social_demokrat" N UTR DEF PL NOM
2084	"man" <MASC> N UTR DEF PL NOM
2078	"kommun" N UTR DEF PL NOM
1706	"moderat" N UTR DEF PL NOM
1661	"företag" N NEU DEF PL NOM
	/---/

För genitivens del ger en sökning i det morfologiskt analyserade materialet följande siffror: 113 257 substantiv (1,9 %) står i definit form genitiv singular och 23 872 substantiv (0,4 %) i definit form genitiv plural. Så här ter sig de tio mest frekventa i respektive grupp:

TABELL 22. *Substantiv: definit singular genitiv.*

4778	"dag" <TIME> <DAY> N UTR DEF SG GEN
4776	"år" <TIME> N NEU DEF SG GEN
2959	"land" N NEU DEF SG GEN
2781	"värld" N UTR DEF SG GEN
2423	"kommun" N UTR DEF SG GEN
2053	"regering" N UTR DEF SG GEN
1792	"gårdag" N UTR DEF SG GEN
1355	"stat" N UTR DEF SG GEN
1071	"stad" N UTR DEF SG GEN
1028	"polis" N UTR DEF SG GEN
	/---/

TABELL 23. *Substantiv: definit plural genitiv.*

784	"barn" <FamRel> N NEU DEF PL GEN
506	"social_demokrat" N UTR DEF PL GEN
469	"moderat" N UTR DEF PL GEN
467	"år" <TIME> N NEU DEF PL GEN
320	"kommun" N UTR DEF PL GEN
264	"företag" N NEU DEF PL GEN
257	"elev" N UTR DEF PL GEN
230	"politiker" N UTR DEF PL GEN
228	"förälder" <FamRel> N UTR DEF PL GEN
224	"dam" N UTR DEF PL GEN
	/---/

I de ovanstående frekvensuppställningarna är inte personnamn inkluderade, eftersom morfologianalysatorn inte klassificerar dem som vare sig definitiva eller indefinita. Det framgår dock på annat vis att 5,4 % av alla substantiv i Göteborgs-Posten 1997 utgörs av förnamn i nominativ och 0,12 % av förnamn i genitiv. De förnamn vars grafiska form i nominativ är identisk med den i genitiv (t.ex. *Niklas* och *Ann-Lis*) utgör 1 % av alla substantiv i tidningstexten. Tio i topp-listan på förnamnsfronten innehåller enbart mansnamn. Det första kvinnliga förnamnet, *Anna*, kommer först på sjuttonde plats.

TABELL 24. *Förnamn.*

8867	**peter" <*c> <FIRST_NAME> <MASC> N ? SG NOM
8682	**anders" <*c> <SWE> <MASC> <FIRST_NAME> <HUMAN> N ? SG NOM/GEN
7013	**lars" <*c> <SWE> <MASC> <FIRST_NAME> <HUMAN> N ? SG NOM/GEN
5842	**magnus" <*c> <SWE> <MASC> <FIRST_NAME> <HUMAN> N ? SG NOM/GEN
5829	**jan" <*c> <SWE> <FIRST_NAME> <MASC> N ? SG NOM
5669	**göran" <*c> <SWE> <FIRST_NAME> <MASC> N ? SG NOM
5541	**stefan" <*c> <SWE> <FIRST_NAME> <MASC> N ? SG NOM
5189	**thomas" <*c> <SWE> <MASC> <FIRST_NAME> <HUMAN> N ? SG NOM/GEN
4995	**johan" <*c> <SWE> <FIRST_NAME> <MASC> N ? SG NOM
4383	**hans" <*c> <MASC> <FIRST_NAME> <HUMAN> N ? SG NOM/GEN /---/
3624	**anna" <*c> <FIRST_NAME> <FEM> N ? SG NOM /---/

På samma vis som den morfologiska analysen kan underlätta sökandet efter frekvensen på en viss böjd form i ett material kan den också få fram antalet förekomster av ord som har en viss avledd form.

4.3. Adjektiv

Det finns totalt drygt 1,77 miljoner adjektiv i tidningstexten. Den största gruppen (45,8 %) utgörs av adjektiv som står i indefinit singular nominativ. Adjektiven i plural nominativ, som har iden-

tisk definit och indefinit form, utgör 20,4 %. Den tredje största gruppen är de definitiva adjektiven i nominativ singular med 16,4 %. Adjektiven i genitivform står för en marginell del av alla adjektiv. Det kan också nämnas att 5,2 % av adjektiven står i komparativ och 3,6 % i superlativ.

TABELL 25. *Adjektiv: indefinit singular nominativ.*

11899	"stor" A UTR INDEF SG NOM
10971	"ny" A UTR INDEF SG NOM
9075	"samtidig" A NEU INDEF SG NOM
7374	"svår" A NEU INDEF SG NOM
7052	"egen" A UTR INDEF SG NOM
6314	"klar" A NEU INDEF SG NOM
6131	"stor" A NEU INDEF SG NOM
5864	"lång" A NEU INDEF SG NOM
5675	"svensk" A UTR INDEF SG NOM
5509	"ny" A NEU INDEF SG NOM
	/---/

TABELL 26. *Adjektiv: definit singular nominativ.*

12523	"ny" A UTR/NEU DEF SG NOM
11631	"svensk" A UTR/NEU DEF SG NOM
6869	"stor" A UTR/NEU DEF SG NOM
4858	"god" <SUP> A UTR/NEU DEF SG NOM
4743	"stor" <SUP> A UTR/NEU DEF SG NOM
3950	"liten" A UTR/NEU DEF SG NOM
2251	"politisk" A UTR/NEU DEF SG NOM
2108	"hög" <SUP> A UTR/NEU DEF SG NOM
2072	"internationell" A UTR/NEU DEF SG NOM
2053	"närmast" <SUP> A UTR-MASC DEF SG NOM
	/---/

TABELL 27. *Adjektiv: plural nominativ.*

13439	"ny" A UTR/NEU DEF/INDEF PL NOM
11711	"stor" A UTR/NEU DEF/INDEF PL NOM
11461	"olik" A UTR/NEU DEF/INDEF PL NOM
8211	"svensk" A UTR/NEU DEF/INDEF PL NOM
4601	"anställa" <V/DER> <PCP2> A UTR/NEU DEF/INDEF PL NOM
4314	"små" A UTR/NEU DEF/INDEF PL NOM
4308	"viss" A UTR/NEU DEF/INDEF PL NOM
4112	"egen" A UTR/NEU DEF/INDEF PL NOM
3810	"ung" A UTR/NEU DEF/INDEF PL NOM
3226	"god" <SUP> A UTR/NEU DEF/INDEF PL NOM
	/---/

5. Slutord

Sätten på vilka man kan frekvenslista ett textmaterial i elektronisk form är många, och den morfologiska analysen möjliggör ytterligare flera sorteringsätt. Med årgång 1997 av Göteborgs-Posten som grund utföll frekvensnedslagen på detta vis – om det är representativt för tidningstext eller dagens svenska i skrift överlag är svårt att avgöra, men vissa språkligt allmängiltiga tendenser kan säkert urskiljas.

Litteratur

A. ORDBÖCKER

Dansk Frekvensordbog. Baseret på danske romaner, ugeblade og aviser 1987–1990. 1992. Red. Henning Bergenholtz. København.

Hyppige Ord i Danske Aviser, Ugeblade og Fagblade. 1986. Red. B. Maegaard & H. Ruus. København.

Íslensk orðtíðnibók. 1991. Red. Jörgen Pind et al. Orðabók Háskólans.

Norsk frekvensordbok. De 10 000 vanligste ord fra norske aviser. 1982. Red. Kolbjørn Heggstad. Universitetsforlaget.

Nusvensk frekvensordbok baserad på tidningstext. Del 1–3. 1970–1975. Red. Sture Allén. Stockholm.

Nusvensk frekvensordbok baserad på tidningstext. Del 4. 1980. Red. Sture Allén et al. Stockholm.

Nynorsk frekvensordbok. Dei vanlegaste orda i skriftleg nynorsk. Hente frå aviser, sakprosa og romanar, 1978–84. 1989. Red. Per Vestbøstad. Bergen.

Svenska Akademiens ordlista över svenska språket. Tofte upplagan. 1998. Svenska Akademien.

Tiotusen i topp. Ordfrekvenser i tidningstext. 1972. Red. Sture Allén. Stockholm.

B. ANNAN LITTERATUR

- Birn, Juhani. 1998. "Swedish Constraint Grammar: A Short Presentation." Tillgänglig på <http://www.lingsoft.fi/doc/swecg/>.
- Karlsson, Fred. 1992. "SWETWOL: A Comprehensive Morphological Analyser for Swedish." *Nordic Journal of Linguistics* 15:1–45.
- Karlsson, Fred et al. 1995. *Constraint Grammar. A Language-Independent System for Parsing Unrestricted text*. Berlin and New York: Mouton de Gruyter.
- Koskenniemi, Kimmo. 1983. *Two-level Morphology. A General Computational Model for Word-form Recognition and Production*. Department of General Linguistics, University of Helsinki, Publication No. 11.

Lars S. Vikør

Purisme på hell? Nye tendensar i nynorsk ordforrådsnormering

One of Norway's two written languages, *Nynorsk*, is marked by a selective purism: it accepts internationalisms readily, but rejects many loans from Danish and Low German. This policy has been modified over time, however, and it is controversial among the users of the language. The paper discusses the nature of Nynorsk purism, the main arguments for and against, and the recent developments in the Norwegian Language Council, which is the highest formal authority in the field of Norwegian language cultivation.

Nynorsken har ord på seg for å vere ei puristisk språkform, men denne purismen er på mange måtar vag og vanskeleg å definere, og dermed òg vanskeleg å praktisere. I nordisk samanheng kan nynorsken plasserast omtrent på midten av ein skala der islandsk står på den eine enden, dansk på den andre. Svensk og norsk bokmål står nær den danske enden, finsk og samisk eit stykke lenger inn på skalaen, færøysk nærmare den islandske enden. Nynorsk ligg midt imellom fordi den er heilt open for internasjonalar, altså ord frå ikkje-germanske språk. Sperrane går mot germanismar og danismar, til dels mot anglisismar òg, men det er eit nytt og internasjonalt fenomen som eg ikkje skal ta opp her. Den andre grunnen til at purismen i nynorsk er vanskeleg, er at det er sterke indre motsetningar blant nynorskfolk om denne purismen, og at han skiftar karakter og blir svekt over tid. Normgrensene, slik dei blir praktiserte i offisielle ordlister og ordbøker, flyttar seg altså gradvis. Nynorsken eksisterer i dag i ulike undervarietetar, der skilnader i lånordreseptivitet er eitt av fleire viktige kriterium.

Purisme er eit vanleg trekk ved språk som er i etableringsfasen og kjempar mot eit dominant anna språk. Det er også vanleg at purismen minkar ettersom språket får ei tryggare og meir konsolidert stilling. Nynorsken er ein god illustrasjon på dette. Ivar Aasen hadde ei sterkt puristisk haldning; han gjekk ikkje så

langt som islendingane, men ønskte at det nye skriftspråket han forma ut, skulle byggje opp eit ordforråd av heimlege røter i all hovudsak. Lånord i dialektane ønskte han såleis å erstatte med nye ord av norske røter. Den første generasjonen av nynorskbrukarar følgde ei slik line, med generell purisme. Men det tok ikkje lang tid før internasjonalarismar begynte å komme inn og vart meir og meir aksepterte; det viste seg vanskeleg å byggje opp eit fagspråk med berre innanlandske røter. For islandsk har det gått, men ein av grunnane var nok at islandsk hadde heile samfunnsapparatet med seg i dette strevet, mens nynorsken sleit med eit dominant bokmål med ein heilt annan ordpolitikk. Ein kan seie at den *generelle* purismen i nynorsken vart gradvis avvikla i dei første tiåra av dette hundreåret, og har vore død i det minste sidan krigen. I staden fekk vi ein *selektiv* purisme; retta mot dansk-tyske lånord, og den vart verande sterk heilt fram til moderne tid, men kunne ikkje hindre at stadig fleire også av slike ord kom inn i bruksspråket og normalen.

Orda og affiksa

Dei orda det dreidde seg om, var av to typar. Dels var det enkeltord som knytte seg til dansk og urban norsk bokmålstale, men ikkje hadde komme inn i bygdemåla, og også vart aktivt motarbeidde av nynorskrørsla; det gjeld ord som dei følgjande (nynorske avløyssarar i parentes):

deilig (herleg, vedunderleg), *desto* (dess), *meget* (svært, sær), *pike* (jente), *rederi* (reiarlag), *skjensel* (skam, vanære), *skjønn* (vakker, ven, fager), *stymper* (stakkar, tufs), *vever* (spedbygd, spinkel; livleg, kvikk).

Slike ord går det ikkje an å halde ute or språket utan anten å pugge dei einskildvis, eller ved å ha det i si eiga stilkjensle at dei ikkje er akseptable. Eg skal komme tilbake til det med stilkjensla seinare.

Den andre gruppa er avleiingar med bestemte prefiks og sufiks, stort sett av tysk, særleg lågtysk, opphav. Dei viktigaste er dei følgjande, velkjende også frå dei andre skandinaviske språka: *an-*, *be-*, *bi-*, *-else*, *er-*, *for-*, *ge-*, *gjen-*, *-heit*, *hen-*, *-ige*, *-messig*, *-vesen* (sjå Omdal og Vikør 1998:132–134). Dei er lettare å kjenne att, men mange slike har vist seg uunnverlege i allmenn språkbruk. Dette er dei orda som blir kalla ”anbeheitelse”, brukt som skjellsord av puristane. Dei er det største problemet for purismen no, for dei er ei heterogen gruppe der dei ymse enkeltorda har svært ulik status i bruksspråket. Ulike miljø og delar av rørsla reagerer så ulikt på dei og avløysarane deira at det er vanskeleg her å gjennomføre ein ordbruk som ikkje anten støyter mot den offisielle normeringa, eller møter misnøye hos den eine eller andre fraksjonen i målrørsla, for å vere unorsk og utvatna på éin kant, eller for å vere stiv og kunstig på ein annan kant.

Normeringssystemet

Normeringsautoriteten på dette området ligg hos Norsk språkråd. Tidlegare – før Norsk språknemnd vart grunnlagd i 1951 – var det slik at dei enkelte ordlisteforfattarane hadde ganske stor fri-dom i val av oppslagsord i ordlistene sine, som vart autoriserte til skolebruk av departementet. Men dei måtte halde seg innanfor ein allmenn konsensus om kva som var akseptabelt – eller særleg uakseptabelt – på nynorsk. Da Staten fekk eit permanent normeringsorgan, Norsk språknemnd, var det naturleg at myndigheita til å godkjenne ordlister og lærebøker generelt vart lagd dit, og slik vart det òg da Norsk språknemnd vart erstatta av Norsk språkråd i 1972.

Systemet er altså slik at Språkrådet får alle ordlistemanus til skolebruk inn til kontroll, og tek stilling til om tvilsame ord skal tillatast som oppslagsord i ordlistene. Denne sensurordninga har vore effektiv forsåvidt som det har danna seg eit mønster for kva ord som er godkjende resp. ikkje godkjende: Når eit ord først er komme inn i ei ordliste med ”godkjent-stempel” på, er det å rekne som formelt tillate i nynorsk og kan fritt brukast vidare.

Vanlege ord i språket som ikkje står i ordlistene, og som har det kjenneteiknet at dei er komne via bokmål frå dansk eller tysk, blir rekna som "forbodne", i den forstand at dei ikkje kan brukast i tekster som Språkrådet har kontroll over, framfor alt lærebøker. I anna språkbruk er det ingen formelle hinder for å bruke dei orda ein vil, men det er ei ganske sterk uformell normering på grunnlag av ordlistene i forlag og andre tekstproduserande institusjonar. Eit ekstra verkemiddel i normeringa er dei såkalla "fornorsningsordlistene", der bokmålsord er oppslagsord og dei "gode" nynorske ekvivalentane oppførte. Den mest autoritative ordlista gjennom mange tiår var Alf Helleviks *Nynorsk ordliste*. Hellevik var òg sentral i Språknemnda og Språkrådet fram til 1980, og han følgde ei varsam midtline i normeringa, med ei gradvis opning for ord som var vanlege i folks talemål og lite markerte i språkkjensla, eller som av andre grunnar verka nødvendige, t.d. semantiske. Samtidig heldt han igjen for eit større frislepp.

Dette systemet fungerte lenge bra på den måten at det vart lite konfliktpåstand av det; nynorskfolk hadde nok av andre saker å slåst om innbyrdes, ikkje minst sjølve rettskrivingsnormalen. Men det står ikkje til å nekte at det verka tungvint for mange brukarar, fordi det var vanskeleg å finne grensene for kva som var godteke eller ikkje utan ved å slå opp i ordlista, og mange av dei forbodne orda hadde ikkje avløyssarord som vart rekna som tilfredsstillande. Denne politikken var – og er – såleis med på å gi nynorsken eit rykte for å vere vanskeleg. Det er likevel mange andre grunnar til at nynorsken gir læringsproblem, ikkje minst at mengda av nynorsk som folk les, er for lita til at det kan feste seg ei automatisert språkkjensle for dei fleste. Dei største problema har nok skoleelevar som skal skrive nynorsk utan skikkeleg bakgrunn i omfattande lesing, under trugsmål om raude strekar og dårlege karakterar. Nettopp det gjer problemet spesielt vanskeleg for nynorskrørsla, for nynorsken er under eit stadig press i skolen, både som sidemål og hovudmål.

Argumenta

Dei sentrale argumenttypane for og imot purismen kan listast opp slik:

For purisme:

- (1) Det nasjonal-historiske argumentet
- (2) Det pedagogiske argumentet
- (3) Det sosial-stilistiske argumentet
- (4) Det demokratisk-stilistiske argumentet
- (5) Det estetisk-tradisjonalistiske argumentet
- (6) Autonomiargumentet

Mot purisme:

- (1) Folkemålsargumentet
- (2) Tilnærmingsargumentet
- (3) Det internasjonale (og nordiske) argumentet
- (4) Det pedagogiske argumentet
- (5) Det vitskapleg-tekniske argumentet
- (6) Det lingvistiske argumentet

Det *nasjonal-historiske* argumentet forklarer seg vel sjølv; det er den mest internasjonalt gangbare motiveringa for all purisme. Det *pedagogiske* argumentet står i begge rekkjene, både for og imot, og er altså tvitydig: pedagogiske argument kunne brukast for purisme i det førre hundreåret da lånord og fremmendord først og fremst var eit kjenneteikn på elitespråk og gjorde skriftspråkstileigninga vanskeleg for vanlege folk. I dag er det derimot slik at dei orda purismen rettar seg mot ofte er vanlege ord i daglegspråket til folk, slik at purismen *skaper* språkkløfter i staden for å rive dei ned. Det *sosial-stilistiske* argumentet går på nynorskrørsla sin karakter av bygdekulturell motrørsle mot det urbane, eit preg som var sterkt i tidlegare fasar, men som no er nokså svekt, men enno heng litt igjen; derfor er det ein puristisk aversjon mot visse ord som er knytte til tradisjonelt urbant borgarskap, jf eksempla mine på s. 426. Det *demokratisk-stilistiske* argumentet går på at mange dansk-tyske lånord er kjenneteikn

på kansellistil, og fremmar ein substantivprega stil i staden for den verbbaserte og talemålsprega som nynorske stiltormer er baserte på. Særleg nominaliseringar på *-heit* og *-else* er blitt sterkt motarbeidde og søkt erstatta av verb- og adjektivbaserte omskrivingar. Det *estetisk-tradisjonalistiske* argumentet går på at nynorsken har utvikla ein litterær stil som har fått status som klassisk, og at det å bevare denne tradisjonen av estetiske grunnar er ein verdi i seg sjølv. *Autonomiargumentet* går på at nynorsken treng purismen for å hevde autonomien sin mot presset frå bokmålet – truleg det argumentet som i dag er mest sentralt for dei fleste som forsvarer purismen.

Motargumenta trur eg forklarar seg sjølv nokså mye. *Folke-målsargumentet* baserer seg på at purismen rettar seg mot lånord i folks vanlege talemål, og dermed gjer nynorsken meir elitistisk, tvert imot nynorskideologien, som hevdar at talemålsnærheit er ein sentral verdi. *Tilnæringsargumentet* baserer seg på eit ønske om gjensidig tilnærming mellom bokmål og nynorsk, eit syn som stod sterkt for nokre tiår sidan, men som har vike plass for autonomiargumentet; i dag er redsla for å bli slukt av bokmålet meir dominerande. Det *internasjonale* og *nordiske* argumentet er sjølv sagt pr. definisjon anti-puristisk. Det *vitskapleg-tekniske* argumentet går på at purismen trugar utviklinga av moderne teknisk og vitskapleg språk på nynorsk, i og med at ein nyordsproduksjon etter islandsk mønster er uopnåeleg i nynorsk samanheng. Problemet er at alle fagfolk i ulike vitskapar må lære faget sitt gjennom bokmål og engelsk, slik at ein nynorsk fagterminologi vil leggje ei ekstra bær på dei så lenge det ikkje finst sterke og separate nynorske fagmiljø i dei fleste vitskapar. Det *lingvistiske* argumentet, til slutt, går ut på at purismen er i strid med språkets vesen, nemleg at ordforrådet prinsipielt er ope, mens grammatikken er eit lukka system. Dette er diskutabelt, sjøvsagt, men eg tek ikkje den diskusjonen her.

Utviklinga etter 1980

Eg skal gi ei kort skisse av utviklinga i saka no i åtti- og nittiåra, da debatten om purismen for alvor blussa opp, framfor alt i Språkrådet og nynorskrørsla, og i norskundervisningsmiljøa (Lauvhjell 1983, Fretland og Vikør 1995 og Faarlund 1998 inneheld viktige innlegg). Tidleg på åttitalet vart problemet følt som så presserande at Språkrådet sette ned ei nemnd til å gjenomgå heile problematikken. Nemnda, kalla Venås-nemnda etter leiaren Kjell Venås, kom fram til ei forsiktig opning, men gav signal som mange etterpå har rekna som tvtydige (tilrådinga frå nemnda er trykt i Lauvhjell 1983:15–49 og Faarlund 1998:87–107). For det første ville nemnda ha ei oppmjuking av rettepraksisen i eksamensarbeid, slik at vanlege talemålsord med dansk-tysk rot ikkje skulle reknast som "feil". Dette vart også presisert i eit rundskriv frå departementet. I tillegg ville nemnda opne for fleire slike ord i ordlistene, blant anna gjekk ho inn for å akseptere suffikset *-heit* prinsipielt, men ho gjekk også inn for at fagnemnda i Språkrådet framleis skulle ha det siste ordet og såleis kontrollere og vurdere kvart nytt ord før det kom inn i listene. Dette vart også endskapan på drøftingane i Språkrådet, som vart avslutta i 1984.

Etter dette var det ro om saka ein del år. Fagnemnda tok for seg nokre ord meir eller mindre kvart år, og sleppte inn ein del i ordlistene. Men som det ofte viser seg: Slike mellomløysingar blir i lengda utilfredsstillande, og utpå nittitalet kom spørsmålet opp som ei stridssak på nytt.

Den utløysande årsaka denne gongen var ein ordlisteforfattar som i 1993 la fram ei lang liste av "forbodne ord" eller bokmålsord som han ville ha lov til å ta inn. Fagnemnda ønskte ikkje å gå inn på dette, så forslaga vart avviste. Men samtidig kom det opp ei ny drøfting av korleis denne grensa mellom tillatne og ikkje-tillatne ord skulle gåast opp. Fart i saka kom det først etter 1996, for det året skjedde det ei personutskifting i fagnemnda som førte til at liberaliseringsvennlege folk kom i dei bestemte posisjonane. Dei sette i gang eit utgreiingsarbeid der ein for det første prøvde å teste ut kor langt Venås-tilrådinga kunne

tolkast når det gjaldt å opne for nye anbeheitelse-ord, og for det andre å arbeide ut nye og meir romslege prinsipp for godkjenning. Fleirtalet i fagnemnda gjekk no inn for ei langt på veg fullstendig liberalisering med berre få avgrensingar. Hovudprinsippet skulle vere at alle ord som var vanlege i folks daglege talemål skulle kunne brukast i nynorsk og stå i ordlistene, med eit par morfologisk baserte avgrensingar (som tilsvarer pkt. 4a–c i regelsettet eg siterer nedanfor; sjå elles Omdal og Vikør 1998: 134–138 om denne prosessen).

På årsmøtet i Språkrådet i 1998 vart saka drøfta i nynorskseksjonen, som delte seg på midten. Dei klare tradisjonalistane var sjølv sagt imot, men også moderate tilhengarar av liberalisering tykte fagnemndfleirtalet gjekk for langt. Dei meinte at ein ut frå vanleg skrifttradisjon og skriftmålskjensle burde halde oppe eit klarare skilje mot typiske bokmålsord, i motsetning til meir umarkerte ord som var vanlege i talemålet. Problemet er at det ikkje er skikkeleg undersøkt korleis ordbruken i talemålet er. Det finst ingen normative sperrer mot visse ordtypar i talemålet, mens eit skriftmål til dels er avhengig av slike sperrer, ikkje minst eitt som skal hevde seg mot eit nært beslekta og dominerande skriftspråk.

Saka vart da utsett eit år, og til årsmøtet 1999 vart det utarbeidd eit kompromissforslag. Det vart vedteke mot berre éi røyst. Eg tek inn vedtaket in extenso her. Sidan vedtaket berre gjeld prinsippa ein skal arbeide etter, omfattar det ikkje eksempla i hakeparentes reint formelt; dei er berre meinte som forklarande tillegg.

Retningslinjer for opptak av tyske og danske importord i nynorske ordlister og ordbøker

A) *Allmenne retningslinjer*

- 1) Ord som har stor utbreiing i norske dialektar, bør ikkje haldast ute frå nynorske ordlister og ordbøker, med avgrensingane nedanfor:
- 2) Ein skal leggje vekt på om ordet alt er i vanleg bruk i skriftleg nynorsk.
- 3) Ein tek ikkje opp ord når dei er synonyme med etablerte nynorskord og ligg nær slike ord i bygnad og skrivemåte. [Døme: *åtvare*, ikkje *advare*.]

- 4) Opptak av ord i éi ordklasse treng ikkje føre til at ein tek opp tilsvarende ordstamme i andre ordklasser. [Døme: *uansett*, men ikkje *ansjå*; *betenkt*, men ikkje *betenkje*.¹]
- 5) Ein markerer ikkje etymologisk samanheng mellom importordet og andre nynorskord dersom det ikkje er vanleg å ha slik formsamheng i dialektane. [Døme: *begavelse*, ikkje *begåvelse*.²]

B) Retningslinjer for særlege ordgrupper

- 1) Ein tek ikkje opp ord med førestavingane *an-* og *be-* der nynorsk i same tyding har same ordrota utan den aktuelle førestavinga, eventuelt med etterfølgjande preposisjon. [Døme: *tvile på*, ikkje *betvile*; *nytte*, ikkje *benytte*.]
- 2) Ein tek ikkje opp verb på *-ige* og avleiingar av dei. [Døme: Ikkje *beskjef-tige*, *avskjedige*.]
- 3) Ord på *-else* kan takast opp berre når ein ikkje har synonyme ord på *-ing*, *-nad* eller utan etterstaving. [Døme: Ikkje *anerkjennelse*, *frigjørelse*, *besvarelse*.]
- 4) For etterstavinga *-heit* gjeld desse restriksjonane:
 - a) Ein tek ikkje opp *-heit*-avleiingar av perfektum partisipp, av presens partisipp og av adjektiv på *-et(e)/-ut*. [Døme: Ikkje *fordømtheit*, *klossetheit*.]
 - b) Ord på *-laus* skal ha *-løyse* som einaste avleiing.
 - c) Ord på *-sam* skal ha *-semd* som einaste avleiing.
 - d) Ord som i nynorsk har etterstavingane *-dom*, *-leik* og *-skap* i allmenn bruk, skal ikkje takast opp med *-heit*.
 - e) Ein godtek ikkje *-heit*-ord der kortavleiingar er i bruk i dialektane. [Døme: *samvit*, *rettferd*, *vanske*; ikkje *samvittigheit*, *rettferdigheit*, *vanskelegheit*.]
- 5) Ein tek ikkje opp fleire substantiv i inkjekjønning på *-ande*. [Døme: Ikkje *for godt befinnande*, *ille befinnande*.]

Dette er altså ikkje reglar for bruk av nynorsk generelt, men det er retningslinjer for fagnemnda i Språkrådet når den skal bedømme forslag til nye dansk-tyske lånord i offisielle ordbøker og ordlister. Utgangspunktet, pkt. 1, er ganske liberalt, som ein ser, men så kjem det ei rekkje avgrensingar nedanfor. Den første, om at ordet bør vere i bruk i vanleg skriftleg nynorsk, er nokså rund

¹ Partisippa *uansett* (som fungerer mest som adverb) og *betenkt* er mye vanlegare i talemålet enn verba *anse* og *betenke* (som i praksis berre er i bruk i formell stil i bokmål). Denne skilnaden i frekvens og stilverdi gjer at ein kan ta opp partisippa utan å akseptere verba.

² Rotordet heiter på nynorsk *gåve*, men på bokmål *gave*. Den importerte avleiinga *begavelse* kan i talemålet aldri ha *-å-*, og derfor tilseier dette prinsippet forma *begavelse* også i nynorsk sjølv om det blir eit formelt avvik i forhold til *gåve*.

sidan den opnar for subjektive bedøimmingar, mens dei andre retningslinene er formulerte slik at dei skal vere så konkrete som mogleg.

På hell?

Korleis dette vil verke i praksis, veit vi ikkje enno. Men ein kan seie noko om tendensen ut ifrå spørsmålet i tittelen, om den nynorske purismen no er på hell. Svaret på det må bli tvitydig.

På den eine sida er det tydeleg at den nynorske purismen har vore på vikande front i heile dette hundreåret. Ei liste over dei orda som tidlegare har vore bannlyste og no er meir eller mindre allment godtekne, ville bli ganske lang. Grunngevinga for purismen har også blitt svekt. Utgangspunktet var eit skarpt skilje mellom norsk og unorsk, der det norske var det som gjekk direkte tilbake på norrønt, det unorske det som var innlånt utanfrå gjennom tidene. I dag er det nok meir eit spørsmål om å bevare det nynorske som ein særskild språktradisjon imot det i og for seg like norske bokmålet. Men nettopp denne kjensla av ein særskild nynorsk språk- og stiltradisjon er svært sterk blant nynorskbrukarar, i alle fall den akademiske delen, men også i varierende grad breiare lag av folket i dei delane av landet der nynorsk er vanleg bruksspråk. Det er såleis eit faktum at mange i og for seg vanlege talemålsord blir unngått i nynorsk skrift på grunn av ei stilkjensle som seier at desse orda ikkje passar, dei er bokmålsord, og dermed *markerte* på same måten som f.eks. svenske og engelske ord er markerte. Dei kan i og for seg brukast, men fungerer da som sitatord. Her er likevel grensene vage, og språkkjensla nokså skiftande frå person til person.

Bruksundersøkingar når det gjeld nynorsk, har vi dessverre altfor lite av, men vi har ei stor frekvensordbok som dekkjer vanleg nynorsk skriftspråk i ei rekkje allmennspråklege sjangrar i åra rundt 1980 (Vestbøstad 1989). Helge Sandøy (i Faarlund 1998: 39–48) har undersøkt forholdet mellom ein del dansk-tyske lånord og dei næraste nynorske synonyma deira i denne frekvensordboka. Det viser seg da ein klar, jamvel ganske sterk,

tendens til å unngå lånorda, med nokre unntak, som *antype*, *beklage*, *opprinneleg*.

Undersøkinga til Sandøy omfattar òg ord som tidlegare har vore forbodne, men seinare tekne inn. Det inntrykket han gir, er at det berre er delvis samsvar mellom den offisielle normeringa og ordbruken; ein del ord som ikkje er tillatne, blir likevel mye brukte, mens andre som har sloppe inn, likevel blir unngått i praksis, gjerne fordi avløyserordet er blitt integrert i brukspråket. Avløyserorda kan også delast i ulike grupper, nemleg slike som er aksepterte og har fått nøytral status, og slike som av mange blir oppfatta som kunstige, litterære, eller overdrive "norske". Døme på det sistnemnde er (med meir vanlege eller umarkerte ord i parentes): *brennhug* eller *eldhug* (begeistring, oppgløding), *fræv* (fruktbar), *hækne* (griskheit eller -leik, grådighet eller -skap), *krimsjuke* (forkjøling), *kverrsetje* (beslagleggje, konfiskere), *mute* (bestikke), *næming* (nybegynnar, nybyrjar), *vernebuing* (beredskap). Men det er ikkje mogleg å bestemme på førehand kva slags ord som blir godtekne som nøytrale, og kva slags som blir stilistisk markerte på ein eller annan måte. Der kan vi berre seie at normerarane spår, språkbrukarane rår.

Også i framtida kjem nok utbreidde dansk-tyske lånord til å bli godtekne i stadig større grad, men kjensla av at det eksisterer særlege nynorske stilnormer som bør respekterast, vil nok også bestå, så ei full leksikalsk utjamning mellom bokmål og nynorsk vil vi ikkje få. Men det kan vel gå i den retninga at klokkartrua på leksikalsk *normering* som universalmiddelet for å bevare det nynorske ordforrådet blir stadig meir svekt. Det er eit stort behov for å utvikle andre og meir positive verkemiddel i traderinga av dei nynorske stilnormene.

Litteratur

- Faarlund, Jan Terje (red.). 1998. *Ordtilfanget i nynorsk – synsmåtar og røyntler. (Norsk språkråds skrifter. 5.)*
- Fretland, Jan Olav og Lars S. Vikør (red.). 1995. *Korleis bør nynorsken sjå ut? Om Norsk språkråd og normeringa av nynorsken.* Noregs Mållag.
- Lauvhjell, Arne (red.). 1983. *Heit strid om nynorsk. Dokument og meiningar 1980–83.* Oslo: Novus Forlag.
- Omdal, Helge og Lars S. Vikør. 1998. *Språknormer i Norge. Normeringsproblematikk i bokmål og nynorsk. (LNU's skriftserie. 100.)* Oslo: Cappelen Akademisk Forlag. [1. utg. 1996.]
- Vestbøstad, Per. 1989. *Nynorsk frekvensordbok. Dei vanlegaste orda i skriftleg nynorsk henta frå aviser, sakprosa og romanar 1978–84.* Bergen: Alma Mater.

Matti Vilppula

Är eller var – tempusproblem i ordboksdefinitioner

In definitions of concepts that no more belong to the present time, dictionaries have sometimes used the present, sometimes the past tense. For instance, *highwayman* may have been defined as 'one who *robs* travellers on a public road' or 'one who *robbed* travellers on a public road'. Of these two alternatives, the first one should be preferred. If the past tense is used, the definition gets an encyclopedic feature. It suggests that sometime earlier people were often robbed on the highway. This matter does not belong to the lexical meaning of the word.

Inledning

”Är eller var?” Frågan togs upp till diskussion av en av redaktörerna vid vår ordbok förra hösten. Hon hade hållit på med en betydelseförklaring och börjat fundera över om predikatet skulle stå i presens eller imperfekt. Det gällde ett ord i stil med *handkvarn*. Borde man alltså säga: ”kvarn som drivs med handkraft” eller ”kvarn som drevs med handkraft”?

Vi arbetar med en omfattande ordbok över de finska dialekterna (*Suomen murteiden sanakirja* 1985–), en ordbok i 20 delar, av vilka fem hittills har utkommit. Ett verk av detta slag innehåller rikligt med etnologiska termer, t.ex. det nyssnämnda *handkvarn*, och termerna syftar på företeelser som börjar vara föråldrade och som en modern människa knappast känner till. I de delar av vår ordbok som redan kommit ut står därför imperfekt mycket ofta som tempus i definitionerna vid sidan av presens.

Jag skall i mitt föredrag behandla de problem som valet mellan presens och imperfekt ger upphov till. Då jag som exempel vill anföra endast autentiska definitioner, inte översättningar, kan jag inte utnyttja definitionerna i vår egen ordbok eller andra finska ordböcker utan har i stället valt exempel ur svenska ordböcker. Jag vill dock framhålla att tempusvalet på intet sätt är något

specifikt svenskt problem. Tvärtom är problemet större på finskt håll redan av den orsaken att man i svenskan ofta kan kringgå det genom att använda olika prepositioner eller "att + infinitiv". Sålunda kan *lie* definieras som 'redskap för slående av gräs', *brandvakt* som 'vakt efter eldsvåda', *taknäver* som 'näver till taktäckning' och *sopborste* som 'borste att sopa upp med'. I finskan är man däremot ofta tvungen att använda relativsats och personform av verb, till exempel: *lie* 'redskap med vilket man slår gräs'.

Det är skäl att framhålla att valet av tempus i definitionerna är ett problem endast när det gäller substantiv, inte ord ur andra ordklasser.

Praxis växlar

Begreppet bakom ordet *kåk* ('skampåle') hör helt klart till historien. Detta är väl orsaken till att två nusvenska ordböcker – *Illustrerad svensk ordbok* och *Svensk ordbok* – bägge använder imperfekt i definitionen:

- (1a) *kåk* hist.: skampåle vid vilken vissa bestraffningar ä g d e rum i allmänhetens åsyn
(*Illustrerad svensk ordbok* 1977)
- (1b) *kåk* 4. typ av skampåle vid vilken bestraffningar ä g d e rum i allmänhetens åsyn (hist.)
(*Svensk ordbok* 1986)

Om vi sedan tar en titt på hur samma ordböcker definierar ordet *skampåle*, som förekommer i de ovannämnda definitionerna, märker vi att ordböckerna valt olika tempusformer. *Illustrerad svensk ordbok* använder imperfekt, *Svensk ordbok* presens:

- (2a) *skampåle* åld.: påle vid vilken förbrytare b a n d s i halsjärn
(*Illustrerad svensk ordbok* 1977)
- (2b) *skampåle* påle på offentlig plats, vid vilken förbrytare b i n d s fast (hist.)
(*Svensk ordbok* 1986)

Lika stor är skillnaden mellan ordböckerna i valet av tempus när det gäller orden *isperiod* eller *istid*:

- (3a) *isperiod* geol.: tid då inlandsisen t ä c k t e väldiga områden, i sht den period då isen t ä c k t e hela norra delen av Europa o. en stor del av Nordamerika (den kvartära isperioden)
(*Illustrerad svensk ordbok* 1977)
- (3b) *istid* geologisk period under vilken stora delar av kontinenten t ä c k s av is *särsk. om den senaste*
(*Svensk ordbok* 1986)

På det hela taget är imperfekt ganska ovanligt i definitionerna i ordböcker över modernt språk. Ordböckerna innehåller ju inte så värst många ord för begrepp som hänför sig enbart till en svunnen tid. Det är annorlunda med dialektordböckerna – i dem är det tämligen lätt att hitta imperfektformer i definitionerna.

Liksom ordböckerna över modernt språk företer också dialektordböckerna olikheter när det gäller tempus i definitioner av ett och samma ord. Dalmålsordboken använder imperfekt i definitionen av ordet *gryt-krok* medan den estlandssvenska dialektordboken använder presens, och när det gäller ordet *arbetspiga* använder den rikssvenska dialektordboken presens och den finlandssvenska imperfekt:

- (4a) *gryt/krok* 1. (med fingerögla försedd) järnkrok, medelst vilken gryta el. kittel l y f t e s av elden
(*Ordbok över folkmålen i övre Dalarna* 1970–)
- (4b) *gryt/krok* som grytan h ä n g e r på; var i äldre tider av trä; är ställbar, så att grytan kan lyftas el. sänkas
(*Ordbok över Nuckömålet* 1951)
- (5a) *arbets/piga* piga som d e l t a r i jordbruksarbete
(*Ordbok över Sveriges dialekter* 1991–)
- (5b) *arbets/piga* piga som d e l t o g i utarbeten
(*Ordbok över Finlands svenska folkmål* 1979–)

Det är förvånande att dialektordböckerna ibland använder imperfekt också i definitioner av ord som åsyftar ett fortfarande aktuellt begrepp och som därför i ordböcker över modernt språk

definieras med presens. Orsaken kan vara att det i dialektordböckerna uttryckligen är fråga om en svunnen tid och inte om de moderna motsvarigheterna till begreppen. Till exempel:

- (6a) *brudsäng* l. säng vari brudparet s o v under bröllopsnatten
(*Ordbok över Finlands svenska folkmål* 1979–)
- (7a) *brödsкура* brödspade med långt skaft, varmed bröden s a t t e s in i ugnen
(*Ordbok över Laumålet på Gotland* 1972–1986)
- (8a) *dags/resa* väglängd, som k u n d e tillryggaläggas på en dag
(*Ordbok över folkmålen i övre Dalarna* 1970–)

Svensk ordbok (1986) använder presens i definitionerna av samma ord eller deras skriftspråkliga motsvarigheter:

- (6b) *brudsäng* säng där brudparet t i l l b r i n g a r bröllopsnatten *särsk. med tanke på fullbordandet av äktenskapet*
- (7b) *bröd/spade* platt, bred spade med långt skaft varmed bröd f ö r s in i ugn för gräddning *vanl. av tunt trä*
- (8b) *dags/resa* sträcka som k a n tillryggaläggas under en dag *med hjälp av ngt fordon*

Det finns ordgrupper inom vilka dialektordböckerna tämligen enhälligt använder presens i definitionerna. En sådan grupp är skällsorden:

- (9a) *apl/hätta* bild., person som a p a r efter
(*Ordbok över Sveriges dialekter* 1991–)
- (9b) *blind/finska* okv. för kvinna som h a r svårt att hitta det hon l e t a r efter
(*Ordbok över Finlands svenska folkmål* 1979–)
- (9c) *bröd/Pål* gosse som ä t e r mycket bröd
(*Ordbok över Laumålet på Gotland* 1972–1986)
- (9d) *lögn/racka* person, som s p r i n g e r med lögner
(*Ordbok över folkmålen i övre Dalarna* 1970–)

Det är svårt att tänka sig imperfekt i stället för presens i dessa fall. Det kan bero på att uttrycken är allmängiltiga till sin natur. De begrepp som de står för är inte knutna till någon speciell tid, utan är ständigt aktuella. Det har alltid funnits och kommer alltid att finnas människor som härmar andra och människor som är tank-spridda, glupska eller skvalleraktiga.

Det finns dock ingen regel utan undantag. Det gotländska ordet *bröd-Pål* (9c) har i Dalarna två synonymer för vilka imperfekt använts i definitionerna:

- (9e) *brödkonung* person, som å t mycket bröd
- (9f) *brödkvarn* person, som å t mycket bröd
(*Ordbok över folkmålen i övre Dalarna 1970-*)

Även i övrigt använder *Ordbok över folkmålen i övre Dalarna* oftare imperfekt i definitionerna, dock inte genomgående, vilket framgår av exempel (9d).

De exempel som hittills tagits upp har främst gällt fall där olika ordböcker använt olika tempus i definitionen av samma ord. Bruket kan emellertid växla också inom en och samma ordbok, där olika tempus ibland används när det gäller synonymer eller ord som hör till samma begreppsfär:

- (10a) *riv/borr* borr varmed man b o r r a d e hål för räfständerna
- (10b) *riv/bågl/borr* borr varmed man b o r r a r hål för "rivbågarna", smalare än "riv-borr"
(*Ordbok över Laumålet på Gotland 1972-1986*)
- (11a) *skam/bjälla* skälla, som s a t t e s på ko, som inte fö l j d e hjorden hem om kvällarna
- (11b) *skam/ko* ko, som e j f ö l j e r med hjorden hem om kvällarna
(*Ordbok över folkmålen i övre Dalarna 1970-*)
- (12a) *annandags/bajen* (spefullt) den som b l e v sist att stiga upp o. i underkläderna o. barfota springa (t.ex. tre varv runt stugan) tidigt på morgonen annandag påsk
- (12b) *annandags/låskan* (spefullt) den som s t i g e r sist upp annandag påsk
(*Ordbok över Sveriges dialekter 1991-*)

SAOB och *rackare*

Det är möjligt att den inkonsekvens jag tycker mig se i användningen av presens och imperfekt endast är skenbar. Kanske tempusanvändningen är genomtänkt och baserar sig på ordböckernas egna redigeringsprinciper. Kanske finns det en logik i växlingen mellan imperfekt och presens, fastän en utomstående har svårt att uppfatta den.

Det följande exemplet är taget ur *Svenska akademiens ordbok* (SAOB). Det gäller ordet *rackare* och definitionerna inom de två första betydelsemomenten. Korstecknet i början av det första momentet anger att det är fråga om en utdöd betydelse, som nysvenskan inte längre känner. Tempusformen är presens. I det andra momentet förhåller det sig annorlunda. Detta moment beskriver en levande betydelse men ett sakförhållande som är föråldrat. Här används imperfekt.

- (13) *rackare* I. (†) person som h a r till yrke I. sysselsättning att bortskafta orenlighet o. d. (från gator I. avträden o. d.); äv. om skorstensfejare.
 2. (om. ä. förh.) person som h a d e till yrke I. sysselsättning att slakta o. flå I. kastrera hästar (I. avliva hundar o. d.) o. att undanskaffa as (jfr I); hudavdragare; äv.: hantlangare åt bödel, bödelsdräng; stundom äv. om bödel; förr med mycket starkt nedsättande innebörd.

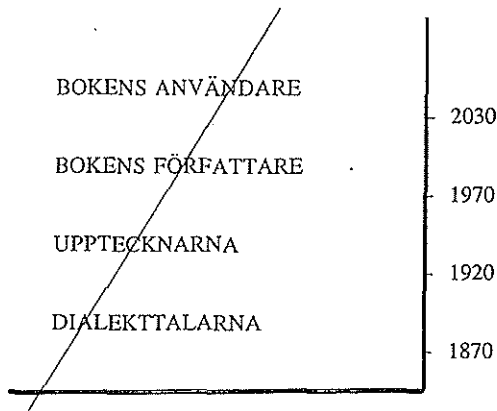
Enligt uppgift från SAOB:s redaktionssekreterare Carl-Erik Lundblad (19.1.1999) baserar sig olikheten i tempus i det ovanstående fallet helt på redigeringsprinciperna för SAOB. Presens används i definitionerna utom vid bruklighetsbeteckningarna "förr" och "om ä. förh.". Efter korstecknet används alltså presens.

Ordbok över de finska dialekterna och tempusvalet

När det gäller användningen av presens och imperfekt i vår finska dialektordbok kan jag säga att praxis har växlat. I våra redigeringsanvisningar nämns ingenting om saken, och redaktö-

rerna har i stor utsträckning själva fått avgöra vilket tempus som passat. Det är klart att redaktörerna också blivit påverkade av upptecknarnas val av tempusform på de ordsedlar som ligger till grund för ordboken.

Slumpmässigheten i valet av tempusformer har börjat besvära oss så mycket att vi beslutat lösa problemet genom att använda endast ett tempus, presens. Jag skall här försöka förklara vilka omständigheter som påverkat vårt beslut.



Detta schema visar de olika skedena i vår ordboks historia. Det är meningen att ordboken skall beskriva det språk som talades på den finska landsbygden i slutet av 1800-talet och början av 1900-talet. Man kan säga att boken är synkronisk såtillvida att den täcker en viss period i de finska dialekternas historia. Materialet, 8 miljoner ordsedlar, insamlades huvudsakligen mellan åren 1920 och 1970. Det insamlades i sin helhet på fältet, enkom för ordbokens behov. Vi har alltså nästan inget material ur skriftliga källor. När insamlingen avslutats började arbetet med manuskriptet, och den första delen av ordboken kom ut år 1985. Den tjugonde och sista delen beräknas bli färdig på 2030-talet.

Snedstreckat på schemat föreställer gränsen mellan presens och imperfekt. Presens är till höger om strecket, imperfekt till vänster. Tiden kring sekelskiftet, som vår ordbok beskriver, hör

nästan helt till presenskategori. De begrepp som människor hade ord för var här och nu. På 1920-talet, när insamlingen av material kom igång, kunde man skönja de första tecknen på att jordbruket började förändras och mekaniseras. Många begrepp som något tiotal år tidigare hade hört till vardagen var nu ett minne blott. De hade förflyttats från presenskategori till imperfektkategori.

Förändringen fortsätter. Vi som skriver ordboken lever hundra år efter den tid som vår ordbok handlar om. För oss hör begrepp som ljuskäring, slända eller slaga utan vidare till imperfektkategori. Ju mera tiden går desto fler blir de begrepp och därmed ord som övergår till imperfektkategori efter att ursprungligen ha hört till presenskategori.

Vi har aldrig funderat på ur vems synvinkel vi beskriver ett ords betydelse. Gör vi det ur dialekttalarens synvinkel så att vi alltid använder presens? Identifierar vi oss med upptecknarna och använder det tempus som de använt? Använder vi presens och imperfekt där det känns naturligt för oss själva? Eller försöker vi ta hänsyn till de framtida användarna av vår ordbok och fundera på vilka begrepp i presenskategori som under de följande årtiondena eller århundradena kommer att förflyttas till imperfektkategori?

Problemet gäller alla dialektordböcker och historiska ordböcker, överhuvudtaget ordböcker som skall vara tidlösa och kunna användas i århundraden. Det är annorlunda med ordböcker över modernt språk, som används på sin höjd några årtionden. De beskriver samma språk som författarna och användarna talar och samma språkstadium som de lever i. Tempus i definitionerna är inget problem för sådana ordböcker; det är nästan alltid presens.

När det gäller evighetsordböcker som den finska dialektordboken beror tempus i definitionerna på valet av synvinkel. Vi har beslutat välja den tidigaste synvinkeln, dialekttalarnas synvinkel, därför att vi därigenom befrias från att använda två olika tempus vilkas inbördes styrkeförhållande ständigt skiftar.

Andra skäl för användningen av presens

Det finns också andra skäl att använda presens. Huruvida begreppet bakom ett ord kommit ur bruk eller inte är inte relevant när det gäller ordets lexikaliska betydelse. Det är encyklopedisk kunskap. Visserligen kan en definition i en ordbok, i synnerhet i en ordbok som anknyter till folkkulturen, väl knappast någonsin vara en ren nominaldefinition helt utan encyklopedisk information, men nominaldefinitionen är ändå det ideal vi försöker uppnå. Sålunda sägs det till exempel i företalet till den tyska *Handwörterbuch der deutschen Gegenwartssprache* (1984:XII) att betydelsedefinitionerna i boken i komprimerad form skall uttrycka det väsentliga och att man såvitt möjligt försökt undvika encyklopediska definitioner, omfattande sakbeskrivningar. ("Die Bedeutungserklärung soll in komprimierter Form das Wesentliche ausdrücken, enzyklopädische Definitionen, d. h. extensive Sachbeschreibungen, werden möglichst vermieden.")

Imperfekt är på sin plats i en definition då det anses nödvändigt att ge encyklopedisk information. I den följande definitionen av ordet *tomte* har imperfektformen *troddes* ingenting att göra med ordets lexikaliska betydelse. Dess uppgift är att förmedla encyklopedisk kunskap, dvs. upplysa om att *tomte* är ett ord som hör till folktron.

- (14a) *tomte* 2. ett väsen i nordisk folktro som *troddes* hålla till på bondgårdar och där sörja för folkets trevnad och välgång
(*Svensk ordbok* 1986)

Man kan emellertid fråga sig om imperfekt är nödvändigt ens i detta fall. En annan ordbok använder nämligen formen *tros* i stället för *troddes* i sin definition av samma ord:

- (14b) *tomte* enl. nordisk folktro ett väsen som *tros* bo på gården o. vaka över dess lycka o. trivsel
(*Illustrerad svensk ordbok* 1977)

Det finns säkert också ord vilkas lexikaliska betydelse förutsätter imperfekt. Sådana är t.ex. de substantiv som avser en förfluten tid

eller tidsperiod. I en definition av ordet *stenålder* kan imperfekt inte ersättas med presens:

- (15) *stenålder* en förhistorisk tidsepok då vapen och verktyg huvudsakligen tillverkades av sten i Norden fram till ca 1500 f. Kr.
(Svensk ordbok 1986)

Påpekas bör dock att det här är fråga om ett undantagsfall. Ordet står mycket nära ett egennamn, vilket framgår bl.a. av att det brukar användas endast i singularis och att *ålder* vanligen står i bestämd form (*åldern*). Däremot kan ordet *istid*, som det tidigare var tal om, förekomma också i pluralis.

I skriftspråksordböcker är det vanligt att en föråldrad betydelse eller ett föråldrat begrepp betecknas med *arkeol.*, *hist.*, *förr.*, *föråldr.*, *om ä. förh.*, *åld.* I de betydelsebeskrivningar som följer efter sådana beteckningar används ofta imperfekt. Man kan fråga sig varför. Uttrycker inte imperfekt detsamma som beteckningen? Blir det inte redundant information?

I dialektordböcker förekommer knappast beteckningar som anger att ett uttryck är föråldrat. De behövs helt enkelt inte, eftersom det ligger i dialektordbokens natur att beskriva en språkform som inte längre används. Läsaren är medveten om detta. Han vet att allt som finns i boken hör till en svunnen tid. Han vet också att den presensform som konsekvent används i bokens definitioner egentligen borde vara imperfekt ur hans synvinkel sett.

Perfekt och pluskvamperfekt i definitioner

Utöver presens och imperfekt förekommer även perfekt i definitionerna, men detta tempus är inte lika problematiskt som imperfekt. Perfekt står närmare presens såtillvida att där imperfekt uttrycker avslutad verksamhet i förfluten tid uttrycker perfekt verksamhet som påbörjats i det förflutna men kan ha fortsatt till nuet. I definitionerna skiljer sig perfekt från imperfekt också genom att perfekt inte nödvändigtvis ger encyklopedisk information. I de följande exemplen utgör perfekt en väsentlig del av betydelsen:

- (16a) *abc/stabbe* barn som inte har lärt sig läsa
(Ordbok över Finlands svenska folkmål 1979–)
- (16b) *sur/mjölk* mjölk som har blivit sur (på gr. av
inverkan från mjölksyrebakterier)
(Illustrerad svensk ordbok 1977)

Det är fråga om processer som framskridit till sitt slut. Man kan kalla någon *abc-stabbe* först när hans klena begåvning blivit uppenbar. Mjölk är *surmjölk* först när den surnat. Åtminstone i det senare fallet kan presens inte komma i fråga.

I svenskan kan hjälpverbet i en bisats utelämnas, vilket inte går för sig i finskan. Av denna orsak kan verbformerna i exemplen

- (17a) *april/bock* person som lurats av aprilskämt
(Ordbok över Sveriges dialekter 1991–)
- (17b) *botten/stock* stock som sjunkit till botten
(Ordbok över Finlands svenska folkmål 1979–)

tolkas på två olika sätt: antingen som perfekt ("person som har lurats av aprilskämt", "stock som har sjunkit till botten") eller som pluskvamperfekt ("person som hade lurats av aprilskämt", "stock som hade sjunkit till botten"). Av dessa två alternativ måste perfekt anses vara bättre. Pluskvamperfekt innebär ett encyklopediskt perspektiv; det anger att seden att kalla någon *april-bock* varit i bruk förr i världen och att en stock som sjunkit till botten förr kallats *botten-stock*.

Imperfekt och pluskvamperfekt gör alltså en definition encyklopedisk. När avsikten är att beskriva en betydelse, inte ett sakförhållande, är den rätta tempusformen antingen presens eller perfekt, allt efter behov.

Litteratur

Handwörterbuch der deutschen Gegenwartsprache: in zwei Bänden. Von einem Autorenkollektiv unter der Leitung von Günther Kempcke. 1984. Berlin Akademie-Verlag.

Illustrerad svensk ordbok. 3 rev. uppl. 1977. Stockholm.

Ordbok över Finlands svenska folkmål. 1979–. Red. Olav Ahlbäck †, Peter Slotte et al. Helsingfors.

Ordbok över folkmålen i övre Dalarna. 1970–. Red. Lars Levander † och Stig Bjöklund. (*Skrifter utgivna genom Landsmåls- och folminnesarkivet i Uppsala. Ser. D:1.*) Uppsala.

Ordbok över Laumålet på Gotland. 1972–1986. Red. Mathias Klintberg † och Herbert Gustavson. (*Skrifter utgivna genom Landsmåls- och folkminnesarkivet i Uppsala. Ser. D:2.*) Uppsala.

Ordbok över Nuckömålet. 1951. Red. Gideon Danell. (*Skrifter utg. av Kungl. Gustav Adolfs akademien. 27.*) Uppsala.

Ordbok över Sveriges dialekter. 1991–. Red. Vidar Reinhammar, Gunnar Nyström et al. Uppsala: Arkivet för ordbok över Sveriges dialekter.

SAOB = *Ordbok över svenska språket, utgiven av Svenska Akademien.* (Svenska Akademiens ordbok.) 1898–. Lund.

Suomen murteiden sanakirja. Kotimaisten kielten tutkimuskeskuksen julkaisuja 36. 1985–. Helsinki.

Svensk ordbok. 1986. Utarbetad vid Språkdata, Göteborgs universitet. Stockholm: Esselte Studium.

Översättning från finska av Monica Äikäs.

Jonas Winnerlöv

Länk och ekvivalens i en mångspråkig lexikografisk relationsdatabas

The paper presents a lexicographic system developed in a standard PC software and based on a multilingual relational database with a structure optimized for the electronic age. Equivalence is created on two independent levels, headword and example, on the basis of relational links between tables for selected pairs of languages. An additional level consisting of analytic bidirectional articles has been created in order to handle culturally specific words and other categories, corresponding roughly to 5 percent of the lemmas, where equivalence cannot be obtained on the two first levels. These three modules are common for a variety of electronic and printed products, for instance a compact pocket dictionary, using a combination of first-level equivalence in simple word lists and analytic articles.

Den snabba utvecklingen inom IT har skapat nya möjligheter för lexikografin. I lätt sökbara textkorpusar med i det närmaste obegränsade datamängder har lexikografen fått en ovärderlig råvara för sitt arbete.

En annan typ av revolution är att standardprogram för PC blivit så kraftfulla att de kan utgöra basen för fullvärdiga lexikografiska system. Samtidigt har dessa blivit så enkla och billiga att språkvetaren kan ha verktyget i egen hand och själv bygga upp en datastruktur från grunden, utan att behöva gå omvägen via dataexperter.

Syftet med denna presentation är att ge en översiktlig beskrivning och analys av ett lexikografiskt system som jag har utvecklat i en relationsdatabas, där de datatekniska länkarna utgör grunden för den språkliga ekvivalensen. Istället för att i detalj gå in på de systemtekniska lösningarna vill jag ge exempel på hur lexikografiska arbetssätt kan utvecklas och på hur nya slags produkter kan skapas effektivt med hjälp av olika kombinationer av ekvivalensnivåerna.

Idén till detta system föddes då jag arbetade som projektledare för utveckling av marknadssystem på Bonniers Affärsinformation

AB, och projektet sjuöskattes i mitt eget företag Scriptum International AB i Moskva 1994 tillsammans med en grupp ryska lingvister och studenter. Systemet (nedan SI-basen) används nu huvudsakligen i mitt eget akademiska arbete.

Databasen i centrum

De flesta av dagens lexikografiska system har skapats specifikt för att hantera de traditionella produktserierna av tryckta ordböcker i olika storlekar och språkkombinationer. När elektroniska produkter som CD-rom skapas av ett sådant material blir de i princip snabbsökta böcker.

Om databasen sätts i centrum finns möjlighet att tänka om. Istället för att göra elektroniska produkter av tryckta böcker eftersträvas en generell grundstruktur på datamängden som direkt kan formas till olika slags elektroniska och tryckta produkter, utan omfattande bearbetning och specialprogrammering.

Det är rimligt att anta att stora ordboksprodukter successivt kommer att bli elektroniska. När kvalificerad information efterfrågas, som vid översättning och annan språklig analys, sitter användaren sannolikt redan vid en dator. Om databasens möjligheter utnyttjas fullt ut kommer den elektroniska produktens fördelar att framstå som ännu tydligare jämfört med den tryckta boken.

Framtidens tryckta produkter kommer rimligtvis att vara mindre och riktade till en viss målgrupp och för vissa situationer, t.ex. läsning i en bekväm fåtölj, på resan, under en lektion eller ett möte.

Relationsdatabasens principer

Den typ av databas som använts här är av relationstyp, liksom de allra flesta moderna system. SI-systemets programvara är idag MS Access, som tillsammans med MS Word och MS Excel ingår i Microsofts Officepaket, alltså en standardprodukt som idag

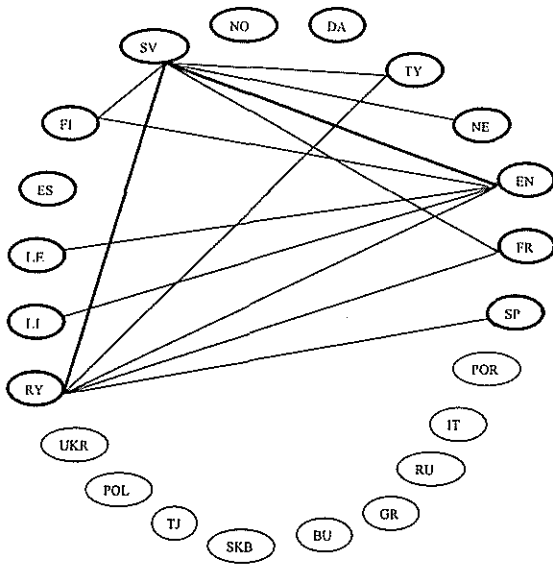
finns installerade på de flesta PC. Grundprincipen för en relationsdatabas är att information inte ska förekomma flera gånger. Istället byggs relationer eller länkar upp mellan dataenheterna.

Typfallet för att illustrera hur en relationsdatabas är uppbyggd brukar vara ett adressregister. I sin enklaste form är ett adressregister en vanlig anteckningsbok, i sin näst enklast sorterbara lappar i ett kartotek, inte helt olikt ett välkänt lexikografiskt arbetssätt. Om adresserna skrivs in i ett vanligt ordbehandlingsprogram tillkommer redigerings- och sökmöjligheterna. Nästa steg kan vara att använda ett registerprogram eller en så kallad enkel databas, där varje kolumn går att sortera på, t. ex förnamn, efternamn, företagsnamn, gatuadress, postnummer, postadress, telnr, stad, land. Här ser man dock att en stor mängd information upprepas. I en relationsdatabas skulle man därför lägga t.ex. organisationsnamn i en separat fil, för att från en given post, t.ex. Göteborgs Universitet, kunna länka till olika personer i en särskild personfil, där varje person (i idealfallet) endast förekommer en gång, oavsett om det handlar om en jobb- eller hemadress. Istället för att skriva ut Göteborg i varje postadress räcker det med postnummer som därifrån kopplar till en ortsfil där stadsnamnet Göteborg endast förekommer en gång, tillsammans med dess riktnummer etc, som på så sätt kan ändras på ett enda ställe. (Göteborg som del av ett organisationsnamn förekommer naturligtvis flera gånger.) Denna typ av lagringsstruktur är platsbesparande, men idag är det antal tecken som kan lagras i stort sett obegränsat även på PC. Det avgörande är att koppla ihop information som det kan finnas behov av att uppdatera eller analysera. Det finns alltså ingen anledning att i ett adressregister koppla ihop alla som heter Mats eller Malmgren, eftersom namnar knappast har något gemensamt. I ett språkligt register är det dock just ordformer av detta slag som ska kopplas ihop.

Mångspråkig databas och tvåspråkiga produkter

För en svensk förläggare som producerar för en svensk målgrupp är det naturligt att ha en svensk ordbas i "navet", och sedan, kan-

ske genom att köpa kopior på andra nationella förläggares ord-baser, bygga upp flera språk som kopplar mot svenskan men inte sinsemellan. SI-basen är däremot uppbyggd som ett språkneutralt nätverk, där alla ingående språk kan kopplas till varandra. När en språkfil byggs upp och förädlas kan det komma en hel rad av produkter till godo, vilket är särskilt viktigt när det gäller nischspråk, där ytterst få produkter annars kan bli lönsamma. I idealfallet sker arbetet direkt mot en central datamängd i ett lokalt nätverk. Om arbete sker mot en kopia på annat håll måste data sedan integreras i den centrala databasen.



FIGUR 1. Figuren visar de europeiska språk som kan komma ifråga. De feta ringarna är de språk som idag är helt utbyggda (svenska, ryska, engelska, tyska och franska) och delvis (danska, norska, finska, estniska, lettiska, litauiska, nederländska och spanska). De länkar som existerar är markerade med streck, varav de feta är unikt förädlade. Varje språk innehåller två oberoende moduler, uppslagsord och exempel.

Även om SI-basen är mångspråkig är slutprodukterna tänkta att i huvudsak vara tvåspråkiga. I de flesta fall har man med ett språkpar i taget att göra, varav det ena är användarens modersmål. En ambition med SI-systemet är just möjligheten att utveckla tvåspråkiga produkter även för mer marginella språkkombinationer, där många användare idag får gå omvägen via t.ex. engelska. Bergenholtz & Svensén (1994) påpekar ”att en flerspråkig allmän ordbok med nämnvärda ambitioner knappast är möjlig”. Terminologiska ordböcker lämpar sig däremot bättre för ett flerspråkigt format, även om sökningen i de flesta fall sker från ett språk till ett annat.

Länk och ekvivalens

I den traditionella ordboken uttrycks varje lemmas ekvivalenter i en artikelstruktur där lexikografen omsorgsfullt väger ihop ekvivalenter på ordnivå och exempelnivå tillsammans med olika slags kommentarer. I den språkliga relationsdatabasen utnyttjas den datatekniska länken för att skapa språklig ekvivalens på olika av varandra oberoende nivåer. Det har visat sig att tre nivåer kan ligga till grund för en mängd lexikografiska produkter:

- (1) *uppslagsord*, där ekvivalens skapas på ordnivå;
- (2) *exempel*, vilket motsvarar ekvivalenser på fras-, sats- och makrosyntagmnivå;
- (3) *artikel*, som utnyttjar textdimensionen, kallas analytisk och skiljer sig i flera avseenden från den traditionella ordbokens artikel.

Datastrukturen

Grundprincipen för datastrukturen är att information av en viss typ, vilket här kallas *grundenhet*, återfinnes i sin egen tabell eller *fil*. I filerna utgör varje grundenhet, t.ex. ett uppslagsord, en *post*.

Varje post innehåller ett antal *fält* med information som intimt hänger ihop med grundenheten. Varje fil kan sägas utgöra en enkel databas. Relationsdatabasen uppstår genom *länkarna* eller *länkfilerna*, vilka sammanbinder filerna genom att unika id-nr från grundenheternas poster paras ihop, tillsammans med ett antal fält som intimt hänger ihop med själva länken.

Uppslagsorden

Den mest centrala grundenheten i systemet är uppslagsordet, vilket i stort motsvarar den traditionella ordbokens lemma. De filer som för varje språk innehåller denna information liknar i sin struktur *Nationalencyklopedins ordbok* (NEO) på CD-rom, där de svenska uppslagsorden är grundenheterna, och fälten innehåller uppgifter om t.ex. ordklass, böjningsändelser, stilvärde, uttal, etymologi. En stor skillnad är att NEO dessutom innehåller definitioner (vilket hittills endast förekommer i undantagsfall i SI-databasen) och utförliga språkprov på samma nivå.

I NEO utgör homografer skilda lemman, vilket de för tillfället inte gör i SI-basen. Mycket arbete kan sparas på att undvika de svåra avgränsningarna mellan homonymi och polysemi. En annan fördel är att sökningar och matchningar underlättas om det är ordformen som utgör grundenheten.

I denna typ av relationsdatabas för allmänna tvåspråkiga produkter är varje språkfil alltså uppbyggd enligt en semasiologisk princip, den utgår från det språkliga tecknet (Svensén 1987). Många flerspråkiga terminologidatabaser, t.ex. Eurodicautom, är däremot uppbyggda enligt en onomasiologisk princip, där ett fast definierat begrepp utgör grundenheten och därför med fördel kan konstrueras som en enkel databas, där översättningen till varje språk får sina egna fält.

En databas för det allmänna ordförrådet enligt onomasiologisk princip skulle ha semem som grundenheter, vilket görs i t.ex. EuroWordNet (Alonge 1998) och dess svenska pendang SWordNet. Att analysera ord på detta sätt är ett mycket krävande arbete,

och upplägget skulle bli tämligen oekonomiskt för tvåspråkig lexikografi.

Att låta mindre språkliga byggstenar än lemman, t.ex. de lexikala morfemen, utgöra grundenheterna, har visat sig ogörligt. Däremot kan centrala lexikala morfem utgöra grundenheter i en annan typ av fil, till vilka uppslagsorden länkas, och på så sätt bilda etymologiska familjer. Ett sådant upplägg har enligt NLO (Bergenholtz & Haraldsson 1997) ”en särskild pedagogisk verdi”. Även graford med frekventa och avvikande ordformer förekommer i särskilda stödfiler, men dessa är självfallet olämpliga som grundenheter.

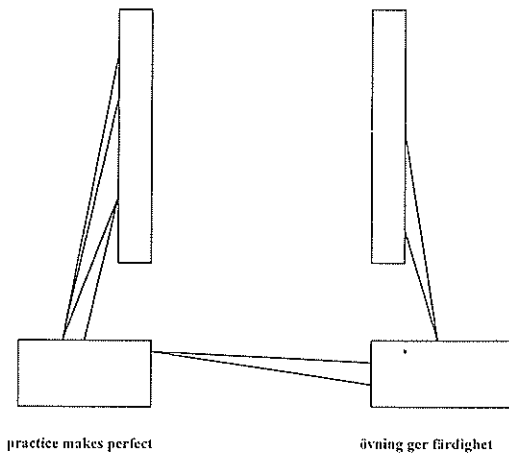
Grundfilerna framställs genom sammanläggningar av varandra oberoende källor, en och tvåspråkiga ordböcker och ordbaser, frekvensundersökningar, läroböcker, pedagogiska ordlistor och annat material som innehåller för inläraren relevanta ord. Antalet källor som ett ord förekommer i och kombinationer av dessa utgör en mycket värdefull information för upprättandet av ”relevanshierarkier” och urval av ord till olika slags produkter. Dessutom har ett antal (ibland diskutabla) ord tillkommit som översättningar från andra språk. Antalet lemma har en tendens att svälla och har på olika sätt begränsats till 50 000–80 000 beroende på språk.

Översättningslänkarna

Ekvivalens på ordnivå skapas genom att relationer eller länkar etableras mellan uppslagsorden för ett givet språkpar i särskilda översättningsfiler, i en-till-många/många-till-en relationer. Översättningslänken innehåller ett antal koder och dessutom kommentarfält, så att kopplingen sker till ett givet uppslagsord i målspråket, t.ex. eng. *hibernate* kopplas till sv. *ide* med *gå i* i ett förstått kommentarfält.

Fördelen med att som här etablera länkar jämfört med att skriva in översättningen i samma tabell är bland annat:

- att det blir enklare att koda översättningen för att sortera, kontrollera, analysera och välja ut dem för olika slags produkter;
- att ”vändning” sker löpande (Corda et al. 1998);
- att en effektiv kontroll av alla ”målspråksord” möjliggörs, eftersom de är länkade till ord i målspråkets grundfil, där systematisk korrekturläsning är gjord och uppdateringar av t.ex. ny stavning kan ske;
- att åtkomst av information om målspråksordet, t.ex. böjning eller uttal, sker direkt och entydigt.



FIGUR 2. *Skiss på länkarna mellan två språk. Ett uppslagsord som engelska practice kopplar till sv praktik, praxis, bruk, övning, träning etc. Sv praktik kopplar förutom till practice även till experience. Practice kopplar till en mängd exempel, varav ett är practice makes perfect, som i sin tur kopplar till övning ger färdighet i den svenska exempelfilen. På den svenska sidan är övning och färdighet länkade till detta exempel. På detta sätt kan man från färdighet komma till perfect, utan att orden är ekvivalenter på ordnivå.*

Exempelnivån

I en annan typ av filer återfinns det som kallas exempel, vilket är alla slags kombinationer av flera än ett uppslagsord. Här finns plats för konstruktionsmönster, kollokationer och andra döda exempel, idiom, ordspråk, kända citat och väl valda autentiska språkprov. En restriktion på 256 tecken har pålagts dessa exempel för att göra dem hanterbara på skärmbilden och även för att begränsa samlingen till så självgående exempel som möjligt.

Dessa exempel länkas till varje uppslagsord som ingår i dem, vilket gör utnyttjandet ekonomiskt i elektroniska produkter. Lexikografen slipper välja och användaren slipper leta. Samma förädlade exempelsamling kan även användas till andra produkter, t.ex. läroböcker, idiomordböcker, citatböcker.

Översättning av exempel görs lämpligen systematiskt av professionella översättare eller genom att förvärva rättigheterna till översättningar gjorda i ett annat sammanhang. I och med att allt är kodat i exempelsamlingen kan ekvivalenser för t.ex. samtliga idiom och ordspråk göras systematiskt. Översättningarna går (särskilt markerade) in i målspråkets exempelfil. Eftersom denna länk som alltid är en-till-flera/flera-till-en är naturligtvis alternativa översättningar möjliga.

Nya arbetssätt

Lexikografi är oavsett system en arbetsintensiv verksamhet. Medan datalingvister strävar efter att få datorn att efterlikna den mänskliga hjärnan är målet här endast att hitta en optimal arbetsfördelning mellan dator och människa, och också mellan lexikografen och andra personer.

Insamling av uppslagsord och exempel kan t.ex. delvis ske av inlärare som en del av språkstudierna. Kontroll och bearbetning av språkfilerna görs självklart av modersmålstalare. Ren korrekturläsning, som här sker vart språk för sig, kan göras av en annan person. Att grammatiskt kategorisera orden kan bli en specialuppgift för en lingvist, en fonetiker kontrollerar ljudskriften.

Ytterst sällan behöver man göra en genomgång av samtliga ord i filerna. I de flesta fall handlar det om att med olika metoder ta fram en delmängd med tveksamma fall.

Lexikografiskt arbete bygger alltid vidare på tidigare verk. I SI-systemet är syftet att i den mån upphovsrättslagstiftningen tillåter digitalisera delar av relevanta källor (alltså aldrig definitioner eller exempel), för att göra konsultationen effektivare och för att på ett systematiskt sätt kunna studera i vilken mån olika lösningar stämmer överens.

Riskerna med att förlita sig på tidigare produkter belyses av de spökord som förekommer i många tvåspråkiga produkter. I vissa fall läggs "virusord" avsiktligt in i ordböcker, för att kunna spåra plagiat (Vadtvedt Fjeld 1994).

När det gäller språkkombinationer där de tvåspråkiga (och kanske också de enspråkiga) produkterna är outvecklade kan man behöva gå igenom andra språk. Om man t.ex. arbetar från lettiska till svenska, kanske man kan behöva söka i en lettisk-tysk, sedan tysk-svensk, eller/och lettisk-rysk, sedan rysk-svensk ordbok. I SI-systemet kan man få upp resultatet av en sådan sökning automatiskt för varje skärmbild. Ju fler översättningsfiler som förädlats, desto mer avancerade blir möjligheterna att utnyttja dessa kopplingar.

I arbetet med den rysk-svenska översättningslänken för kompaktordboken (beskriven nedan) användes kopplingar via engelska, tyska och franska tillbaka till svenska, för att i största möjliga mån undvika de undermåliga rysk-svenska källorna. I vissa fall blev återkopplingarna helt entydiga, i många andra fall upptäcktes nya betydelser, och ofta kunde återkopplingarna ge inspiration till fräscha översättningar, även om ibland helt orelaterade ord kunde komma upp. Varje uppslagsord har en egen inmatningsbild som täcker hela skärmen och där all tillgänglig information från den ryska filen och från alla länkade stödfiler kan göras synliga. Felaktigheter och luckor kan korrigeras direkt.

Med denna metod visade det sig möjligt att gå igenom 100 ord per timme, något som också var möjligt för engelsk-svenska, även om arbets sättet där var annorlunda. Ungefär 95 procent av de ryska respektive engelska orden hittade en acceptabel ekvi-

valens på ordnivå, med den traditionella ordbokens måttstock, som i och för sig kan ifrågasättas (Toporowska Gronostaj 1993). Övriga ord kodades preliminärt för en systematisk bearbetning när grundfilens alla ord har gått igenom.

Tre ekvivalensnivåer i samspel

Målet är att de länkar som byggs på ord- och exempelnivå i största möjliga mån ska kunna vara gemensamma för samtliga produkter, såväl elektroniska som tryckta, stora som små, mer specialiserade som allmänna.

På ordnivå eftersträvas ett stramt urval av översättningar av ordets mest prototypiska betydelser. Användaren ska snabbt kunna bilda sig en uppfattning om ordets "väsen". Översättningarna presenteras snarare i etymologisk ordning än den frekvensordning som blivit allt vanligare och som motiveras av att vissa användare inte gör sig möda att läsa igenom långa ordboksartiklar och helt enkelt väljer den första översättningen. Här är denna risk mindre, eftersom orden följer direkt efter varandra, utan infogade exempel och andra kommentarer. För att begränsa antalet översättningar har principen följts att inte ta med mer än 7 ord, inspirerat av det faktum att korttidsminnet sägs kunna hantera 7 enheter, och av den journalistiska tumregeln att rubriker inte bör innehålla mer än 7 ord.

I den elektroniska standardprodukten samverkar ordnivån med exempelnivån. På exempelnivån ska man dels visa ekvivalenser för större språkliga enheter än ett ord, t ex idiom, dels ge inspiration från verkligt fungerade exempel hur ord kan översättas i en vidare kontext.

Genom att låta dessa två nivåer tala för sig själva kan man rationalisera det mödosamma arbetet att för varje uppslagsord författa en ordboksartikel. Istället består det lexikografiska arbetet i att välja ekvivalenter på ordnivå och att bedöma om redan befintliga exempel är tillfredsställande. Sökande efter nya och bättre exempel sker inte vart ord för sig utan för större grupper av ord av en viss typ. Exempelfilen rymmer i princip hur många poster

som helst, men för användarvänlighetens skull bör länkarna mellan ord- och exempelfil redigeras, så att ett rimligt och representativt urval blir synligt, inte minst för att skapa ett mervärde över de ”råa” parallellkorpora som i framtiden kanske kommer att bli tillgängliga via Internet även för slutanvändare.

Dessa två nivåer täcker dock inte in samtliga ord, och under arbetets gång har behovet av ytterligare en nivå uppstått, något som lösts genom de så kallade analytiska artiklarna. De ord som inte fått en fullgod ekvivalens på ord- eller exempelnivå låter sig tämligen naturligt grupperas tematiskt, t.ex. kulturspecifika företeelser som utbildning, mat, familjeord, helgdagar, religion etc. I en sådan artikel kan begreppen förklaras utförligt, ställas mot andra termer i samma grupp och möjliga ekvivalenser i målspråket. Artiklarna är bidirektionella, samma artikel ska alltså kunna användas såväl produktivt som receptivt, men artiklarna är avsedda endast för en av språkparets två språkgemenskaper. Denna typ av artikel, som kan omfatta upp till 100 olika uppslagsord, kan även vara en lämplig form för andra ord som användaren sannolikt söker vid samma tillfälle: datatermer, ekonomi och finans, militära grundbegrepp och andra semantiska fält. Även ord som inte låter sig fångas av 7 översättningar kan vara lämpliga att presentera i artikelform, t.ex. homonymer, för vilka även etymologiska samband kan utredas.

Förutom att artikeln blir platsbesparande i tryck och tillför användaren ett mervärde blir den också mer effektiv att utarbeta. Lexikograferna kan arbeta i par med var sitt modersmål och experter kan kallas in för att utreda ett ämnesområde, där orden är givna och den utredande stilen är enkel att anamma.

Produktutveckling

Alla tre nivåerna är självklara i den elektroniska produkten. På uppslagsordets skärmbild kan man (förutom valfri information som hör till uppslagsordet) se översättningarna på ordnivå till andra språk, de exempel där uppslagsordet ingår och dessa ex-

empels översättning till andra språk, samt länkarna till de analytiska samlingsartiklar där ordet kan ingå.

Exempelnivån tar dock stor plats i tryck. I den elektroniska versionen kan man hämta upp samma exempel på flera ställen. I en tryckt produkt skulle man antingen bli tvungen att trycka långa exempel upprepade gånger, eller att samla dem i en viss del och konstruera ett referenssystem, som lär bli tämligen användarvänligt.

Om den elektroniska produkten framför allt är tänkt som en "skrivbordsprodukt" bokstavligt sett, lämpar sig den tryckta produkten då man behöver en snabbreferens eller något man gärna läser.

Problemet med traditionella fickordböcker är att dess naturliga begränsning på cirka 15 000 uppslagsord i vardera riktningen (och frekvensprincip som urval) gör att de svåra och specifika ord man kan behöva i en given situation saknas, medan den innehåller mycket som åtminstone en lite mer avancerad ordboksanvändare redan känner väl till.

I detta projekt har en modell för en kompaktordbok skapats. Den innehåller två helt olika ordlistor för den receptiva och produktiva delen, samt ett urval av analytiska artiklar.

Målet är att reducera mängden uppslagsord till cirka 15 000 i den receptiva och cirka 25 000 i den produktiva delen. Dessa ordlistor består då enbart av direkta översättningar och/eller en referens till en specialartikel. Vägledande för urvalet är sannolikheten att den tänkta användaren kommer att söka ordet och vikten av att ordet verkligen finns i ordboken.

Från den receptiva delen kan man t.ex. utsluta:

- de elementära orden för en engelsk-svensk ordbok avsedd för mer avancerade användare borde man kunna utsluta den större delen av definitionsvokabulären från t.ex. Oxford Advanced Learner's Dictionary. Dessa frekventa ord är i regel de allra mest plats- och arbetskrävande i traditionella ordböcker.
- kognater, alltså de ord med gemensamt ursprung som har en liknande form och i stort sett samma betydelse. Extremfall är

de skandinaviska språken, där det är vedertaget att endast lista de ord som skiljer sig mellan språken. Finskan ligger å andra sidan nära det motsatta extremfallet, med ytterst få kognater. Språk som tyska, engelska och franska ligger däremellan, och har en betydande andel ord som knappast sökes i en receptiv del av en ordbok.

- andra än huvudordet i en ordfamilj (Nation & Bauer 1994), t.ex. för *internal*: *internally*, *internalize*, *internalization* (men knappast *intern*).
- transparent sammansatta ord, t.ex. *dairy farm*. En stor del av dessa är samtidigt att betrakta som kognater, t.ex. *database*.

I den receptiva delen inkluderas i gengäld fler abstrakta ord och lågfrekventa ord som kan vara av vikt för förståelse av instruktioner, juridiska texter, matsedlar etc.

I den produktiva delen måste man inkludera:

- de flesta högfrekventa ord, eftersom även de mest elementära orden tillfälligtvis kan falla ur minnet.
- de flesta kognater, eftersom användaren sällan kan vara säker på att det snarlika främmande ordet (vars exakta form dessutom kan vara mycket svår att minnas) verkligen är en lämplig översättning.
- avsevärt fler av de receptivt transparenta orden, eftersom de knappast är det ur ett produktivt perspektiv.

Däremot kan antalet abstrakta ord begränsas, eftersom användaren kan söka ordet genom en synonym eller en närbesläktad L1-form. Den produktiva delen kan i denna enkla form inte ge särskilt utförlig information om användningen av L2-formen, utan är mer tänkt som en kandidatlista och en inkörspport till en större enspråkig ordbok.

Ur båda delarna skulle man kunna utesluta siffror, geografiska namn, kalenderuppgifter och en stor del av de grammatiska orden, t.ex. pronomen och prepositioner. Dessa återfinns istället i särskilda tabeller tillsammans med de analytiska artiklarna. Ett sådant upplägg förutsätter naturligtvis att användaren är väl för-

trogen med ordbokens upplägg, vilket sällan är självklart, trots ambitiösa förord och instruktioner.

Principerna för dessa ordlistor är starkt påverkade av det tryckta formatets begränsningar. Naturligtvis vore det bättre att kunna ta med samtliga ord. En framtida produkt skulle kunna vara en bok med de analytiska artiklarna och tabellerna, samt ett urval av läsvärda exempel. Ordlistorna med snabböversättningar och/eller referenser till artiklarna skulle kunna ges av ett elektroniskt sökregister infällt i bokpärmen.

Slutsats

Drivkraften bakom detta projekt har varit att komma fram till en optimal datastruktur för ett allmänt lexikografiskt material i den digitala världen, oberoende av tidigare system, organisationsformer, produkter och arbetssätt.

I en mycket kraftfull och användarvänlig relationsdatabas för PC har en mångspråkig datastruktur experimenterats fram. Uppslagsord och exempel, som återfinns i separata filer för varje språk, fungerar på oberoende nivåer och sammanlänkas parvis med andra språk för att skapa översättningar eller språklig ekvivalens. Denna struktur underlättar avsevärt det lexikografiska arbetet och öppnar möjligheterna till olika slags analyser och elektroniska specialprodukter, inte minst som referens för att hantera den enorma mängd språklig råvara som idag finns tillgänglig på Internet. I den elektroniska konsumentprodukten, som i någon form sannolikt kommer att ersätta stora tryckta produkter, får ordnivån med de mest prototypiska översättningarna samverka med den hårt redigerade parallellkorpus som uppstår ur exempelnivån. Man frångår alltså den traditionella och extremt arbetsintensiva ordboksartikeln. Istället skapas en ny typ av analytisk bidirektionell artikel med upp till 100 uppslagsord för de cirka 5 procent av databasens ord som inte kan översättas på ordnivå eller exempelnivå utan kommentar. Dessa analytiska artiklar kan även användas i tryckt form, förslagsvis i en kompektordbok, med ordlistor i vilka man istället för de mest

frekventa väljer ut de ord användaren sannolikt kommer att behöva slå upp. En sådan tryckt produkt kan också vara reklam-bärare för elektroniska produkter och massdistribueras till nya målgrupper. Tack vare datastrukturen i relationsdatabasen och det kraftigt förenklade lexikografiska arbetet möjliggörs två-språkig ordboksproduktion även för mer marginella språkkombinationer.

Litteratur

OALD 1995 = *Oxford Advanced Learner's Dictionary*. 1995.

Editor: Jonathan Crowther. Oxford University Press.

Nationalencyklopedins ordbok på CD-rom. 1997. Göteborgs universitet. Institutionen för språkvetenskaplig databehandling.

Höganäs: Bra böcker.

Eurodicautom: <http://eurodic.echo.lu>

EuroWordNet: <http://www.hum.uva.nl/~ewn>

SWordNet: <http://galaxy.ling.lu.se/projects/Swordnet>

Alonge, A. 1998. "Encoding data on lexicalization of semantic components: the EuroWordNet 'relational' model." In: *EURALEX Proceedings*, eds. Fontenelle, T. et al. Liège.

Bauer, L. & Nation, P. 1994. "Word families." *International Journal of Lexicography* 6/4:253–279.

Bergenholtz, H., Svensén, B. 1994. "Systematisk inledning till Nordisk lexikografisk ordbok." *LexicoNordica* 1:149–186.

Bergenholtz, H. et al. (red.) 1997. *Nordisk leksikografisk ordbok*. Oslo: Universitetsforlaget.

Corda, A. et al. 1998. "Automatic Reversal of a Bilingual Dictionary: Implications for Lexicographic Work." In: *EURALEX Proceedings*, eds. Fontenelle, T. et al. Liège.

Svensén, B. 1987. *Handbok i Lexikografi*. Stockholm: Esselte Studium.

Toporowska Gronostaj, M. 1993. "Mot en mångdimensionell tvåspråkig ordbok." *Nordiske studier i leksikografi*. 2:96–122.

Vatvedt Fjeld, R. 1994. "Om ordbokskriminalitet og etikk i leksikografisk arbeid." *LexicoNordica* 1:27–42.

Författarna

- Anna Helga Hannesdóttir* Göteborgs universitet, Institutionen för svenska språket, Box 200, SE-405 30 Göteborg
- Maria Arnstad* Stockholms universitet, Institutionen för lingvistik, SE-106 91 Stockholm
- Antti Arpe* Lingsoft Ab Fabriksgatan 27–29 D, FIN-00150 Helsingfors
- Juhani Birn* Lingsoft Ab Fabriksgatan 27–29 D, FIN-00150 Helsingfors
- Tove Bjørneset* Universitetet i Bergen, HIT-senteret, Allégt. 27, N-5007 Bergen
- Erica Brewitz* Helgagatan 36, SE-116 58 Stockholm
- Anna-Lena Bucher* Terminologiceentrum TNC AB, Västra vägen 7B, Terrassen; SE-169 61 Solna
- Yvonne Cederholm* Göteborgs universitet, Institutionen för svenska språket, Box 200, SE-405 30 Göteborg
- Ulla Clausén* Svenska språknämnden, Box 20057, SE-104 60 Stockholm
- Elżbieta Dura* Göteborgs universitet, Institutionen för slaviska språk, Box 200, SE-405 30 Göteborg
- Øystein Eek* Kunnskapsforlaget, Postboks 6736, N-0130 Oslo
- Guðrún Kvaran* Orðabók Háskóla Íslands, Neshaga 16, IS-107 Reykjavík
- Åsa Holmér* Terminologiceentrum TNC AB, Västra vägen 7B, Terrassen, SE-169 61 Solna
- Sofie Johansson Kokkinakis* Göteborgs universitet, Institutionen för svenska språket, Box 200, SE-405 30 Göteborg
- Dimitrios Kokkinakis* Göteborgs universitet, Institutionen för svenska språket, Box 200, SE-405 30 Göteborg
- Ulrika Kvist Darnell* Svenska språknämnden, Box 20057, SE-104 60 Stockholm
- Godelieve Laureys* Universiteit Gent, Seminarie voor Skandinavistiek, Rozier 44, B-9000 Gent

- Henrik Lorentzen* Den Danske ordbog, Christians Brygge 1, DK-1219 København K.
- Nina Martola* Forskningscentralen för de inhemska språken, Sörnäs strandväg 25, FIN-00500 Helsingfors
- Sjur Nørstebø Moshagen* Lingsoft Ab, Fabriksgatan 27–29 D, FIN-00150 Helsingfors
- Torbjørn Nordgård* Universitetet i Trondheim, Lingvistisk institutt, N-7491 Trondheim
- Eva Orava* Lingsoft Ab, Fabriksgatan 27–29 D, FIN-00150 Helsingfors
- Christian-Emil Ore* Universitetet i Oslo, Dokumentasjonsprosjektet, Postboks 1123, Blindern, N-0315 Oslo
- Riitta Palkki* Forskningscentralen för de inhemska språken, Sörnäs strandväg 25, FIN-00500 Helsingfors
- Bo Ralph* Göteborgs universitet, Institutionen för svenska språket, Box 200, SE-405 30 Göteborg
- Gudrun Rawoens* Universiteit Gent, Seminarie voor Skandinavistiek, Rozier 44, B-9000 Gent
- Lena Rogström* Göteborgs universitet, Institutionen för svenska språket, Box 200, SE-405 30 Göteborg
- John Sinclair* The Tuscan Word Centre, Azienda Casanova, Vellano, 409-51010 Pescia, Italien
- Emma Sköldberg* Göteborgs universitet, Institutionen för svenska språket, Box 200, SE-405 30 Göteborg
- Bo Svensén* Svenska Akademien, Box 2118, SE-103 13 Stockholm
- Lars Törnqvist* Älvkarleövägen 13, SE-115 43 Stockholm
- Ruth Vatvedt Fjeld* Universitetet i Oslo, Avdeling for leksikografi, INL, Postboks 1001 Blindern, N-0315 Oslo
- Fredrik Westerlund* Lingsoft Ab, Fabriksgatan 27–29 D, FIN-00150 Helsingfors
- Lars Vikør* Universitetet i Oslo, Avdeling for leksikografi, INL, Postboks 1001, Blindern, N-0315 Oslo
- Matti Vilppula* Forskningscentralen för de inhemska språken, Sörnäs strandväg 25, FIN-00500 Helsingfors
- Jonas Winnerlöv* Drottninggatan 101, SE-111 60 Stockholm