

Indholdsfortegnelse

<i>Henning Bergenholtz/Sven-Göran Malmgren</i> Forord	1
--	---

Tematiske bidrag

<i>Henning Bergenholtz</i> Korpusbaseret leksikografi	5
--	---

<i>Eiríkur Rögnvaldsson</i> Brugen af et gammelislandsk tekstkorpus i leksikografisk arbejde	19
--	----

<i>Martin Gellerstam:</i> Rätt och fel i korpusen	35
--	----

<i>Henrik Holmboe</i> Genbrug af korpora	49
---	----

<i>Tove Jacobsen</i> Oversættelseskorpora og tospråklige ordbøger	57
--	----

<i>Lars G. Johnsen & Torbjørn Nordgård</i> Korpusbasert leksikonbygging	69
--	----

<i>Göran Kjellmer</i> Idiomen, kollokationerna och lexikonet	79
---	----

<i>Carl-Erik Lundbladh</i> Värdering av SAOB:s korpus och språkprovssamling	91
--	----

<i>Nina Martola</i> Finlandssvenska, finsk-svensk tvåspråkslexikografi och korpusar	105
---	-----

<i>Ole Norling-Christensen</i> Klassifikation af korpustekster, og kvantitative mål for sammensætningen af et almensprogligt korpus	121
---	-----

<i>Liisa Nuutinen</i> Korpus över äldre finska	129
---	-----

Jette Pedersen
Sammensætning af et tekstkorpus til en bilingval teknisk fagordbog
og forslag til metodiske lemmaselektionskriterier139

Ikke-tematiske bidrag

Hans Kristian Mikkelsen
Leksikografiske kvalitetsindikatorer. Selvfølgelig som
udgangspunkt for evaluering153

Recensioner

Henning Bergenholtz
Norsk Riksmålsordbok195

Ilse Cantell/Nina Martola
Nationalencyklopedins ordbok209

Karl Johan Danell
Jeannie Enwall/Ruth Lötmarker: *Fransk-svensk ordbok*223

Ela Dura, Sven-Göran Malmgren & Maria Toporowska-Gronostaj
Jacek Kubicki: *Słownik szwedzko-polski/Svensk-polsk ordbok.*
Jacek Kubicki: *Słownik polsko-szwedzki/Polsk-svensk ordbok*229

Birger Lohse
Kåre Nilsson: *Norsk-portugisisk ordbok*243

Sölve Ohlander
"The Big Four." "Learners' dictionaries" inför 2000-talet.
[Rec. af *Cambridge International Dictionary of English*;
Collins COBUILD English Dictionary; *Longman Dictionary*
of Contemporary English; *Oxford Advanced Learner's*
Dictionary of Current English]257

Sölve Ohlander
Ingvar Stenström, *Interlingua-svensk ordbok.*
25.000 internationella ord279

Jette Pedersen, i samarbejde med *Tine Pedersen*
Jens Axelsen: *Dansk-Engelsk Ordbog*287

Klaas Ruppel
Det rör på sig inom finsk-tysk lexikografi
[Rec. af Aino Kärnä: *Saksa-suomi-opiskelusanakirja*]295

Ingrid Simonnæs
Wilhem Gubba: *Juridisk ordbog tysk-dansk*307

Marketta Sundman
Ilse Cantell, Nina Martola, Birgitta Romppanen,
Mats-Peter Sundström: *Suomi-ruotsi opiskelusanakirja*315

Trond Trosterud
Turid Farbregd/Hannele Seppinen: *Finsk-norsk ordbok*335

Omtaler

Franziskus Geeb
Magnar Brekke, Øivin Andersen, Trine Dahl,
Johan Myking (udg.): *Applications and Implications
of current LSP Research. Proceedings of the 9th
Europoean Symposium on LSP*345

Martin Gellerstam
Birgitta Lindgren et al.: *Skandinavisk ordbok*347

Gu>var>ur Már Gunnlaugsson
Íslendskur-føroyiskur or>alisti349

Indgåede bøger351

Meddelelser fra foreningen353

Henning Bergenholtz/Sven-Göran Malmgren

Forord

I alle tre hidtidige udgaver af tidsskriftet har redaktionen haft den på samme tid problematiske og fordelagtige situation, at der har været flere bidrag, end vi har kunnet optage. For at et tidsskrift skal kunne eksistere, må mindst to forudsætninger være opfyldt: et rigeligt antal tilbud om artikler, diskussionsbidrag og anmeldelser og et tilstrækkeligt antal abonnenter. Til hver temadel har de fleste artikler dog bestået af skriftlige versioner af nogle af de foredrag, som har været holdt i forbindelse med de symposier, som har været afholdt på Schæffergården. Mængden af ikke-tematiske artikler har derimod indtil nu været mindre end mængden af tematiske bidrag. Derfor er der flere grunde. Dels har redaktionen hidtil ikke gjort forsøg på at opfordre til yderligere bidrag, bl.a. med henblik på nødvendigheden af at begrænse omfanget af de enkelte numre. Dels har tidsskriftet en naturlig konkurrent i Nordiske Studier i Leksikografi, som indtil nu er udkommet hver andet år i tilslutning til foreningens konferencer. Bind tre udgivet af Ásta Svavarsdóttir/Guðrún Kvaran/Jón Hilmar Jónsson med bidrag fra konferencen i Reykjavík '95 indeholder således 42 artikler.

Vi har derfor følgende situation: På den ene side ser vi os nødsaget til at påpege, at den hidtidige praksis ved antagelse af artikler vil blive yderligere strammet, så omfanget af de enkelte numre af Lexico-Nordica ikke overstiger 250 sider. Vi vil på den anden side samtidig tilskynde til indsendelse af uopfordrede bidrag til den ikke-tematiske del. Derudover forventer og håber vi på en fortsat indsendelse af anmeldelser, som udgør en væsentlig del af tidsskriftets profil. Her vil vi pege på, at den stigende strøm af CD-ROM-ordbøger indtil nu ikke har ført til en tilsvarende mængde af anmeldelser.

Anmeldelser er i øvrigt objekt for en stigende interesse med forslag til en slags normering af varedeklARATIONER, omtaler og af anmeldelser for almindelige ordbogsbrugere og for metaleksikografer (bl.a. Thomsen 1996). En af artiklerne i dette nummer kan forstås både som et bidrag til yderligere overvejelser om leksikografiske kvalitetsbedømmelser. Mikkelsens mål er "at få kortlagt omfanget og arten af de eksplicite selvforståelsesoplysninger, idet en sådan oversigt vil kunne tjene som værktøjskasse for den, som skal evaluere en ordbog". Denne selvforståelse findes i ordbøgernes forord og bruger-

vejledning, som her får en positiv revurdering. Disse fortekster læses i øvrigt også i højere grad af andre end metaleksikografer (hvilket mange forlagsleksikografer betvivler). Dette påvises af Wolf (1994:335–338), som har analyseret svar fra 288 informanter. Disse informanter deles op i FA (Facharbeiter), FS (Berufstätiger mit Fachschulabschluß), HS (Berufstätiger mit Hochschul-/Universitätsabschluß) og angiver alt have læst forord og evt. brugervejledning i følgende omfang:

antal informanter i %			
læst forteksten	HS	FS	FA
ja	66	74	55
nej	34	26	45

Andelen af informanter, der har læst forteksten, anser Wolf (1994: 336) for utilfredsstillende. Vi deler ikke denne interpretation. Hvis virkelig dette resultat kan overføres til andre brugerundersøgelser, taler alt for at lægge større og ikke mindre vægt på en udførlig brugervejledning. Det gælder desuden også, hvis tallene for nej- og ja-svar blev byttet om.

Artiklen af Mikkelsen giver anledning til endnu en kommentar. Denne indeholder sammen med en bedømmelse af forteksten i 155 danske fagordbøger bibliografiske angivelser til disse ordbøger. Lignende oversigter foreligger sandsynligvis også for andre ordbøger i Norden. Det må være en opgave for den leksikografiske forskning i Norden at tage initiativ til en sådan bibliografisk oversigt. Med udgangspunkt i Haugen (1985) kan man fortsætte med at optage samtlige relevante titler i Claes/Bakema (1994), en bibliografi over hollandske ordbøger. Denne bibliografi udmærker sig bl.a. ved at have et eget register over de sprog, som indgår i de anførte ordbøger. Heraf er der til de enkelte nordiske sprog følgende antal titler:

dansk: 183
 islandsk: 16
 finsk: 46
 norsk: 68
 svensk: 167

En pendant til den hollandske bibliografi med en systematisk bibliografi over nordiske ordbøger vil i første række kunne gøre stor nytte i Norden, men vil også i relation til alle andre sprog kunne være af stor værdi.

Temadelen i LexicoNordica 3 behandler den leksikografiske brug af korpora. Over for de indbudte deltagere i konferencen på Schæffergården i februar 1996 blev temavalget indkredset på følgende måde: "For tyve år siden var det ikke bare ingen selvfølgelighed at inddrage et tekstkorpus som den vigtigste del af en ordbogs empiriske basis. Det er det måske stadigvæk ikke i alle kredse, der findes fx en del tyske leksikografer og leksikologer, som har en række indvendinger. Men alligevel kan man i almindelighed sige, at følgende beskrivelse svarer til den generelle tendens: Kvaliteten af en ordbog er bl.a. direkte proportional med kvaliteten af den empiriske basis. På trods af denne principielt positive tendens foreligger der et væsentligt metodisk deficit. Hermed menes overvejelser om sammensætning af korpora til forskellige ordbogstyper og metodiske overvejelser om systematiske analyser af foreliggende korpora. Derimod foreligger en række instruktive bidrag om de forskellige værktøjer, leksikografen kan inddrage, således at den mere edb-mæssige problematik ikke skal stå i centrum ved det planlagte symposium." Efter afholdelsen af symposiet er denne beskrivelse ikke mere helt korrekt. Bidragene til Euralex-symposiet i Göteborg (Gellerstam et al. 1996), i det første nummer af tidsskriftet "Corpus Linguistics" og i det sidste nummer af "International Journal of Lexicography" indeholder flere artikler, som ikke koncentrerer sig om computerprogrammer og andre mere tekniske forhold.

Rent redaktionelt har vi tre tilføjelser. Vi vil for det første tilskynde potentielle bidragsydere til at tage udgangspunkt i manuskriptreglerne, som til dette nummer er blevet ændret uvæsentligt. En ændring er, at deadline er flyttet frem til 1. juni. Til gengælde vil vi for eftertiden holde os strengt til denne tidsfrist, så at bidrag, der indleveres efter denne dato, vil blive afvist.

Dernæst vil vi over for de forlag, som i sidste nummer af tidsskriftet havde indleveret annoncer, beklage den dårlige tekniske kvalitet af netop denne del. Vi har for fremtiden indført en yderligere kvalitetskontrol af forlagsreklamerne.

Endelig vil vi takke Nordisk Språksekretariat i Oslo, specielt Ståle Løland og Rikke Hauge, for en uvurderlig hjælp i forbindelse med forberedelse, trykning og distribution også af dette nummer af LexicoNordica. Beslutningen om nedlæggelse af sprogsekretariatet kommer uden tvivl til at være forbundet med ændringer i forudsætningerne for dette samarbejde. Vi vil her udtrykke ønske om samarbejde under de nye rammer, forbundet med ett fortsat godt samarbejde med de hidtidige medarbejdere på Nordisk Språksekretariat i Oslo.

Litteratur

- Ásta Svavarsdóttir/Guðrún Kvaran/Jón Hilmar Jónsson (red.): *Nordiske studier i leksikografi 3. Rapport fra Konference om leksikografi i Norden. Reykjavik 7.–10. juni 1995*. Reykjavík: Nordisk forening for leksikografi 1995.
- Claes, Frans/Peter Bakema 1995: *A Bibliography of Dutch Dictionaries*. Tübingen: Niemeyer.
- Gellerstam, Martin/Jerker Järborg/Sven-Göran Malmgren/Kerstin Norén/Lena Rogström/Catarina Röjder Pappmehl (eds.) (1996): *Euralex '96. Proceedings I–II. Papers submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden*. Göteborg: Göteborg University.
- Haugen, Eva L. 1985: *A Bibliography of Scandinavian Dictionaries*. New York: Kraus.
- Mikkelsen, Hans Kristian: Leksikografiske kvalitetsindikatorer. Selvfølgelig som udgangspunkt for evaluering. I: dette tidsskrift.
- Thomsen, K.T. 1996: Om varedeklARATIONER for ordbøger. I: *LEDA. Nyhedsbrev nr. 21 – sept. 1996*, 17–26.
- Wolf, Birgit: Wörterbuch und Benutzer – Versuch einer empirischen Untersuchung. I: Ursula Brauße/Dieter Viehweger (Hrsg.) 1994: *Lexikontheorie und Wörterbuch. Wege der Verbindung von lexikologischer Forschung und lexikographischer Praxis*. Tübingen: Niemeyer, 295–389.

Henning Bergenholtz

Korpusbaseret leksikografi

The first computerised corpora (Brown, LOB and Limas) can be described as poly-functional corpora of partial texts. They consisted of a certain number of small, brutally cut parts of texts, all of the same length, and they were intended as a linguistic basis for all linguistic research fields one could think of. But they could not fulfil this aim. Not only lexicography, but especially this science requires corpora with whole texts that have been collected for a special purpose. Whole texts here understood as texts taken in their entirety or at least parts of texts that may be seen as a whole, for instance a complete chapter.

1. Tilbageblik

I 1977 kunne man med god ret skrive om en ny udvikling inden for leksikografien: "Lexikographen legen neuerdings verstärkten wert auf satzübergreifende textkorpora" (Wiegand 1977:129). Udtrykket "tekstkorpora" var så nyt, at Wiegand følte, at det måtte forklares, idet han tilføjede attributtet "satzübergreifend" (går ud over sætningsgrænsen). Hvis Wiegand i stedet for **tekstkorpus** havde brugt udtrykket **korpus**, havde tilføjelsen været mere nødvendig, idet korpus traditionelt ikke bare er blevet brugt om en samling af tekster, men også om en samling af belæg, desuden om en samling af selvdannede eksempler, ja endda om en samling af kollokationer eller ord. Der er i virkeligheden tale om mindst tre forskellige måder at brugen termen korpus på.

For det første har vi dem, der ved et korpus forstår en samling af sproglige enheder under sætningsniveau, dvs. af ord og kollokationer. Denne gruppe er ikke relevant for korpusdiskussionen i denne sammenhæng. For det andet kan man udskille de leksikografer og lingvister, som enten antager samme status for selvlavede eksempler og tekstbelæg (Greenbaum 1984:194) eller advarer mod brugen af et tekstkorpus eller af belæg (Zöfgen 1986 og Pasch 1992:284). Diskussionen med sådanne belæg- og korpus skeptikere er blevet ført i anden sammenhæng (se Bergenholtz/Mugdan 1989 og 1990) og vil ikke blive gentaget her. Endelig er der de leksikografer, der exciperer belæg fra tekster til en seddelsamling eller anvender et tekstkorpus. Det er de to sidste gruppe, som står i centrum i dette bidrag.

I denne sammenhæng er Wiegands udtryk "neuerdings" (i den sidste tid) og "verstärkten wert" (større vægt) særligt interessante. Det var selvfølgelig ikke noget helt nyt, at leksikografer helt eller delvis brugte

foreliggende tekster som empirisk basis. I lingvistikens historie har det i større eller mindre grad været usus for bl.a. retorikker og grammatikere; og belægsamlinger har i flere hundrede år været det væsentlige empiriske grundlag for store monolingvale betydningsordbøger. Men det er først i dele af den unggrammatiske skole, at denne fremgangsmåde bliver fremlagt som et metodisk ideal for specielt den grammatiske og i videre forstand al lingvistisk forskning. Behaghel beskriver i forordet til sin firebinds tyske syntaks netop modsætningen til de grammatikere, som har samlet så mange belæg som muligt og så vidt muligt tager højde for dem alle. På denne måde, siger Behaghel, inddrager man i virkeligheden først og fremmest det påfaldende, det sjældne og undtagelserne. Alt det regelmæssige i sproget bliver derimod taget relativt mindre hensyn til. Og én ting kan man overhovedet ikke sige noget om, nemlig at formodede forekomster slet ikke kan findes. Eksempler på det, Behaghel taler om, findes der mange af. Et af de særlig tydelige er problemet kasuskongruens i tyske nominalfraser. Her har en lang række grammatikere af sådanne eksempelsamlinger ladet sig forlede til at udtale sig både om middelhøjtysk, tidligt nyhøjtysk og den nuværende sprogbrug, at det må anses for at være meget udbredt med en manglende kasuskongruens, sml. *mit Herrn Müller, dem berühmten Maler* og *mit Herrn Müller, des berühmten Malers*. Den sidste eksempeltype uden kasuskongruens kan findes, men ret sjældent, ret nøjagtig i ca. 3,6% af tilfældene (Bergenholtz 1985).

2. Empirisk basis i leksikografien

I den leksikografiske verden kender vi tilsvarende resultater af belægsamlinger. Hvert belæg er ganske vist et virkeligt eksempel på konkret sprogbrug. Men de, der har søgt og skrevet belæggene af, har i høj grad ladet sig fange af det særlige, det overraskende. Om fx seddel-samlingerne for Duden-ordbøgerne eller ved det Danske Sprognævn er der derfor også med et gran af sandhed blevet sagt, at de er præget af sproglige perversiteter. I stedet for en ukontrolleret belægindsamling foreslår Behaghel følgende metode for den grammatiske forskning:

Bei diesen Untersuchungen habe ich das Verfahren beobachtet, das sich bei meiner Arbeit über die Zeitfolge bewährt hat, das Verfahren der Stichproben, das gewisse Stücke gewisser Denkmäler vollständig auszubehuten sucht. Wer danach andere Stücke und andere Quellen durchmustert, wird vielleicht wertvolle Ergänzungen bieten können, aber das von mir Gefundene kaum gänzlich umwerfen. (Behaghel 1923:VIII).

Der er to afgørende udtryk i dette citat, som er relevante for den leksikografiske diskussion. Det er for det første kravet om en fuldstændig analyse af hele stikprøven. Hermed adskiller fremgangsmåden sig positivt fra metoden med den myreflittige samlen, som resulterer i et tilfældigt og ikke-kontrollerbart udvalg af alle potentielle belæg, og hvor man især ikke kan forestille sig, at to forskellige excerptorer skulle kunne komme frem til en bare nogenlunde ensartet selektion. Behaghels krav er dog metodisk ikke uden problemer for leksikografisk arbejde. Det kan opfyldes i mange grammatiske undersøgelser, fx ved eksemplet appositioners kasuskongruens. Det tog omkring 400 timer at gennemlæse og afmærke alle appositioner i et korpus med to mio. tekstord, men det drejede sig trods alt om ikke mere end godt 6.000 belæg. I leksikografien vil et sådant korpus ved mange problemstillinger være alt for lille. Men alligevel vil man selv ved et så beskedent korpus have så mange belæg til særlig hyppige funktionsord, at arbejdets omfang med at tage nøjagtig højde for alle forekomster vil udgøre en unødvendig stor arbejdsbelastning. Med andre ord vil vi have det problem, at vi til forskellige leksikografiske problemstillinger vil være tvunget til at gøre brug af en større eller mindre del af det sammenstillede korpus eller evt. have forskellige korpora til forskellige problemstillinger. I virkeligheden svarer dette præcis til Behaghels fremgangsmåde. Han har i praksis alt efter problemstilling haft forskellige "Denkmäler", men det anser han for noget så selvfølgelig, at der ikke bliver redegjort nøjagtigt herfor.

Det andet afgørende, men samtidig noget uklare udtryk i ovennævnte citat er "visse dele af visse tekster": Hvor mange tekster drejer det sig om? Hvordan er dette tekstudvalg truffet? Hvor store dele af teksterne er medtaget? Og er det hele tiden den samme del af teksterne, fx det første afsnit eller kapitel? Også herom siger Behaghel intet.

3. Teoretiske og metodiske overvejelser

De nævnte problemer udgør nogle af de centrale problemer, som korpussammensætningen og korpusanalysen også i dag står over for. En af de første bøger, som udelukkende tematiserer et tekstkorpus' status, dets sammensætning og analysemetoder, er "Empirische Textwissenschaft, Aufbau und Auswertung von Text-Corpora". Denne bog har ikke fundet videre genklang, sandsynligvis fordi bogen har en tysk titel og de fleste af bidragene er skrevet på tysk. Som særligt centrale anser jeg bidragene af van de Velde (1979), Bausch (1979), Bungarten (1979) og Rieger (1979). Men disse og de fleste andre artikler i denne bog har i

den videnskabelige korpusdiskussion, som i særlig grad har været præget af anglistere, ikke spillet den centrale rolle, som de havde fortjent og som sandsynligvis kunne have haft påvirket et til dels problematisk diskussionsforløb i positiv retning. Det drejer sig dels om den statistisk uholdbare tale om 'repræsentative korpora', dels om et korpus' videnskabsteoretiske status i lingvistikken og i andre fag.

Særligt misvisende er den udbredte betegnelse "korpuslingvistik", som efterhånden har bredt sig fra anglisternes rækker til resten af lingvistikken og i øvrigt har givet navn til et nyt tidsskrift. Hermed lader man en særlig form for empirisk basis udgøre en særlig lingvistisk disciplin. Med samme tankegang kunne man også opfinde en intuitionslingvistik, spørgeskemalingvistik eller belægsamlingslingvistik. Tilsvarende kunne man have en korpusleksikografi, en intuitionsleksikografi eller måske i særlig grad en skrive-af-leksikografi. Betegnelsen **korpuslingvistik** er i øvrigt dobbelt misvisende. Ikke alene hylder man en betænkelig habeas-corpus-holdning, som fører til, at det at sammenstille korpus skulle være et særligt sprogvidenskabeligt mål i sig selv. Men man ser også bort fra, at tekstkorpora desuden sammen sættes og bruges af andre end lingvister, sml. fx Mergenthaler (1979).

På sin vis er denne habeas-corpus-holdning ikke spor bedre end den tilsvarende radikale afvisning af brugen af korpora, sml. fx: "That is a complete waste of your time and the government's money. You are a native speaker of English; in ten minutes you can produce more illustrations of any point in English grammar than you will find in many millions of words of random text.". (Robert Lees 1962, mundtlig diskussion, citeret i Francis 1979:110). Francis kan i sin modreplik have ret i, at Chomsky hverken dengang eller senere har forelagt en holdbar definition af competence eller performance, og at han ikke kan forstå, hvordan competence kan undersøges uden hensyntagen til virkelig performance, dvs. tekster. Men Francis' glæde over korpus-skeptikernes nederlag var mere en foregriben af en senere udvikling end en beskrivelse af en afsluttet debat, idet diskussionen om værdien af tekstkorpusbaserede undersøgelser vedblev at være noget følelsesladet, sml. fx beskrivelsen af korpusanalyser som "eine überflüssige Zeremonie" (Itkonen 1976:65).

4. Sættning af korpus

Rent historisk kan man fastslå, at de første korpora, begyndende med det amerikanske Brown-Corpus, over det engelske LOB-Corpus og det tyske Limas-Korpus er totalsproglige korpora, dvs. korpora bestående tekster fra både fagsprog og almensprog. Hvert af disse korpora tager

hensyn til en sagklassifikation, som fx ved Limas-Korpus går ud fra opdelingen i Frankfurter Nationalbibliographie og fra en vægtning efter mængden af udkomne tekster i de der brugte 33 "sagområder". Hver deltekst har en størrelse på 2000 tekstord. Korpus, som er på en mio. tekstord, omfatter således i alt 500 enkelttekster, fx 17 medicinske, 6 matematiske og 64 skønlitterære tekster. En sådan spredning gør det fx muligt at foretage frekvenssammenligninger mellem de hyppigst forekommende ord og fraser inden for de forskellige sagområder. Således kan Johansson (1979) i en undersøgelse af Brown-Corpus påvise tre gange så mange definite nominalfraser med efterfølgende *of*-konstruktioner i naturvidenskabelige og tekniske tekster som i fiktionale tekster.

Med Brown-Corpus og de andre små korpora på en mio. tekstord foreligger der helt sikkert det, som Wiegand kaldte "korpora, som går ud over sætningsgrænsen". Nøjagtigere leksikalske undersøgelser af sprogbrugen i enkelte fag- eller sagområder er selvfølgelig slet ikke mulige i et korpus med en så svag dækning af de enkelte områder. Hverken omfanget eller sammensætningen gør det muligt at foretage bare nogenlunde sikre undersøgelser med henblik på en total- eller en almensproglig ordbog. Den undersøgelse, Wiegand refererede til, antog et omfang på omkring 50 mio. tekstord som et tilstrækkeligt stort korpus til en planlagt stor, ny totalsproglig ordbog, sml. hertil også Mentrup (1979:185).

En yderligere indskrænkning af mulige tekstlingvistiske og også leksikalske undersøgelser ligger i de tekstuelte brutale snit i udgangsteksterne, når hver korpustekst skæres ud i en ganske bestemt fast størrelse. Herved bliver tekstuelle sammenhænge i begyndelsen eller slutningen af en korpustekst af og til så uklare, at en sikker interpretation og dermed leksikografisk brug af disse tekstdele bliver umulig.

De hidtil nævnte korpora kan henregnes til typen **deltekstkorpus**. Disse korpora danner en færdig enhed, som ikke mere kan og skal ændres. En helt anden type er **heltekstkorpus**, som ikke kan skelnes skarpt fra et tekstbibliotek hhv. **tekstbank**. En særlig type heltekstkorpus er det såkaldte **monitor corpus**, som er blevet oprettet og brugt i forbindelse med Cobuild-projektet, sml. Renouf (1991). Et monitorkorpus optager så mange tekster som muligt, de modtages normalt direkte fra forlag o.l. og indgår i deres helhed i korpus, som kan "rulle" hen over skærmen. Alt efter behov bruges større eller mindre dele af det til enhver tid foreliggende tekstmateriale, som hele tiden bliver udvidet, ligesom tekster, som ikke længere er aktuelle, slettes eller overføres til et reservekorpus. På mange måde minder argumentationen om den, vi indledningsvis fandt hos Behagel.

De førstnævnte korpora (Brown, Limas og LOB) blev af deres oprettere forstået og beskrevet som **polyfunktionale** korpora til brug i alle eller en lang række lingvistiske forskningsområder. Sådanne potentielt alsidige brugsmuligheder kan man utvivlsomt også forudsætte, når der ikke angives specifikke enkeltproblemer, fx hos Glas (1975). I praksis er den forudsætte polyfunktionalitet hverken realistisk eller hensigtsmæssig. Til et bestemt problem må man sørge for fremskaffelsen af præcis den form for empirisk basis, som må anses for hensigtsmæssig. Et sådant **monofunktionalt** korpus foreligger fx med Maegaard-Ruus-korporaene, som ganske vist ikke uden store indskrænkninger kan bruges til andet end sprogstatiske undersøgelser, men som på den anden side er sammensat under særlig hensyntagen til i det mindste ét formål og også brugbare hertil. Derimod er en lang række polyfunktionale deltekstkorpora kun virkeligt brugbare til grammatiske undersøgelser af meget almen karakter, men ikke til opnåelse af nøjagtigere indsigt i grammatiske problemområder og slet ikke til leksikografiske undersøgelser.

5. Korpora som empirisk basis for almensproglige ordbøger

Mange, måske de fleste nyere europæiske og amerikanske ordbøger, beskriver sig selv direkte som totalsproglige ordbøger, dvs. ordbøger, der både tager hensyn til almensproglig og i større grad også til fagsproglig brug. Nogle går endda så vidt, at de som et af deres ordbogs formål ser en hjælp til forståelse mellem fagområderne, dvs. fagfolk imellem og mellem fagfolk og lægfolk, sml. "Durch die starke Berücksichtigung der Fachsprachen wird es auch eine sichere Basis für die Verständigung zwischen den Fachbereichen schaffen." (DUDEN-GWB, Vorwort, side 2). Man kan ganske vist godt forestille sig en stor interdisciplinær ordbog, som både tager virkelig hensyn til fagsprogene og til almensproget, men i det tilfælde vil man ikke kunne nøjes med inddragelse af leksikografer med en lingvistisk baggrund. For alle de angiveligt totalsproglige ordbøger, som er udkommet indtil nu, gælder det, som Paolo Beni i en anmeldelse af Accademia della Crusca's ordbog skrev i 1612 (se herom Hausmann 1989). Man bør i en optimal alment brugbar ordbog, mener Beni, foretage et meget bredt udvalg af stikord, som bruges i almensproget. Derimod frarådes optagelse af stikord fra de faglige ordforråd. Denne tankegang kan føres tilbage til den empiriske basis, som optimalt bør lægges til grund for en almen ordbog. Sådanne ordbøger udarbejdes af lingvistiske leksikografer, som ganske vist kan rådføre sig med fageksperter, men dog ikke har forudsætning for at udvælge de relevante fagtekster og endnu mindre

for at analysere alle mulige fagtekster. Det betyder ikke, at fagtermer helt skal udelukkes fra lemmalisten til en almensproglig ordbog. En lang række fagtermer indgår i det almene sprog. Og sådanne termer bør principielt også optages i lemmalisten. Det betyder heller ikke, at den leksikografiske definition ikke skal være fagligt korrekt, som fx COBUILD anser som en nødvendighed ud fra deres deskriptive argumentation, sml. "Hence we have explained the technical words according to the way we use them in ordinary English." (COBUILD, Introduction, p. XX). Men det betyder, at man bør indskrænke sig til et udvalg af sådanne tekster, som formodes hørt eller læst af en stor del af et bestemt sprogs modersmålsbrugere.

I forbindelse med konceptionen af det danske polyfunktionale deltekstkorpus DK87–90 (sml. Bergenholtz 1988) blev der således opstillet følgende udgangspunkt, som i nogen grad kan overføres til leksikografiske sammenhæng:

1. der skal ikke medtages fagtekster
2. korpus skal kun indeholde førstegangsudgivelser
3. der skal ikke medtages oversættelser
4. tekstudvalget skal være på forkant med sprogudviklingen
5. tekstrecipienterne skal udgøre en væsentlig del af befolkningen

Disse principper medfører (1), at der ikke medtages tekster af typen fagmand → fagmand eller fagmand → semifagmand. Der medtages almensproglige tekster, som muligvis er skrevet af en fagmand eller semifagmand, men hvor den forudsete tekstrecipient er en lægmand. Endvidere gælder (2), at kun udgivelser fra det pågældende år medtages, dvs. fx ikke H.C.Andersen-tekster. Sådanne tekster læses ganske vist i betydelig grad, men de genspejler ikke uden indskrænkninger den nutidige sprogbrug. Det samme gælder (3) for oversættelser, som ofte bærer tydeligt præg af deres oprindelige sprog. Mængden af aviser, dagblade, reklamer osv. er ganske vist betydeligt større end mængden af fiktionale tekster, sidstnævnte anses dog (4) for at være på forkant med udviklingen. Endelig (5) udelukkes tekster, som kun henvender sig til børn eller unge. Disse forudsætninger førte til følgende repræsentation af tekstarter:

1. romaner og noveller (50% af alle tekster)
2. aviser (25% af alle tekster)
3. ugeblade (25% af alle tekster)

Hver af disse tre tekstarter læses af mere end 50% af befolkningen, de er i modsætning til ikke-medtagne børnebøger og fagtekster ikke skrevet for en bestemt, forholdsvis begrænset del af den danske befolkning. Der er således gode grunde til at anse disse tre tekstarter for en væsentlig del af det almindelige danske skriftsprog. Derimod er procentopdelingen ikke uproblematisk; en opdeling mellem de tre tekstarter i forholdet 1:1:1 havde også været mulig. Jeg valgte at give romaner og noveller en særlig stor vægt, fordi de (på grund af dialogerne) i højere grad end aviser og ugeblade bærer præg af udviklingen i det talte sprog.

6. Korpora som empirisk basis til fagordbøger

De fleste foreliggende fagsproglige korpora er polyfunktionale heltekstkorpora. Et af dem, det såkaldte **Limas-Kfz-Korpus**, indeholdende tekster om automobiler hentet fra lærebøger, brugsanvisninger og bilfagblade, er aldrig blevet brugt til noget som helst (se Bergenholtz/Pedersen 1994:165). Andre korpora er efter en del udlugning af ikke-faglige, altså almensproglige tekster, blevet brugt under udarbejdelse af en bilingval ordbog (GENTEKNOLOGISK ORDBOG). Men mængden af udarbejdede fagordbøger, der bygger på et gennemanalyseret tekstkorpus som en del af ordbogens empiriske basis, er indtil nu ikke stor (ud over den ovenfor nævnte ordbog kendes kun PUMPE-TEKNOLOGISK ORDBOG). Både tekstselektion til og omfang af korpus udgør særlige problemer, som her ikke kan behandles, der henvises til den noget mere udførlige diskussion i Bergenholtz/Pedersen (1994).

Man kan skelne mellem et flerfagskorpus (fx Siliakus 1979), et enkeltfagskorpus (fx Bergenholtz/Kaufmann 1991) og et delfagskorpus (fx Dyrberg et al. 1988). Ligesom udarbejdelse af flerfagsfagordbøger vil sammensætning og analyse af flerfagskorpora være forbundet med en lang række problemer og i mange tilfælde ikke give tilfredsstillende resultater. For enkeltfags- og delfagskorpora må der gælde følgende: "korpus bør sammensættes i samråd med fageksperter" (Terminologi-afdelingen 1987:9). Ikke bare er lægfolk ikke i stand til at overse et fags systematik og dække hele faget ind, de er heller ikke i stand til at skelne klart mellem tekster for forskellige målgrupper. Fx har Lauridsen/Riiber/Søndergård (1991) medtaget tekster fra fire grupper:

1. fra fageksperter til fageksperter
2. fra fageksperter til lægfolk
3. fra lægfolk til lægfolk
4. fra lægfolk til fageksperter

Til dette korpus har Kaufmann (1993) og Stummann (1993), begge fageksperter (molekylærbiologer) og samtidig deltagere i et fagleksikografisk projekt, haft en lang række indvendinger. Alle tekster tilordnet gruppe 1 må betegnes som fejlklassificerede, en dansk fagekspert skriver på engelsk, hvis målgruppen er fageksperter. En dansk molekylærbiolog skriver evt. på dansk for semifagfolk, dvs. for fagfolk fra nabodiscipliner. De fleste af disse tekster var tilordnet gruppe 1, men visse andre gruppe 2. Til gruppe 3 og 4 indvendte molekylærbiologerne, at det for dem var umuligt at se forskel på disse grupper af tekster. I øvrigt kunne de ikke betegne lægfolks udsagn om genteknologi som fagtekster, men som almensproglige tekster.

Fageksperter bør dog ikke kun medvirke ved sammensætning af korpus, deres medvirken er i lige så høj grad en nødvendighed under arbejdet med korpus, hvis leksikografen ikke selv har en vis grad af fagkompetence. Der må dog her skelnes mellem forskellige problemstillinger. Hvor en lingvist uden større problemer vil kunne foretage en fleksionsmorfologisk analyse af et fagsprogskorpus, vil et vist kendskab til faglige sammenhæng være nødvendigt ved selektion af kollokationer, og et temmeligt stort kendskab til faget under udvælgelse af evt. eksempler. Optimalt vil det her være, at der under hele arbejdet med sammensætning og analyse af korpus til stadighed foregår et tæt samarbejde mellem en leksikograf og en fagekspert, således at intet arbejdsstrin udføres alene af en fagekspert, men heller intet alene af en leksikograf (men man kunne dog forestille sig, at leksikografen både er fagsprogsekspert og fagekspert).

7. Analyse af korpus

Under hensyntagen til 60ernes og 70ernes tekniske muligheder var et korpus på en mio. tekstord sikkert et "large corpus" (Francis 1979). Men det er nu næsten helt rørende, når en ordbog påberåber sig sin store nøjagtighed under henvisning til et sådant lille polyfunktionalt deltekstkorpus, fx i forordet til den første udgave af AMERICAN HERITAGE DICTIONARY. Med en sådan basis vil man selvfølgelig kunne få nogen hjælp ved en almensproglig ordbog, men dette korpus kan helt sikkert ikke være en tilstrækkelig empirisk basis i alle problematiske sammenhæng: betydningsangivelser, grammatiske angivelser, kollokationsangivelser, oplysninger om ortografiske varianter osv. osv. Men også ordbøger, som bygger på et større korpus, har tilsyneladende det problem, at korpus ganske vist foreligger, men at de pågældende leksikografer af forskellige grunde ikke har brugt det. Det

er blevet sagt eksplicit i forbindelse med TRÉSOR DE LA LANGUE FRANÇAISE, at det i praksis i de fleste tilfælde kun har været muligt at tage højde for mindre end en procent af tekstmaterialet. Hvor stor en del af korpus, der er blevet brugt af korpusbaserede ordbøger, kan vi ikke vide. Principielt kan man mest naturligt læse forteksten til ordbøger, der henviser til det brugte korpus, sådan, at korpus er blevet inddraget under hele det leksikografiske arbejde. Den leksikografiske virkelighed ser dog ud til at være en anden – i hvert fald, hvis Rösel's (1995) analyser gælder for mere end de undersøgte kollokationer: De korpusbaserede ordbøger indeholder kun få og i forhold til Rösels tekstmateriale tilsyneladende tilfældigt udvalgte og slet ikke alle vigtige kollokationer. Den ordbog, som har de mest relevante og også de fleste kollokationer er den ikke-korpusbaserede BBI-DICTIONARY.

Nu er det selvfølgelig ikke sådan, at meget på alle måder er bedre end lidt. Det afgørende må være en hensyntagen til de forudsete ordbogsfunktioner. En L1 -> L2-oversættelsesordbog skal helt sikkert omfatte flere grammatiske angivelser end det er tilfældet for en L2 -> L1-ordbog. Når det drejer sig om kollokationer, kan man opstille lignende argumenter. Efter min mening og erfaring kommer der et yderligere argument til: Hvis mængden af relevante kollokationsangivelser til et bestemt lemma bliver for stor, fx mere end en eller måske to spalter, må man have et noget snævrere selektionskriterium end ved andre ordbogsartikler. I det sidste tilfælde kunne man fx i en bilingval ordbog i sådanne tilfælde i højere grad bruge det ellers noget problematiske kriterium: er kollokationen umiddelbart forståelig og helt problemløs at oversætte. Der er her brug for yderligere metodiske overvejelser, som i fagleksikografisk sammenhæng vil kunne bygge på Pedersen (1995), hvorimod man i almensproglig leksikografisk sammenhæng snarere vil kunne sige, hvad man ikke skal gå ud fra, nemlig den teoretisk svagt funderede og også praktisk yderst problematiske teori, som den er fremlagt af Hausmann (1985). Her bør man i højere grad gå ud fra Kjellmer (1982) og Sinclair (1991). Dog i modsætning til den fuldstændige optagelse af alle kollokationer i KJELLMER vil en relatering til brugergruppe og ordbogsfunktion kunne være medvirkende til den nødvendige selektion. Den tilsvarende problematik gør sig ligeledes gældende ved valg af eksempler og ved valg hhv. fravalg af bl.a. de ortografiske, grammatiske, dialektale og stilistiske varianter, som forekommer i korpus.

8. Litteratur

8.1 Ordbøger

- COBUILD = *Collins COBUILD English Language Dictionary*. Editor in Chief: John Sinclair, Managing Editor: Patrick Hanks. London/Glasgow: Collins 1987.
- DUDEN-GWB = *Duden. Das große Wörterbuch der deutschen Sprache in sechs Bänden*. Hrsg. u. bearb. vom Wissenschaftlichen Rat und den Mitarbeitern der Dudenredaktion unter Leitung von Günther Drosdowski. Mannheim/Wien/Zürich: Bibliographisches Institut. Bd. 1 A–Ci 1976, Bd. 2 Cl–F 1976, Bd. 3 G–Kal 1977, Bd. 4 Kam–N 1978, Bd. 5 O–So 1980, Bd. 6 Sp–Z 1981.
- GENTEKNOLOGISK ORDBOG = Uwe Kaufmann/Henning Bergenholtz: *Genteknologisk ordbog. Dansk-engelsk/engelsk-dansk molekylærbiologi og DNA-teknologi*. København: Gad 1992.
- KJELLMER = Göran Kjellmer: *A Dictionary of English Collocations. Basis on the Brown Corpus. In Three Volumes*. Oxford: Clarendon Press 1995.
- PUMPETEKNOLOGISK ORDBOG = Jette Pedersen: *A Grundfos Basic Dictionary of Pump Technology and Related Terminology*. Århus/Bjerringbro 1995.

8.2 Anden litteratur

- Bausch, Karl-Heinz 1979: Intuition und Datenerhebung in der Linguistik. Zur pragmatischen Basis linguistischer Methodologie. I: Bergenholtz/Schaeder 1979, 71–88.
- Behaghel, Otto 1923: *Deutsche Syntax. Eine geschichtliche Darstellung. Bd. 1: Die Wortklassen und Wortformen. A: Nomen. Pronomen*. Heidelberg: Winter.
- Bergenholtz, Henning 1985: Kongruenz der Apposition. I: *Beiträge zur Geschichte der deutschen Sprache und Literatur (Tübingen)* 107, 21–44.
- Bergenholtz, Henning 1988: DK87: Et korpus med dansk almensprog. I: *Hermes* 1, 229–237.
- Bergenholtz, Henning/Uwe Kaufmann (eds.) 1991: *Gene Technology Corpus*. Århus/København.
- Bergenholtz, Henning/Joachim Mugdan 1989: Korpusproblematik in der Computerlinguistik: Konstruktionsprinzipien und Repräsentativität. I: *Computational Linguistics. Computerlinguistik. An International Handbook on Computer Oriented Language Research and Applications. Ein internationales Handbuch zur computerunterstützten Sprachforschung und ihrer Anwendungen*, hrsg. von István

- S. Bátori/Winfried Lenders/Wolfgang Putschke. Berlin/New York: de Gruyter, 141–149.
- Bergenholtz, Henning/Joachim Mugdan 1990: Formen und Probleme der Datenerhebung II: Gegenwartsbezogene synchronische Wörterbücher. I: *Wörterbücher. Dictionaries. Dictionnaires. Ein internationales Handbuch zur Lexikographie. Zweiter Teilband*, hrsg. von Franz Josef Hausmann/Oskar Reichmann/Herbert Ernst Wiegand/Ladislav Zgusta. Berlin/New York: de Gruyter, 1611–1625.
- Bergenholtz, Henning/Burkhard Schaefer 1977: Deskriptive Lexikographie. I: *zeitschrift für germanistische linguistik* 5, 2–33.
- Bergenholtz, Henning/Burkhard Schaefer (Hrsg.) 1979: *Empirische Textwissenschaft. Aufbau und Auswertung von Text-Corpora*. Königstein/Ts.: Scriptor.
- Bergenholtz, Henning/Jette Pedersen 1994: Zusammensetzung von Textkorpora für die Fachlexikographie. I: Burkhard Schaefer/Henning Bergenholtz (Hrsg.): *Fachlexikographie. Fachwissen und seine Repräsentation in Wörterbüchern*. Tübingen: Narr, 161–176.
- Bungarten, Theo 1979: Das Korpus als empirische Grundlage in der Linguistik und Literaturwissenschaft. I: Bergenholtz/Schaefer 1979, 28–51.
- Dyrberg, Gunhild/Dorrit Faber/Steffen Leo Hansen/Joan Tournay 1988: Etablering af et juridisk tekstkorpus. I: *Hermes* 1, 209–227.
- Francis, W. Nelson 1979: Problems of Assembling and Computerizing Large Corpora. I: Bergenholtz/Schaefer 1979, 110–123.
- Glas, Reinhold 1975: Das LIMAS-Korpus, ein Textkorpus für die deutsche Gegenwartssprache. I: *Linguistische Berichte* 40, 63–66.
- Greenbaum, Sidney 1984: Corpus Analysis and Elicitation Tests. I: *Corpus Linguistics. Recent Developments in the Use of Computer Corpora in English Language Research*, ed. by Jan Aarts and Willem Meijs. Amsterdam: Rodopi, 193–201.
- Hausmann, Franz Josef 1985: Kollokationen im deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels. I: Henning Bergenholtz/Joachim Mugdan (Hrsg.): *Lexikographie und Grammatik. Akten des Essener Kolloquiums 1984*. Tübingen: Niemeyer, 118–129.
- Hausmann, Franz Josef 1989: Kleine Weltgeschichte der Metalexikographie. I: *Wörterbücher in der Diskussion. Vorträge aus dem Heidelberger Lexikographischen Kolloquium*, hrsg. von Herbert Ernst Wiegand. Tübingen: Niemeyer, 75–109.

- Itkonen, Isa 1976: Was für eine Wissenschaft ist die Linguistik eigentlich? I: Dieter Wunderlich (Hrsg.): *Wissenschaftstheorie der Linguistik*. Kronberg: Athenäum, 56–76.
- Johansson, Stig 1979: The Use of a Corpus in Register Analysis: The Case of Learned and Scientific English. I: Bergenholtz/Schaeder 1979, 281–293.
- Kaufmann, Uwe 1993: Anvendelse af det danske genteknologiske tekstkorpus ved udarbejdelsen af Genteknologisk ordbog, med specielt henblik på udvælgelsen af eksempler. I: Gert Engel (red.): *Proceedings af seminar om korpuslingvistik i fagsprogsforskningen. Hindsgavl Slot 26. og 27. nov. 92*. [Kolding], 56–68.
- Kjellmer, Göran 1982: Some problems in relation to the study of collocations in the Brown Corpus. I: Stig Johansson (ed.): *Computer Corpora in English Language Research*. Bergen: Norwegian Computing Centre for the Humanities, 25–33.
- Lauridsen, Ole/Theis Riiber/Henning Søndergaard 1991: Erstellung eines dänischen und eines deutschen Textkorpus – Fachsprache der Gentechnik. I: *Hermes 6*, 125–137.
- Maegaard, Bente/Hanne Ruus 1980: Danske almindelige ord: rangfrekvenslister og deres brug. I: *SAML 1*, 5–22.
- Mentrup, Wolfgang 1979: Überlegungen zur Zusammenstellung und Verwendung eines Korpus für ein großes interdisziplinäres Wörterbuch der deutschen Sprache. I: Bergenholtz/Schaeder 1979, 182–203.
- Mergenthaler, Erhard 1979: Das Textkorpus in der psychoanalytischen Forschung. I: Bergenholtz/Schaeder 1979, 131–147.
- Pasch, Renate 1992: Es lebe das lexikographische Beispiel! (Probleme der lexikographischen Beschreibung wahrheitsfunktionaler Satzverknüpfers mit Kontextbeschränkungen. I: *Lexikontheorie und Wörterbuch. Wege der Verbindung von lexikologischer Forschung und lexikographischer Praxis*, hrsg. von Ursula Brauße/Dieter Viehweger. Tübingen: Niemeyer, 245–293.
- Pedersen, Jette 1995: The Identification and Selection of Collocations in Technical Dictionaries. I: *Lexicographica 11*, 50–73.
- Quirk, Randolph/Jan Svartvik 1979: A Corpus of Modern English. I: Bergenholtz/Schaeder 1979, 204–218.
- Renouf, Antoinette 1991: The Establishment and Use of Text Corpora at Birmingham University. I: *Hermes 7*, 71–80.
- Rieger, Burkhard 1979: Repräsentativität. Von der Unangemessenheit eines Begriffs zur Kennzeichnung eines Problems linguistischer Korpusbildung. I: Bergenholtz/Schaeder 1979, 52–70.
- Rösel, Petr 1995: Kollokationen und Sublemmabestand. Ist- und Soll-Stand in englischen monolingualen und in bilingualen Wörterbü-

- chern der Sprachrichtung Englisch-Deutsch. I: *Lexicographica* 11, 172–195.
- Siliakus, Hendricus Johannes 1979: In Search of a Common Vocabulary for the Social Sciences and the Humanities – a Report. I: Bergenholtz/Schaeder 1979, 148–170.
- Sinclair, John 1991: *Corpus. Concordance. Collocation*. Oxford: Oxford University Press.
- Stumman, Bjarne 1993: Anvendelsesmuligheder og fagligt indhold af det danske genteknologiske tekstkorpus. I: Gert Engel (red.): *Proceedings af seminar om korpuslingvistik i fagsprogsforskningen. Hindsgavl Slot 26. og 27. nov. 92*. [Kolding], 69–74.
- Terminologiafdelingen 1987: *Pilotprojekt vedrørende database til terminologisk information og generering af ordbøger*. København: Handelshøjskolen i København.
- van de Velde, Roger G. 1979: Probleme der linguistischen Theoriebildung einer empirischen Textwissenschaft. I: Bergenholtz/Schaeder 1979, 10–27.
- Wiegand, Herbert Ernst 1977: [referat af Bergenholtz/Schaeder 1977]. I: *Zeitschrift für germanistische linguistik* 5, 129.
- Wolski, Werner 1986: Partikeln im Wörterbuch. Eine Fallstudie am Beispiel von *doch*. I: *Lexicographica* 2, 244–270.
- Zöfgen, Ekkehard 1986: Kollokation – Kontextualisierung – (Beleg-) Satz. Anmerkungen zu Theorie und Praxis des lexikographischen Beispiels. I: A.Barrera-Vidal/H.Kleineidam/M.Raupach (Hrsg.): *Französische Sprachlehre und bon usage. Festschrift für Hans-Wilhelm Klein zum 75. Geburtstag*. München: Hueber, 219–238.

Eiríkur Rögnvaldsson

Brugen af et gammelislandsk tekstkorpus i leksikografisk arbejde

In this paper I give a short description of a project that I have been engaged in for the last few years. This is a corpus of Old Icelandic texts, and an analysis of the vocabulary of these texts. I will describe the corpus and a new concordance to the Icelandic Family Sagas, which is based on the corpus. I will also point out several problems which we are bound to encounter, if we try to make a lexicographic description which is exclusively based on this corpus.

1. Tekstkorpuset

I nogle år har jeg og nogle kolleger arbejdet på et korpus med gammelislandske tekster (se Eiríkur Rögnvaldsson og Örnólfur Thorsson 1989). Dette arbejde begyndte med udgivelsen af de islandske sagaer med moderne islandsk retskrivning i årene 1985–1986. I denne udgave findes der omkring 40 sagaer, nogle af dem i to versioner, og desuden omkring fyre kortere fortællinger, såkaldteflættir. I de sidste ti år har vi tilføjet Sturlunga saga, Heimskringla og Grágás, som er blevet tastet ind i forberedning af nye udgaver. Desuden har vi skannet ind Landnámabók og de fire grammatiske afhandlinger i Snorris Edda, og nogle få andre tekster. Dermed har vi fået en væsentlig del af alle opbevarede gammelislandske prosatekster. Korpusets størrelse er nu næsten to millioner ord, og næsten 60.000 forskellige ordformer. Vi savner dog helt eller delvis nogle litterære genrer. Vi har fx kun tre sagaer om hellige mænd og kvinder, to fornaldarsögur, og ingen af bispesagaerne.

Et sådant tekstkorpus kan bruges til mange forskellige formål. Indtil nu har det været tilgængelig i WordCruncher format, så at det har været muligt at søge efter enkelte ord eller kombinationer af flere ord i hele korpuset på én gang. Man kan fx bruge korpuset til at finde ud af, hvor i teksten enkelte emner behandles, og på denne måde har forskere fra forskellige fagområder allerede haft en stor gavn af korpuset. Men korpuset er også blevet benyttet til forskellig lingvistisk forskning. For det første kan man nævne frekvensstudier (Eiríkur Rögnvaldsson 1990, Örnólfur Thorsson 1993), for det andet syntaktiske studier (Eiríkur

Rögvaldsson 1991, 1994–95, 1995a), og til sidst leksikografisk beskrivelse.

Brugen af korpuset er også afhængig af de konkordanser, som er blevet udarbejdet på dets grundlag. Det er kun konkordansen til de islandske sagaer som er helt færdig, men konkordanser til Sturlunga, Heimskringla og Landnáma er næsten færdige. Konkordansen til de islandske sagaer er langt den største – større end de andre tilsammen. Vi har brugt WordCruncher til at lave grundlaget til konkordansen, men det meste af arbejdet måtte dog gøres manuelt. Dette er en lemmatiseret KWIC-konkordanse, hvor alle eksempler om hver enkelt ordform er alfabetisk ordnet efter det efterfølgende ord i teksten. I (1) kan man se et eksempel fra konkordansen.

(1)

abbast so; abba>ist (1)

Njála 124;275

Skarphé>inn hló a> og spur>i hví hún **abba>ist** upp á arfasátuna. Kerlingin mælti

a>alborinn lo; a>albornir (1)

Egla 9;378

fram a> flytja, a> vi> séum menn **a>albornir.**" fiórólfur svarar flá stygglega:

a>alfestur kv; a>alfestur (1)

Njála 144;313

festum helganda nema fjörbaugur e>a **a>alfestur** komi fram a> féránsdómi, en alsekjan

a>allega ao; a>allega (1)

Grett 81;1078

fram á fjör>inn. En svo sem fleir komu **a>allega** fram á fjör>inn og á djúpi> hæg>ist

a>alskáli kk; a>alskálann (1)

Eyrb 43;591

flá inn í forhúsi>. En er hann gekk í **a>alskálann** vildi hann fara hljó>lega flví a> hann

a>djúp hk; a>djúp (1)

Hávís 9;1316

og hreinsa allt inn a> landi. Var flar **a>djúp** miki>. Mátti flar fljóta skúta e>a

a>djúpur lo; a>djúpt (1)

Krók 14;1528

á framanver>an sævarbakkann. fiar var **a>djúpt** mjög og ekki útfiri. fiessi skur>ur var

a>dráttama>ur kk; sjá a>dráttarma>ur kk

a>dráttarma>ur kk; a>dráttama>ur (1); **a>dráttarma>ur** (2)

Eyrb 11;543

sex tigu frelsingja. Hann var mikill **a>dráttama>ur** og var jafnan í fiskiró>rum. Hann lét

Fljót 1;675

og í öllu ójafna>arma>ur. Hann var **a>dráttarma>ur** mikill a> búinu bæ>i af fjör>um ne>an

Grett 25;994

a> Lækjamóti og bjó flar. fiorgils var **a>dráttarma>ur** mikill og fór á Strandir hvert ár.

a>dráttur kk; a>drátt (1); **a>drátta** (1); **a>dráttar** (1)

Hávís 2;1301

ma>ur og hinn frálegasti. Haf>i hann Grett 72;1064	a>drátt a> búi fleirra. fia> var eitthvert
a> hann vildi fara til lands til Grett 16;972	a>drátta en fleir Illugi og Glaumur skyldu eftir
Ásmundur, "fyrir sakir starfa og	a>dráttar en Grettir vill ekki starfa. Er hann

a>drætti hk; a>drætti (1)

Hrafn 11;1410

hin fyrstu misseri. Hann haf>i miki> **a>drætti** af fiskimönnum. Hrafnkell gekk mjög a>

Konkordansen til de islandske sagaer har i de sidste par år været tilgængelig på en computer på Lingvistisk institut ved Islands Universitet. Nu har bogforlaget Mál og menning lige udgivet teksten til alle sagaerne sammen med den lemmatiserede konkordans på CD-ROM. Desuden findes der forskellige ordlister på CD-en, fx en liste hvor alle substantiver i teksten er semantisk klassificeret. Der er forbindelser mellem teksten og konkordansen, så at det er muligt at fx klikke på et bestemt ord i teksten for at se alle forekomster af dette ord i konkordansen. Man kan også vælge et bestemt ord i konkordansen og se det relevante sted i teksten.

Som jeg har nævnt, har konkordansen til de islandske sagaer allerede opfordret folk til mange forskellige studier i det gamle islandske sprog. Jeg vil nu prøve at beskrive den leksikografiske del af disse studier lidt nærmere.

2. Leksikografisk beskrivelse

I de sidste år har vi arbejdet på en leksikografisk beskrivelse af ordforrådet i sagaerne. Vores mål er at notere alt, som vi synes er interessant ved hvert enkelt ord, i bøjning, syntaktisk position, betydning, distribution o.s.v. Jeg har faktisk kun været konsulent i dette arbejde; det er Gu>rún Ingólfssdóttir og Bergljót Kristjánsdóttir som er ansvarlige for beskrivelsen. Hvad jeg kommer at sige herefter er for det meste baseret på deres erfaringer og observationer; se også Gu>rún Ingólfssdóttir (1995).

Bearbejdelsen af en leksikografisk beskrivelse for ældre stadier af sproget er noget forskellig fra beskrivelsen af et moderne, levende sprog. Når man arbejder med ældre sprogstadier, er man helt afhængig af de opbevarede tekster; man kan ikke bruge sin egen sproglige intuition eller spørge informanter. Men selv om beskrivelsen altid baseres på tekster, kan resultatet naturligvis blive forskelligt, afhængig af om man baserer beskrivelsen på et seddelarkiv eller om man benytter et tekstkorpus.

De opbevarede tekster er naturligvis begrænsede både i mængde og typer. Som jeg allerede har nævnt er størstedelen af gammelislandske tekster en slags narrativ litteratur, og de handler for det meste om temmelig ensidige emner. Dette afsløres ganske tydeligt når man ser på en liste over de hyppigste ord i sagaerne. Det allerhyppigste substantiv er *mand*, og de næste er *konge* og *skib*. Andre hyppige substantiver er fx *søn*, *bror*, *far*, *datter*, *mor*; *hånd*, *hoved*, *fod*; *dag*, *sommer*, *nat*, *morgen*, *aften*, *forår*, *vinter*, *sværd*, *spyd*, *skjold*. Blandt de hyppigste adjektiver finder man ord, som ofte bruges, når mænd beskrives: *stærk*, *ung*, *gammel*, *klog*, *populær*, *smuk*, *rig*, *udmærket* o.s.v. Hyppige verber er fx *sige*, *svare*, *spørge*, *tale*; *komme*, *gå*, *ride*, *stå*, *sidde*, *løbe*; *dræbe*, *hugge*.

Man siger ofte at de islandske sagaer, Sturlunga og Heimskringla er skrevet i den såkaldte folkelige stil. Der eksisterer også gammel litteratur som er skrevet i lærd stil. Det er navnlig sagaer om hellige mænd, homilier og den slags litteratur, som enten er blevet oversat fra latin eller er under stærk indflydelse fra latinsk litteratur. Når man bygger op et gammelislandsk tekstkorpus må man derfor være sikker på at de vigtigste typer er repræsenteret. Men hvis man sammenligner gammelislandske tekster med nutidssproget er det naturligvis klart at et gammelislandsk korpus bliver stærkt begrænset. Vi må huske at den skriftlige tradition ser ud til at have været meget stærk. De fleste narrative tekster anses for at være skrevet i det trettende og fjortende århundrede, selv om mange af dem handler om begivenheder fra det niende og tiende århundrede. Det er da klart at de er mere eller mindre baseret på mundtlig tradition, og selv om der er delte meninger, om hvorvidt sagaerne eksisterede, før de blev skrevet, er det klart, at dette har en væsentlig indflydelse på deres ordforråd og sprogbrug.

Derfor er det ofte meget svært at evaluere tekstenes vidnesbyrd. De eksempler af et bestemt ord som man finder i teksterne behøver slet ikke være typiske for ordets brug i gammelislandsk – og vi har ingen mulighed for at vide om de er. Spørgsmålet er derfor hvor meget vi kan konkludere om ordets betydning og brug, udover det som eksemplerne viser direkte. Der findes ni og tyve eksempler af ordet *sandur* 'sand' i sagaerne, men ikke ét af dem er i nominativ. Spørgsmålet er så: Er denne distribution noget der bør nævnes i en leksikografisk beskrivelse? Næppe, fordi eksemplerne ikke har noget andet tilfælles. Det må anses som rent tilfældigt, at der ikke forekommer en sætning, hvor *sandur* er subjekt eller prædikat. Situationen er helt anderledes med ord som fx *skotfæri* og *litklæ>i*. Disse ord forekommer heller ikke i nominativ, men eksemplerne har også andre ting tilfælles; i alle

eksempler styres ordene af præpositionen *í*. Dette er noget der bør nævnes.

Men der er også andre faktorer, som gør situationen endnu mere problematisk. Størstedelen af både gammelislandsk ordforråd og bøjningssystem er blevet bevaret i nutidssproget, og derfor er det relativt let for nutidens Islændinge at læse gamle tekster, i det mindste sagaerne. Men dette indebærer den fare, at vi uden betænkning fastslår, at et ord som vi finder i en gammel tekst, og som vi kender fra vores sprog, betyder det samme i den gamle tekst, som det gør i nutidssproget, og virker på samme måde, både syntaktisk og morfologisk. Derfor fristes vi somme tider til at fastslå noget, som faktisk ikke har nogen direkt støtte i opbevarede tekster - noget som ved nærmere granskning viser sig at være fejl.

Ét eksempel af denne art er ordet *skipverji*. I islandsk nutidssprog betyder det 'en af Besætningen på et Skib; Matros ...', ifølge Sigfús Blöndals ordbog fra 1920. I sagaerne findes der seks og fyrre eksempler af dette ord. Når man studerer disse eksempler, lægger man mærke til, at i næsten alle af dem bliver der talt om *skipverja sína*, *skipverja hans* o.s.v. Dette stemmer også med Fritznens definition; 'Mand som er sammen med en (*e-s*) på et Fartøi ...'. I nutidssproget har ordet altså en mere almen betydning, men de fleste Islændinge vil ikke lægge mærke til forskellen når de læser sagaerne. Og der findes vældig mange eksempler af denne slags.

3. Kollokationer

Det som var mest iøjnefaldende når vi begyndte at arbejde med konkordansen til de islandske sagaer var alle slags kollokationer, som vi ikke havde haft viden om. I Fritznens ordbog bliver kollokationer sjældent nævnt. Naturligvis kan man ofte se mønstre i Fritznens eksempler, men siden han ikke siger noget om disse mønstre, kan man ikke vide, hvor typiske hans eksempler er. Jeg har allerede beskrevet nogle interessante kollokationer i sagaerne (Eiríkur Rögnvaldsson 1995b), men her kan jeg tilføje nogle andre. Men det er naturligvis klart at selv om man finder den samme ordforbindelse på nogle steder er det ikke altid af lingvistisk interesse. Det kan fx godt ske at én saga låner fra en anden. Man må også huske at de fleste af sagaerne skildrer den samme slags aktiviteter, og det er derfor helt naturligt at de samme ord bruges i beskrivelsen.

Adjektiverne *gullrekinn* og *silfurrekinn* er vældig interessante. Ifølge Fritznen betyder de 'indlagt med guld' og 'indlagt med sølv'. Man

kunne måske vente at forskellige ting kunne være indlagt med guld eller sølv. Men det viser sig, at det næsten udelukkende er økser, der er indlagt med sølv, og i langt de fleste eksempler om *gullrekinn* er det spyd der er indlagt med guld, selv om der også findes nogle få økser. Der er kun ét sværd der er *gullreki*, selv om sværd synes at have været de hyppigste våben i sagatiden.

(2)

gullrekinn lo; gullreki> (11); gullrekin (1); gullrekinn (3); gullrekna (6); gullreknu

(2)

Njála 86;222

gaf jarl Kára sver> gott og spjót
Laxd 37;1587grám feldi og haf>i í hendi bryntröll
Njála 84;220og fagurt. Sjá ma>ur haf>i spjót
fiór> 8;2030um jólin. Hann haf>i gefi> fiór>i spjót
Njála 68;203Starka>arson nafna sínum spjót
Laxd 21;1565Ólafi til skips og gaf honum spjót
Gullfi 8;1127veitir flá snarpa atgöngu. Hann haf>i
Eyrb 13;545Hann haf>i búi> sver> og
HallM 4;1198feir settu úti spjót sín. Grís átti
HallÓ 2;1226en settu úti spjót sín. Grís átti
VaLjó 8;1839flá var>veitti hann spjóti>. fia> var
Laxd 21;1563Hann var gyr>ur sver>i og voru
Flóam 9;733sver>i og haf>i spjót miki> í hendi og
Laxd 44;1604helgi. Hann haf>i í hendi spjót og
Egla 84;509í hendi krókaspjót, var flar
fiórSH 1;2061og drengilega og fligg af mér eina öxi
VígGl 8;1917Hann tók flá feldinn blá og spjóti>
Laxd 24;1570Ólafur tekur í hönd sér spjóti>
Vatn 43;1897jarli og hann gaf honum öxi
VígGl 25;1943Gissuri feldinn blá en Ásgrími spjóti>
Laxd 29;1575En a> skilna>i gaf jarl honum öxi
VígGl 6;1913í skautfeldi blám og lék sér a> spjóti
fiórhv 7;2059**gullreki**> en Helga gullhring og skikkju en Grími**gullreki**> er Haraldur konungur gaf**gullreki**> í hendi. Hann spur>i: "Hverjir**gullreki**> og heiti> honum sínu li>i hvar sem**gullreki**> og rei> heim sí>an. Ger>u fleir me> sér**gullreki**> og sver> búi> og miki> fé anna>. Ólafur**gullreki**> spjót í hendi. Hann hleypur í flokk**gullreki**> spjót, myrkblán skjöld og mjög gylltan,**gullreki**> spjót. Nú sátu fleir a> málunum og**gullreki**> spjót. feir tóku tal sitt og flutti Már**gullreki**>. feir spur>u hva>an honum kæmi fla>**gullrekin** hjöltin. Hann haf>i krókaspjót í hendi**gullrekinn** á falurinn. feir fe>gar höf>u**gullrekinn** falurinn á. Allir menn hans voru í**gullrekinn** falurinn. Hann var sver>i gyr>ur. fiar**gullrekna**. Hana sómir flér a> bera."**gullrekna** í hönd sér, lét sö>la hest sinn.**gullrekna**, konungsnaut, gengur nú heiman og**gullrekna** og gó> klæ>i og kva>st vera skyldu**gullrekna** og skildust vinir. Um veturinn**gullrekna** og var fla> hin mesta gersemi, skildust**gullreknu**, gekk sí>an a> honum og kvaddi hann en

lék sveinninn Helgi fiorgilsson sér a> **gullreknu** spjóti er fiørsteinn fagri haf-i sett

(3)

silfurrekinn lo; silfurrekin (1); silfurrekinn (1); silfurrekna (5)

Svarf 24;1817

kasta> ni>ur glófum og flar lá hjá öx

silfurrekin. Karl tekur upp og leggur í kné sér

Grett 48;1027

í hendi og öngvir krókarnir á og var

silfurrekinn falurinn á. Hann settist ni>ur og drap

Gunnl 8;1178

vel ortur. Jarl gaf honum brei>öxi,

silfurrekna alla flar er bæta flótti, a> kvæ>islaunum

Njála 138;299

her>um sér og gullhla> um höfu> og öxi

silfurrekna í hendi. Bjarni mælti: "Hér ber

Njála 147;325

rei> í blárrí kápu og haf-i litla öxi

silfurrekna í hendi. En er fleir komu í túni> flá

HallÓ 4;1230

og gaf honum klæ>i gó> og mikla öxi

silfurrekna og bau> honum me> sér a> vera um

HallM 5;1201

flakka>i honum og gaf honum exi mikla

silfurrekna og gó> klæ>i og bau> honum me> sér a>

Det er naturligvis muligt, at spyd var de fornemste våben, og derfor var det spydene, der var indlagt med guld. Men så må man tage hensyn til at sagaerne ikke blev skrevet ned før tre eller fire hundrede år efter sagatiden. Derfor skildrer de sandsynligvis ikke våbnene som de var i sagatiden, men som skriverne tænkte sig, at de var. Det vil sige at for skriverne i det trettende eller fjortende århundrede var *gullrekinn* et passende adjektiv for spyd, og *silfurrekinn* et passende adjektiv for økse. Derfor er dette efter min mening blevet et faktum om sproget, men ikke kun et faktum om samfundet. Hvis dette er sådan, tror jeg, at disse oplysninger bør findes i en leksikografisk beskrivelse.

Jeg kan blot nævne nogle forskellige typer af kollokationer. Adjektivet *sí>búinn* står i 11 af 12 eksempler med verbet *ver>a* 'blive', og tager ofte adverbiet *heldur*. Substantivene *skartsma>ur* og *skraut-menni* tager altid adjektivet *mikill*, enten i positiv eller superlativ. Adverbiet *stórilla* står næsten altid enten med verbet *líka* eller verbet *una*. Substantivet *skotfæri* styres altid af præpositionen *í*, og står altid med verbet *koma*. Og der findes mange flere typer, og mange flere eksempler af hver type (se også Eiríkur Rögnvaldsson 1993).

4. Semantisk beskrivelse

En væsentlig del af de islandske sagaers ordforråd er ord, der kun forekommer én gang. Når det kommer til en leksikografisk beskrivelse af disse ord, gør det sandsynligvis ingen forskel, om beskrivelsen base-res på et seddelarkiv eller direkt på et korpus. Men i de hyppige ord kan dette gøre en stor forskel. Det er ikke sandsynligt, at alle forekomster af

disse ord findes i seddelarkivet, og den som kan benytte sig et tekstkorpus ved beskrivelsen har derfor meget bedre oversigt over disse ord. Det kan ændre beskrivelsen på to måder. På den ene side er det muligt at få en nøjagtigere beskrivelse, siden den kan baseres på flere eksempler. Man kan fx tage hensyn til detaljer i betydning eller syntaktisk position som kun forekommer i ét eller få eksempler, og som måske ikke vil være repræsenteret i seddelarkivet.

På den anden side kan det ændre beskrivelsen betydeligt at generaliseringer bliver unødvendige. Når leksikografen har adgang til alle eksemplerne kan han basere sin beskrivelse på netop disse eksempler, og behøver ikke at tage hensyn til andre eksempler som muligvis kunne findes i tekster. Dette kan være af en stor betydning, fx når man skal lave en semantisk beskrivelse. Vi kan fx tænke os et adjektiv som ser ud til at have en generel betydning. I alle vores eksempler i seddelarkivet står det vistnok kun med nogle få substantiver, men leksikografen tør ikke tage chancen at lade det være en del af beskrivelsen, siden han ikke kan garantere, at der ikke findes anden slags eksempler i teksterne, hvor ordet også står med andre typer af substantiver. Men når man har et korpus og kan basere beskrivelsen på alle eksemplerne viser det sig at der ikke findes nogen eksempler af en anden type. Så kan man lade denne begrænsning være en del af ordets leksikografiske beskrivelse.

For at tage et konkret eksempel kan vi se på ordet *i>ur* 'indvolde':

(4)

i>ur hk ft; i>rin (12); i>runum (2); i>urin (1)

GíslS 36;896

til hans me> spjótum svo a> út falla
Fljót 17;706

i>rin en hann sveipar a> sér i>runum og

hann af sér klæ>um. fiá falla út
GíslL 27;951

i>rin. Hann sest flá ni>ur og lét fiorkell flar

feir særa hann á hol svo a> út falla
Laxd 49;1614

i>rin. Hann sveipar nú a> sér i>runum me>

fló barist um hri> svo a> úti lágu
LjósA 3;1729

i>rin. Í flessi svipan hjó Kjartan fót af

einn. Hann var eigi a> óákafari fló a>
LjósC 19;1688

i>rin lægju úti. Gu>mundur hopa>i undan og

einn. Hann var eigi a> óákafari fló a>
LjósC 19;1688

i>rin lægju úti. Gu>mundur hopa>i undan og

vi> ef flú florir flví a> nú liggja úti
LjósA 3;1729

i>rin mín. fiar hefir flú jafngjarn á veri> er

flú nú hinga> Gu>mundur. Úti liggja nú
Gullfi 15;1136

i>rin mín." Sí>an drápu fleir hann.

renndi ofan í kvi>inn svo a> út féllu
Grett 66;1057

i>rin og létust fleir flar bá>ir fe>gar af

alla bringspalina og kvi>inn svo a>
Laxd 55;1621

i>rin steyptust úr honum ofan í ána og

hélt a> sér kyrtlinum a> eigi hlypu út
Hei> 30;1384

i>rin. fiá hljóp Steinflór Ólafsson a> Bolla

fióroddi og fellur hann og liggja úti

i>rin. fiorbjörn sér nú frændur sína, hir>ir

GíslL 27;951	falla i>rin. Hann sveipar nú a> sér	i>runum me> skyrtnunni og bindur svo a> sér me>
GíslS 36;896	út falla i>rin en hann sveipar a> sér	i>runum og skyrtnunni og bindur a> fyrir ne>an
Svarf 5;1786	bringspölu og hljóp á hol svo út féllu	i>urin . Sveipa>i eg flá a> mér klæ>unum og svo

I næsten alle eksemplerne finder man kombinationen "indvoldene ligger ude" eller "indvoldene falder ud", d.v.s. ud af bughulen. Man kan da spekulere på, om dette måske er en del af ordets betydning. Det er slet ikke umuligt at *i>ur* kun brugtes i denne sammenhæng, og at hvis indvoldene var på deres plads så brugte man et andet ord. På den anden side må man tage hensyn til at sagaerne handler ikke mindst om al slags konflikter mellem personer og slægter, og disse konflikter leder ganske ofte til at nogen bliver såret eller dræbt. Derfor er det også en mulighed at *i>ur* slet ikke var begrænset som det ser ud til; årsagen kan simpelthen være at man ikke behøvede dette ord i andre sammenhæng.

Vi kan også se på ordet *kálfi*, som ifølge Fritzner betyder 'Indlæggen, det tykke Kjød på Læggens Bagside ...':

(5)

kálfi **kk**; kálfa (1); kálfann (5); kálfi (3)

GíslL 21;928	sk̄tur flanga> til spjóti og kemur í	kálfa Gísla. Hann sendir aftur spjóti> og
GíslS 20;874	sk̄tur spjóti á gagngert og kemur í	kálfann á Gísla. Hann sendir aftur spjóti> og
GíslS 27;883	sk̄tur eftir honum spjóti og kom í	kálfann á honum og skar út úr og var> fla>
Fósth 24;834	höggi> á læri> og reist ofan læri> og	kálfann . Hljóp öxin ni>ur í flröskuldinn. Ljótur
GíslL 24;938	flá spjóti eftir honum og kemur í	kálfann og var> fla> miki> sár. Gísli kippir
Fósth 24;836	fyrir ne>an knésfótinn og beit úr	kálfann úr fætinum. Kár haltrar heim og ver>ur
Har> 39;1294	bá>a fætur, annan flar sem mjóstur var	kálfi en annan í ristarli>. Sí>an hljóp
Eyrb 18;552	svo a> fiórarinn hjó fót af fióri flar er	kálfi var digrastur en drap bá>a förunauta
Njála 145;318	til Skafta og kom fyrir ne>an fla> er	kálfi var digrastur og svo í gegnum bá>a

Som eksemplerne viser bruges *kálfi* kun i skildringer af sår. Her er situationen naturligtvis helt anderledes end med *i>ur*. Indlæggen sidder jo på sin plads selv om den er såret, og det virker højst usandsynligt at man har haft et specielt ord for at betegne en såret indlæg. Når man kikker på andre ord fra det samme semantiske område, ord som *hryggur*, *hol* o.fl., så ser man at det samme gælder for dem; de fore-

kommer næsten udelukkende i skildringer af sår. Derfor er der næppe nogen grund til at tro at *i>ur* kun kunne bruges om indvolde der ikke var på deres plads; i det mindste giver teksten os ikke lov til at hævde det.

Lad os næst se på eksemplerne om to ord for klædningsstykker; *hattur* og *brók*. Fritznens definitioner findes i (6):

- (6) *hattr*, m Hat, Hætte, = *hötr*, som det synes brugt om alle Slags Hovedbedækning, saavel særskilt, som forbunden med eller tilhørende den Overklædning, som en Person er iført.

brók, f. (G. *brókar*, N.Pl. *brækr*) Brog, Bure, Klædningsstykke som skjuler Legemet fra Beltestedet over Laarene og nedover Knæet, og bestaar af to Halvdele, der tilsammen betegnes ved Plur. *brækr*, medens derimod hver Halvdel betegnes ved Sing. *brók*, som dog ofte anvendes enstydigt med *brækr*.

Det viser sig at i alle tilfælde er det mænd, som har en hat på hovedet, men dette siger Fritznér ikke noget om. Han nævner heller ikke om *brók* bruges af mænd eller kvinder, eller måske begge køn. Men eksemplerne viser at med to undtagelser er det kun mænd som er iført *brækur*. Den ene af de to kvinder blev kaldt *Bróka-Au>ur*. I sagaen læser man, at hendes mand forlangte skilsmisse, og det er ret klart at *Au>ur* anses for at være lesbisk. Spørgsmålet er nu, om disse oplysninger skal findes i en leksikografisk beskrivelse. Er det en vigtig semantisk faktor at det kun er mænd der bruger *brækur* og *hattar*?

I (7) finder vi Fritznens definitioner af adjektiverne *skapstór* og *skapflungur*. Eksemplerne findes i (8); og det er bemærkelsesværdig at begge bruges i de fleste tilfælde i skildringer af kvinder. Husk at sagaerne handler for det meste om mænd, så denne distribution er ikke den man ville vente. Men hvad viser denne distribution; en semantisk afgrænsing eller forfatterens eller skriverens idéer om kvinder?

- (7) *skapstórr*, adj. = skapmikill ...

skapmikill, adj. af Sindsbeskaffenhed ikke som Folk i Alm., men saadan, at man hæver sig over dem baade ved de Indtryk, som en modtager af sine Oplevelser, og ved de Foretagender, hvorpaa man indlader sig ...

skapflungur, adj. modsat *skapléttr*; *e-m er skapflungt* : en er ilde tilmode, i slet Humør.

- (8)

skapstór lo; skapstór (13); skapstórir (2); skapstærri (1)

Njála 97;240

bónor>i>. Hún kva>st vera kona **skapstór** "og veit eg eigi hversu mér er hent við

Eyrb 28;568

og sköruleg, ofláti mikill og heldur **skapstór**. En er Styr fann tal fleirra flá

Egla 79;489

kona og kvenna mest, vitur og heldur **skapstór** en hversdaglega kyrrlát. Egill kunni

Fljót 6;684 vel en við þá menn var hún heldur Njála 9;136 hent," sagði Ósvífur, "hún er kona Harðv 1;1296 Hún var ekkja. Hún var skörungur og fiór 1;2008 kvenna vænust og ofláti mikill og Laxd 9;1543 að sér og mart vel kunnandi og heldur Harð 2;1254 Hún var skörungur mikill, skjótor og Fljót 10;688 gefa, svo var hún örlát. Hún var Laxd 31;1577 miklu fé. fiórur var vitur kona og Eirík 3;521 bera flessi mál upp flví að fiorbjörn er Eyrb 18;550 fleirra, drengileg kona og heldur Eyrb 25;563 Hygg eg að fleir verði flér stírir og Laxd 35;1583 við fleiri manna ráð flví að menn eru Laxd 65;1636 tjá mundu "flví að fiorgils er miklu	skapstór en fless í milli fálát og steigurlát en skapstór en flú ert harðlyndur og óvæginn." skapstór . Grímur hét son hennar er flar óx upp skapstór . Hún var allra kvenna høgust fleirra er skapstór jafnan. Vel var um samfarar fleirra skapstór og harðúig í öllu. fiar óx upp skapstór og skörungur mikill, skafinn drengur skapstór og skörungur mikill. Hallur hét son skapstór og fló metnaðarmaður mikill." skapstór . Styr var héraðríkur og hafði fjölmennt skapstórir flegar er flér kaupist við. Hygg eg flá skapstórir fleir er sér mun flykja misboði í flessu? skapstærri maður en hann muni hér að smáhlutum
---	---

skapflungur lo; skapflungt (9)

Njála 116;261 augum sér og grét. Flosi mælti: " HallM 10;1214 kemur Grís til seljanna. Kolfinnu var Fóstb 8;791 reiddist hann við er honum var áður Vopnf 13;1998 hví hún gréti eða hví henni væri svo Njála 6;132 Hann fagnaði henni vel en henni var Fljót 20;719 Snæfr hún nú til rekkju sinnar og er HallÓ 11;1247 Hallfreður flá í brottu. Kolfinnu var Njála 11;138 frá Hallgerði að hún sat úti og var Njála 10;136 honum hvað ætlað var og var henni	Skapflungt er flér nú frændkona er flú grætur en fló skapflungt . Grís sá flá. Hann kvað skapflungt . Hann ríður flá yfir ána að Torfa og skapflungt . Hún kvaðst gráta drauma sína. skapflungt nokkuð. Og er hann fann flá mælti skapflungt sem von var til. Sofa menn af nóttina. skapflungt . flá sá Grís og kvað: skapflungt . fíjóstólfur gekk að og sá að hún skapflungt . fíjóstólfur mælti: "Ger flú flér
---	--

En anden type af eksempler kan man finde i ordene *búráð* og *búsforráð*.
Fritzners definitioner kan man se i (9), og eksemplerne i (10):

(9) *búráð*, n. Husets, Husholdningens Bestyrelse ...
búsforráð, n. = *búráð* ...

(10)

búráð hk ft; búráðum (3)

Njála 14;143

Hallgeri ef hún vildi taka við Njála 34;163	búráum "Eigi vil eg fla," segir hún.
og atkvæðamikil. Þorgerður tók við Njála 34;163	búráum a) Grjótá og var góð húsfréyjja.
flá vel. Hallgerður tók við	búráum og var fengsöm og atkvæðamikil.

búsforráð hk ft; búsforráð (1)

Grett 28;997

flá við honum sæmilega. Atli hafði flá **búsforráð**. Féll vel á með fleim bræðrum. flá

Ifølge Fritznens er disse to ord synonymmer. Når man ser på eksemplerne viser det sig at det førstnævnte bruges om en kvinde, mens det sidstnævnte bruges om en mand. Vi ved jo at ifølge loven skulle kvinder bestyre husholdningen indendørs, mens mænd skulle bestyre al udendørs arbejde. Derfor er det slet ikke umuligt at disse to ord netop ikke er synonymmer, og det første betyder indendørs bestyrelse, men det sidste betyder udendørs bestyrelse. Se også Guðrún Ingólfssdóttir (1993).

Lad os nu se på ordene *hvíld* og *sjatna*. Fritznens definition findes i (11), og eksemplerne i (12):

(11) *hvíld*, f. Hvile, Ro i Modsætning til Arbejde og Anstrengelse ...

sjatna, v. (a) **1)** synke sammen ... **2)** fordøies ... **3)** lægge sig, falde bort, ophøre ... om *úfriðr*, *úflokki* ...

(12)

hvíld kv; hvíld (12); hvíldar (5); hvíldir (1)

GíslS 35;896

en fleir sárir og móðir. Verður nú **hvíld** á aðsókninni. fá eggjar Eyjólfur menn
Reykð 13;1750og riðu snemma aftans heiman, taka **hvíld** á Fljótsheiði gagnvert Öndólfssstöðum og
Vopnf 18;2004skammt til túngarðsins fá tóku fleir **hvíld**. Bjarni mælti: "Eigi mun eg
Njála 145;319dauður til jarðar. fá var **hvíld** nokkur á um bardagann. Snorri góði kom
Vatn 7;1851hans menn voru. Og er kveldaði var á **hvíld** nokkur bardaganum. Ingimundur
Heið 31;1386og er óvígur. Nú verður á **hvíld** nokkur og binda menn sár sín. Nú er sén
Heið 31;1385og hann fellur. Nú er á **hvíld** nokkur og nú er sén reið manna sex. flar
Fóstb 24;833hjá skipinu og kemst á kjöl, tekur flar **hvíld** og litast um og hyggur að ef hann sjái
Flóam 13;737fiorgils ofan á hann. Tekur hann flar flá **hvíld** og nær síðan öxi sinni. Höggur fiorgils
Njála 77;212og fór svo fram um hríð. fleir tóku **hvíld** og sóttu að í annað sinn. Gunnar skaut

Njála 30;157 báir um daginn. Kolskeggur tók Reykd 12;1749 eigi nþtt um fla. Og í einhverri GunKe 5;1151 og vetra>i, en flestum flótti mál FóstF 1;845 heita flér a> flú komir til nokkurar Egla 65;469 hin fyrsta hrí. fiá beiddist Ljótur GunKV 16;1164 svo a> hann mæddist og beiddist GunKe 9;1158 a> hann var ma>ur yngri og beiddi Örn Fóstb 24;837 sker flau er á lei>inni voru og tók flar	hvíld um daginn á skipi Gunnars og sér Gunnar hvíld flá er menn var>i minnst flá lystur hvíldar af löngu sjóvolki. Og var fla> af rá>ð hvíldar eftir sjöund flína og munt flú eigi mega hvíldar . Egill lét fla> og vera. Nema fleir flá hvíldar um stund og fla> veitti Gunnar honum. hvíldar . fleir hvíldust og studdust fram á vopn hvíldir . Og er skammt var til meginlands flá
--	--

sjatna so; sjatna (7)

Hei> 37;1392 vitrum mönnum flá líklegast til a> Hei> 36;1391 er vænst er a> best megi haldast og LjósA 1;1717 Ófeigur mælti: "Illa mun LjósC 1;1655 Ófeigur mælti: "Illa mun Hei> 37;1392 á fébætur. Væntu fleir a> heldur mundi Njála 92;234 allir vita a> eigi mundi svo gert Flóam 32;762 ba> hann utan fara "og megi flá	sjatna mundi fleirra ofsi, svo mikill sem var, sjatna mætti ófri>urinn." fiá sjatna ofsi fleirra bræ>ra en koma mun eg hér sjatna ofsi fleirra bræ>ra en koma mun eg hér sjatna ófri>urinn og flótti sér fla> eigi minni sjatna . Runólfur son Úlfs aurgo>a austur sjatna flessi óflokki er í millum ykkar er."
---	--

Hvis man ser nøjagtig på eksemplerne om *hvíld*, så viser det sig, at i næsten alle tilfælde er der tale om en pause i kampen. Der ser ud til, at ordets fundamentale betydning er 'våbenhvile'; men det fremgår ikke af Fritzners definition. Situationen er næsten den samme i ordet *sjatna*. I alle belæggene bruges ordet om ophør af krig eller kamp. Men i dette tilfælde bliver denne betydning nævnt i Fritzners ordbog. Man kan da spørge, om der findes en forklaring på denne forskel. Kan man sige, at i *sjatna* drejer det sig om betydning, men i *hvíld* drejer det sig om sprogbrug? Det ved jeg ikke.

Det sidste ord som jeg vil behandle er *laug*, som Fritzner definerer bl.a. på følgende måde:

(13) *laug*, f. ... 2) varm Kilde som har dannet et Bækken, Basin skikket til Badning ...

(14)

laug kv; laug (8); lauga (1); laugar (17); laugarinnar (1); laugina (4); lauginni (3); laugu (2); laugunni (1)
Grett 50;1031

dag nokkuð eftir jóla a> Grettir fór í Svarf 28;1824	laug einn saman. fiorgeir vissi flá og
hafði borist. Karl lætur gera henni Kjaln 9;1447	laug og fá henni góð klæði. flá gekk Karl
heim til Kollafjarðar og var Ólöf við Svarf 28;1823	laug og heilsaði Búa. Hann tók
hana til skips og lét gera henni BandM 7;11	laug og klæddi hana góðum klæðum og gerði
Og um vorið hittast fleiri fægar við Kjaln 9;1447	laug og spyr Ófeigur tíunda. Oddur lést
hjá flér." Hún gerði honum flá Fljót 17;707	laug og strauk hvert bein á honum. Síðan
hádegisskeið. flá hafði Kórekur búið Grett 75;1068	laug . Tekur hann nú og fægir sár Gunnsteins
gekk til bæjar a> Reykjum og fór í BjHít 9;86	laug flvi honum var orðið nokkuð
ger kerlaug flvi a> eigi er annarra fiórð 5;2020	lauga kostur í Noregi. Konungur og hans menn
flá var einn dag a> Ormur gekk til fiórð 3;2015	laugar a> Sigríður frá Ósi var flar fyrir og
mikill. Hann gekk jafnan til fiórð 3;2015	laugar a> skemmta sér. Og einn dag gengu
Og einn dag gengu fleiri Skeggi til VígGl 11;1921	laugar eftir vanda og lágu við laugina og
En fiorgímur fiórisson fór flann dag til VígGl 19;1932	laugar er fleirra var utan von með máltil og
ríða með sér og kveðst mundu ríða til fiórð 3;2015	laugar . Hann stefnir suður úr garði um
flenna dag gekk Sigríður frá Ósi til Egla 88;516	laugar með léreft sín og bjóst í flann tíma
bað flá taka sér hest "vil eg fara til Laxd 33;1579	laugar ." Og er Egill var búinn gekk
og dvelst flar um hríð. Guðrún kom til BandK 7;35	laugar og fagnar vel Gesti frænda sínum.
misserin. Nú ríður Oddur oft til Njála 116;261	laugar og hittir föður sinn og batnar
Síðan voru borð tekin en Flosi tók fiórð 12;2039	laugar og lið hans. Flosi hugði a>
sama dag hafði Ásbjörn gengið til LjósA 2;1726	laugar og sex menn með honum. Nú er a>
En um morguninn ríðu fleiri til LjósA 2;1726	laugar og töluðu mart.
vildara a> ríða með mér um daga til Egla 88;517	laugar og vinna ekki?" Hann kvað sér
kastað í fé sínu. Fyrir sunnan ána eru LjósC 18;1685	laugar og flar skammt frá jarðholur stórar og
flykir flér nokkuð hægra a> ríða til Kjaln 18;1458	laugar um daga?" Hann kvað flá vísst enn
Búi gengu flá austur undir fjallið til Vatn 44;1898	laugar . fiar voru vellir fagrir. Síðan
í anddyri. fiá var Glævir flar og tók LjósA 2;1727	laugar . fiá gengu menn hjá honum með
ríður fiursteinn með Guðmundi oft til BjHít 9;86	laugarinnar . Og fannst mönnum mjög orð um flá og
sín á völinn en tjaldað var yfir Kjaln 16;1457	laugina . En flá var mönnum flá títt a> hafa
kveðst mundu eiga flar dvöl nokkura við laugina og binda sár sín. Var nú svo gert.	

BjHít 9;86	
Noregi. Konungur og hans menn fóru í fiór 3;2015	laugina og lög-u menn klæ-i sín á völlinn en
til laugar eftir vanda og lágu vi	laugina og tölu-ust me-. fienna dag gekk
Grett 50;1031	
ef eg ræ- á hann flá er hann fer frá	lauginni . "Ekki er mér um fla-," sag-i
Grett 50;1031	
öxina. Grettir gekk flá ne-an frá	lauginni og er fleir fundust mælti fiorgeir: "Er
Grett 75;1069	
nokku- kalt. Baka-ist hann lengi í	lauginni um nóttina og fór sí-an í stofu. fiar
fiór 12;2040	
Reykja. Í flann tíma gekk Ásbjörn frá	laugu og sá mannare-i-na. Ásbjörn mælti
Laxd 39;1592	
Jafnan bar svo til a- Gu-rún var a-	laugu . flótti Kjartani gott a- tala vi-
Reykð 16;1757	
og sjá fleir a- menn ganga brott úr	laugunni og kenna a- flar var Steingrímur og

Når man læser sagaerne, så ser man, at det ikke er kun for at bade når man *gengur til laugar*. I de fleste tilfælde er det to eller flere som går, og *laug* er ofte et vigtigt mødested, hvor vigtige beslutninger bliver truffet. At *ganga til laugar* har derfor en vigtig konnotation som burde nævnes i beskrivelsen.

5. Tolkning

En leksikograf, som har adgang til et tekstkorpus for det sprogstadium, som han eller hun vil beskrive, er i princippet i stand til at lave en meget nøjagtig syntaktisk og semantisk beskrivelse af enkelte ord. Men selv om en begrænsning af enkelte ords distribution eller betydning er mulig, hvis man har adgang til alle eksempler, så betyder det ikke, at den altid er naturlig eller rigtig. Den kan godt være en illusion, som forklares af de begrænsede teksttyper som vi har. Selvom det kan lyde paradoksalt så kan en beskrivelse, som kun baseres på nogle få eksempler, plus leksikografens intuition og viden, være rigtigere end en beskrivelse som baseres på alle eksemplerne i korpuset.

I sådanne tilfælde er det leksikografens arbejde at tolke eksemplerne. I denne tolkning må han både se på parallelle eksempler i gammel-islandsk og også se på nutidssproget. Hvis vi kommer igen til eksemplet med et adjektiv, som kun står med nogle få substantiver, så kan man spørge: Findes der i korpuset andre adjektiver, som ser ud til at have en distribution der er begrænset i samme grad? Hvis de gør, så er det naturligvis et godt argument for den tolkning at distributionen ikke er tilfældig, og skal derfor dokumenteres. Man kan også spørge: Er dette adjektivs distribution også begrænset på den samme måde i nutidssproget? Hvis den er, så er det naturligvis et argument for at den samme

begrænsing var til stede i gammelisländsk, selv om det naturligvis ikke er noget bevis. Men hvis nutidssproget ikke viser en sådan begrænset distribution, så kan det være svært at sige om begrænsingen i gammelisländsk er en illusion, eller om der er sket en semantisk ændring.

Det er ofte vældig svært at finde den rigtige udvej. I og for sig kan man naturligvis sige, at hvis et adjektiv kun forekommer med en bestemt type substantiver i gamle tekster, så har vi ikke lov til at sige noget andet om dette adjektivs distribution; d.v.s. vores beskrivelse må vise denne begrænsing. Det kan naturligvis gøres på mere end én måde. Én mulighed er at give en almen semantisk beskrivelse, men lade eksemplerne vise ordets brug. På den måde lader man brugeren om at drage sin egen konklusion. En anden mulighed er at beskrivelsen viser visse begrænsinger i ordets distribution, uden at sige noget om generelle konklusioner.

Man må også tænke på beskrivelsens formål. Der er naturligvis en fundamental forskel på en beskrivelse af gammelisländsk på den ene side og en beskrivelse af nutidssprog på den anden. Den førstnævnte beskriver visse tekster, og skal hjælpe brugerne til at forstå disse tekster bedre. Derfor kan man måske sige at det er naturligt at gå så langt som muligt når man beskriver gammelisländsk; man behøver ikke tage hensyn til andre tekster end de som man arbejder med. Hvis et bestemt ord fx kun forekommer som subjekt, så skal man nævne dette i beskrivelsen, selv om der ikke er nogen grund til at tro at ordet i virkeligheden har været begrænset på denne måde. Andre muligheder er simpelthen irrelevante; foremålet med beskrivelsen er kun at være en nøgle til teksterne, men ikke til selve sproget.

Bibliografi

- Eiríkur Rögnvaldsson 1990: Orðstöðulykill Íslendinga sagna. I: *Skáldskaparmál* 1, 54–61.
- Eiríkur Rögnvaldsson 1991: Quirky Subjects in Old Icelandic. I: Halldór Ármann Sigurðsson (ed.): *Papers from the Twelfth Scandinavian Conference of Linguistics*, 369–378. Institute of Linguistics, University of Iceland, Reykjavík.
- Eiríkur Rögnvaldsson 1993: Collocations and the Minimalist Framework. I: *Lambda*, 18. 107–118.

- Eiríkur Rögnvaldsson 1994–95: Breytileg orðaröð í sagnlið. I: *Íslenskt mál*, 16–17.
- Eiríkur Rögnvaldsson. 1995a Old Icelandic: A Non-Configurational Language? I: *NOWELE* 26, 3–29
- Eiríkur Rögnvaldsson 1995b: A Concordance to Old Icelandic Texts and its Lexicographic value. I: *Nordiske studier i leksikografi* 3, 123–135.
- Eiríkur Rögnvaldsson/Örnólfur Thorsson 1989: Fornir textar í tölvubanka. I: *Saga News* 4, 19–24.
- Fritzner, Johan 1954: *Ordbog over Det gamle norske Sprog*. Nytt uforandret opptrykk av. 2. utgave (1883 1896) med et bind tillegg og rettelser redigert av Didrik Arup Seip og Trygve Knudsen. Oslo: Tryggve Juul Møller forlag.
- Guðrún Ingólfssdóttir 1992: Að utan: Um búandi konur í Íslendinga sögum. I: *Skáldskaparmál* 2, 124–134.
- Guðrún Ingólfssdóttir 1995: A Dictionary of the Icelandic Family Sagas and Semantic Classification. I: *Nordiske studier i leksikografi* 3, 167–174.
- Örnólfur Thorsson 1993: *Orð af orði*. MA-opgave. Háskóla Íslands, Reykjavík.

Martin Gellerstam

Rätt och fel i korpusen

Today, language corpora are necessary requisites for the compiling of dictionaries. But to what extent should they have an impact on the lexical description? Opinions differ between lexicographers who use the corpus more or less as an inspiration for lexical description and those who rely heavily on corpus data (e.g. the wording of examples in Collins COBUILD). To what extent is a corpus reliable as a source for lexical data? This question is discussed against the background of a new Dutch spelling guide based on a 50 million word corpus.

1. Korpus och ordböcker

Att använda korpus när man gör ordböcker – över dagens eller gårdagens språk – är idag närmast en självklarhet. Historiska ordböcker har länge gjort det även om korpusarna kanske inte alltid varit så stora och de har härbärgerats på papperslappar snarare än i digital form. Men någon form av korpus har funnits vilket man i dagens korpusboom ibland glömmer bort. Enspråkiga ordböcker över det moderna språket har tidigare inte varit lika noga med att redovisa en korpus. Det viktiga har varit att man har fått med aktuella ord, ofta genom att konsultera andra ordböcker och nyordssamlingar. Textexemplen har företrädesvis konstruerats av redaktionen. Detsamma har tidigare gällt tvåspråkiga ordböcker.

Läget har på senare tid radikalt förändrats, bland annat genom korpuslingvistikens och datateknikens intåg i lexikografin. Men även om det idag är självklart att man skall ha en korpus så är det inte självklart hur långt man skall använda den. Till skillnad från de historiska ordböckerna så handskas nuspråkiga ordböcker lite mindre varsamt med exemplen: man talar vanligen inte om var de kommer ifrån (direkta beläggsuppgifter saknas) och man anpassar dem normalt för att de inte skall vara för långa eller komplicerade eller på annat sätt ägnade att förbrylla läsaren. I de få fall – jag tänker särskilt på Collins COBUILD – där man låter korpusen ha vitsord också ifråga om exempel, utsätter man sig för hård kritik från lexikografer som visserligen använder korpusar men inte går så långt i renlärighet. Man vill bygga på korpusen men – i likhet med Linje två i den svenska kärnkraftsomröstningen – göra det "med förnuft".

Men hur långt har egentligen korpusen vitsord? Man kan säga att det finns en skala från korpus skeptiker, som vet vad som är den goda normen och inte låter verkligheten störa, till korpusaktivister som är beredda att låta korpusen bestämma allt. Korpus skeptikerna finns framför allt i länder med starkt normerande tradition, möjligen kopplat till en rätt stor skillnad mellan tal och skrift. Särskilt den västeuropeiska traditionen har varit starkt pro-korpus, något som kanske inte bara har en ideologisk bakgrund utan också bottnar i bättre datautrustning än vad som finns i många andra länder. Vi som anser oss vara korpusanvändare med förnuft är kanske inte särskilt rädda för kritik från korpus skeptikernas sida. Snarare behöver vi värja oss mot föreställningen att korpusen skulle kunna användas mer än vad vi är beredda att göra. Låt oss materialisera en lite mer ortodox inställning ifråga om korpusanvändning och se hur en person som John Sinclair formulerar tankarna bakom COBUILD (Sinclair 1987). I förordet till ordboken tar Sinclair upp frågan om korpus ("usage") och auktoritet ("authority") och diskuterar förhållandet mellan dessa två begrepp:

These concepts must support each other or noone will respect either of them. If /.../ authority is not backed up by usage, then noone will respect it. It will be seen as unrealistic, arbitrary, old-fashioned and a barrier to free expression. Similarly, noone will respect usage if it is merely an unedited record of what people say and write. Unless they have the support of authority, people will be unable to distinguish between good and bad usage; it will not be possible to teach or use the language with any confidence. /.../ Any successful record of the language such as a dictionary is itself a contribution to authority. People tend to believe that dictionaries tell them what is or is not allowed. Actually, the rules of a language are very flexible, but it is difficult to show all this in a dictionary. However much the editors may say they are only recording and following usage, there is no doubt whatsoever that they take thousands and thousands of decisions which contain an element of subjective judgement.

Möjligen blir man lite förvånad över hur självklart Sinclair avvisar en "unedited record of what people say and write" och hänvisar till det som i andra sammanhang (Zgusta 1996) kallats lexikografens "kreativitet". Och vilka är då närmare bestämt dessa "thousands and thousands of decisions"? Här kommer vi in på normativa resonemang och vi ska därför först titta lite på något som man skulle kunna beteckna som normerarens dilemma.

2. Språkvårdens Moment 22

Varför är normeraren – här kallad språkvårdaren – så defensiv, varför bestämmer man inte helt enkelt vad som är det "rätta bruket" och sedan ser till att bestämmelserna följs? Det finns exempel på att språkvårdsorgan följt en sådan maxim och spåren förskräcker. Ett ofta citerat exempel (se t.ex. Bengt Sigurds bidrag i Allén & Loman & Sigurd 1986) är stavningsreformen 1906. Då avvisade Svenska Akademien förenklingen av stavningen *-dt* i ord som *rödt*, *kastadt* osv. (till *rött*, *kastat* osv.). Man framhärjade långt efter det att svenska folket – och framför allt skolan – anammat den nya stavningen och tvingades till slut att i en senare upplaga av ordlistan ange sin egen form som ett förslag i andra hand.

Bilden av äldre, kloka språkvårdare som "bestämmer" hur språket skall se ut är naturligtvis falsk: dessa personers klokhed går i bästa fall ut på att konsolidera det som de flesta är eniga om och vid alternativa möjligheter förelå den lösning som man anser mest konsekvent och ändamålsenlig utifrån systemet. Detta är mycket nog och alldeles tillräckligt för språkvårdens existensberättigande. Man behöver inte måla upp ett språkligt Moment 22 där det som är språkligt korrekt är det som man i efterhand kunde konstatera var språkligt korrekt. Och i den mån som denna ovisshet ändå finns så är det något som man får lära sig att leva med. Språket är ett levande och föränderligt system. En av de första medlemmarna i Svenska Akademien, Anders von Höpken, uttryckte saken så här: "Det finns inget annat språk än det som folk i allmänhet talar och skriver". En korpus är ett utsnitt ur språket eller en del av språket som visar hur "folk i allmänhet talar och skriver". Är det inte lexikografens uppgift att återge just detta? Vari ligger de "tusentals beslut" som gör att ordboken skiljer sig från det språk som talas och skrivs? Var klämmer egentligen skon? Är det alla de tillfälliga ordbildningarna som aldrig finner vägen in i ordböcker (*Göteborgsläkare*), är det icke accepterade uttryck (*jag slog han*), är det icke accepterade uttal (*elektri'ker*), är det missuppfattningar som håller på att bli standard (*situationstecken* istället för *citationstecken*)?

Vi har hittills talat allmänt om korpusens användning för lexikografiska ändamål. Innan vi går vidare behöver vi precisera dessa ändamål. En korpus är bevismaterialet vid konstruktionen av en ordbok men inte verktygen att bygga ordboken. En aldrig så bra korpus hjälper dig inte att formulera definitionerna i ordboken eller den grammatiska beskrivningen. Korpusen kan enbart hjälpa dig att välja mellan olika frekventa uppslagsord, leverera autentiska och frekventa exempel på meningar, syntaktiska typer, kollokationer och avledningar till uppslagsordet. I en korpus kan saknas vissa betydelser hos ett ord, du

kan inte räkna med att hitta alla böjningsformer till ett ord osv. I fortsättningen skall vi huvudsakligen diskutera användningen av en korpus som underlag för en rättskrivningsordlista, alltså en ordlista över språket som ger många exempel på ord och visar hur de stavas och böjs, kanske också i vilka stilistiska sammanhang de förekommer. En sådan ordlista finns redan och vi skall se närmare på hur den ser ut.

3. En korpusbaserad rättskrivningsordlista

I Holland utgavs i december 1995 en rättskrivningsordlista (*Woordenlijste Nederlandse taal*) baserad på en holländsk 50-miljonerordskorpus. Korpusen som ligger till grund för ordlistan heter Corpus of Present-Day Dutch och består av en enspråkig fulltextkorpus med 1600 texter, 50 milj. löpord och 700.000 olika ord från tiden 1979–1990.

Idén är alltså att göra en stavningsordlista med uppgifter om böjning utifrån en stor korpus. Korpusens storlek och fördelning gör det möjligt att ställa krav på frekvens och fördelning hos de ord som tas med i ordlistan. Ett tröskelvärde som använts i detta avseende är att ord(former) för att komma med i ordlistan skall ha minst två belägg i två deltexter. Korpusen används för tre saker: urval, stavning och böjning.

Skulle en sådan idé kunna anpassas till att gälla *Svenska Akademiens ordlista* (SAOL)? Frågan gäller hur långt man kan låta korpusen bestämma det lexikaliska innehållet och var ens egna normativa tankar kommer in och tar ifrån korpusen beslutanderätten. Vi skall nu se närmare på fördelarna och nackdelarna med att låta en korpus få en central roll vid produktionen av en ordbok.

4. Korpusens urval är ovedersägligt

Fördelen med att ha en korpus som grund för en ordbok är att urvalet i alla avseenden är ovedersägligt. Inte i den meningen att korpusen såsom stickprov ur en viss population skulle vara oomtvistad. Men om man väl accepterar det faktum att ordboken bygger på en viss korpus så är invändningar ifråga om urval, omfattning, språktyp etc. meningslösa. Om jag gör en ordbok över Strindbergs språk baserat på författarens samlade produktion så är det poänglöst att invända mot det ordförråd som blir resultatet. På samma sätt kan man stödja sig på en korpus över allmänspråket – även om ett sådant stickprov alltid kan utsättas för kritik. Invänder någon att ett vanligt ord inte finns med så kan man svar att det kanske finns i språket fast inte i just det här stickprovet. Gör man korpusen tillräckligt stor (så att den rentav omfattade allt som är skrivet

på ett språk under en viss tid, alltså hela populationen) så skulle man kunna säga att det som inte finns med knappast är värt att redogöra för i en ordbok. Och mellan ett litet stickprov (som t.ex. Brownkorpusen) och hela den skriftspråkliga populationen under en viss tid finns naturligtvis mellannivåer. En sådan är exempelvis den holländska 50-miljonerordskorpusen som vi just talade om.

Om korpusen inte alltid är tillräcklig så är den nödvändig för de flesta avgöranden om språkbruk. Ett bra exempel på korpusens oundgänglighet är förekomsten av variantformer och variantböjning. Hur fördelar sig former som *sjal* eller *schal*, *jogging* eller *joggnig*, *chef-läkare* eller *chefsläkare*? Det är lätt att exemplifiera vikten av korpusen som korrektiv genom att jämföra med existerande ordböcker. I Tabell 1 ges exempel på ett antal frekventa alternativformer som i olika lexikaliska källor (här SAOL11 och SOB) är mer eller mindre anpassade till fördelningen i korpusen (som dock representerar ett något senare språkbruk).

Uppgifterna i SAOL11 och SOB är någorlunda likartade men angivelserna om alternativens respektive vanlighet (där *eller* jämställer men *även* innebär ett andrahandsalternativ) är olika anpassade till korpusdata. Skillnaderna är mindre en fråga om SAOL:s normerande inslag kontra SOB:s rent beskrivande utan snarare en olika långt gående anpassning till bruket.

Tabell 1. Hopskrivning eller särskrivning. Korpusfrekvenser i Press 95 och rekommendationer i SAOL11 och Svensk ordbok.

Press 95		SAOL 11 (rek.)	SOB (rek)
efter hand 96	efterhand 159	efter hand äv. efterhand	efterhand äv. två ord
i dag >2000	idag 928	i dag el. idag	i dag äv. idag
i fred 68	ifred 22	i fred äv. ifred	i fred el. i fred
i gång 520	igång 631	i gång äv. igång	i gång el. igång
till buds 31	tillbuds 0	till buds äv. tillbuds	till buds äv. tillbuds
till synes 180	tillsynes 5	till synes äv. tillsynes	till synes äv. tillsynes
till godo 36	tillgodo 22	till godo äv. tillgodo	till godo äv. tillgodo
till rätta 106	tillrätta 57	till rätta el. tillrätta	till rätta äv. tillrätta

till känna 16	tillkänna 3	till känna el. tillkänna	till känna äv. tillkänna
---------------	-------------	-----------------------------	-----------------------------

Att bygga en ordbok på korpusgrund är för övrigt ingen nyhet för just den nämnda holländska ordlistan. Ett annat och mer känt exempel är *The American Heritage School Dictionary* (1972) som baseras på *The American Heritage Word Frequency Book* (1971), frukten av en textgenomgång av ett stort antal läroböcker i olika skolämnen i USA. I frekvensräkningen ges uppgift om frekvens, spridning över olika (skol)ämnen. Korpusen innehåller också ett stort urval exempel på ords användning.

Under senare år har som tidigare nämnts *Collins COBUILD* varit det mest kända exemplet på en ordbok som fast baseras på korpusmaterial. Eftersom korpusen här har använts betydligt mer konsekvent än i de flesta andra ordböcker i England och internationellt skall vi först undersöka de argument som talar mot en långtgående användning av en korpus för lexikografiska ändamål.

5. Argument mot en korpusbaserad ordbok

Den kritik som brukar riktas mot ett alltför troget utnyttjande av en korpus i ordbokssammanhang är följande.

För det första kan det invändas att korpusen är full av *tillfälliga bildningar*, framför allt en mängd sammansättningar som är "självklara". Skall en ordlista inkludera ord som *göteborgsläkare*, *huvudtalare*, *3000-årig* och *tysk-fransk*? Tillfälligheter kan också gälla på satsplanet: många konstruktörer av ordböcker anser sig inte kunna använda exempel som de går och står i texten utan anpassar kontextberoende meningar efter vad läsaren anses kunna ha nytta av. Denna senare aspekt på tillfälligheter i korpusen skall här inte vidare diskuteras. Även om det är en viktig del i frågan om hur långt en korpus kan användas så har denna fråga för det första redan flitigt diskuterats (se t.ex. Malmgren 1994), för det andra berörs här huvudsakligen lexikografiska källor av typ rättskrivningsordlistor och slutligen handlar användningen av meningsexempel mer om pedagogik än om rätt och fel.

För det andra kan korpusen vara ett *snedvidet* eller *alltför litet stickprov* ur den språkliga populationen. Detta får återverkningar på alltifrån ordurvalet till fördelningen av grammatiska och stilistiska kategorier: det blir stora luckor i ordförrådet, semantiska fält är inte utfyllda (t.ex. alla månaders namn är med utom *november*), det saknas

information om böjningsformer för att inte tala om den morfologiska variationen och vissa stilregister kan saknas eller vara dåligt företrädda.

För det tredje kan man invända att *korpusen innehåller många fel*. Det kan röra sig om enkla saker som korrekturfel men också om s.k. normluckefel enligt Ulf Telemans förslag till terminologi (Teleman 1979), dvs. allt som strider mot språknormen (i stavning, ordbildning, ordböjning, syntax och semantik), stilfel uppkomna genom sammanblandning av tal- och skriftnormen och fel som innebär att den språkliga kompetensen helt enkelt strejkar (anakoluter etc.).

Låt oss se lite närmare på dessa synpunkter.

(a) *Tillfälliga ordbildningar*

En korpus har i språk som svenskan och tyskan många sammansättningar och avledningar som är tillfällighetsbildningar på olika sätt. Det kan röra sig om sammansättningar med proprier, avledningar av siffror, sammansättningar gjorda för stundens bruk (*ölkostnader, bilköp* och *slumskildring*). Problemet med tillfälliga ordbildningar löser man inte enbart med att utöka korpusen och sätta en lägsta gräns för hur vanliga orden skall vara för att tas med i lexikonet. Också tillfälliga bildningar kan vara frekventa i vissa texter. Om en misstänkt brottsling av tidningspressen allmänt kallas *bombmannen* så är detta knappast något skäl för att införa ordet *bombman* i en ordbok. Enbart frekvens är uppenbarligen ett otillräckligt kriterium som behöver kompletteras med någon form av spridningsuppgift. Ett sådant kan se ut på olika sätt, alltifrån ett enkelt krav att ett ord skall förekomma i minst två textkategorier (hur nu dessa ser ut) till ett spridningsindex som ger en statistisk sammanfattning av hur väl spritt ordet är i olika textkategorier.

Man kan alltså på denna punkt behöva precisera vilken roll en korpus bör spela för urvalet av uppslagsord. Det handlar inte om att kritiskt redovisa allt som finns i korpusen (det kan man göra i en frekvensundersökning) utan om att hitta formella kriterier att välja ut de ord som kan ses som lexikaliserade enheter. För att återkoppla till en frekvensundersökning som ligger nära till hands (Nusvensk frekvensordbok) så är det den redovisning av korpusen som kallas basvokabulären – resultatet av ords frekvens och spridning – som skall gå vidare till ordboken.

Principen illustreras enkelt i följande översikt över förekomsten av ordet *rutin* i några olika källor (åtta olika korpusar i Språkbanken). Här rör det sig alltså inte om ordets fördelning på olika ämneskategorier utan helt enkelt om förekomsten i några olika korpusar. Exemplet

illustrerar enbart principen att ordet skall ha en viss spridning – frekvensen redovisas inte – för att få ingå i en viss ordbok. Av Språkbankens 130-tal belägg på ordet och dess böjningsformer visade sig följande ord (uppslagsord) förekomma i minst två källor:

Tabell 2. Förekomst av ordet *rutin* i mer än en källa i Språkbanken vid Göteborgs universitet

rutin	rutinarbete	rutinmässig
sjukhusrutin	rutinartad	rutinsak
vardagsrutin	rutinbesök	rutinundersökning
dagsrutin	rutinbetonad	rutinuppdrag
inköpsrutin	rutinerad	rutinåtgärd
orutin	rutinjobb	rutinärende
orutinerad	rutinkontroll	

Orden som uppfyller kraven på spridning är på det hela taget väl etablerade i svenskan och man slipper ifrån en rad tillfälliga bildningar i Språkbanksmaterialet som *brödjobbsrutin* och *daghemsrutin*. Men samtidigt kan man konstatera att flera av de ord som på grund av spridningen kvalificerat sig för ordboken inte kan anses mer lexikaliserade än de ord som avvisats av samma skäl. De kan kanske anses vara vanligare men är de mer "lexikaliserade"? Är *sjukhusrutin* mer lexikaliserat än *daghemsrutin*? Det finns inget enkelt sätt att ange graden av lexikalisering. Kanske är spridningen i kombination med frekvens det enda operationellt användbara sättet att ange om ett ord skall beskrivas i en ordbok.

Här finns också en avgörande skillnad mellan ordboken med sin semantiska beskrivning av ordförrådet och rättskrivnings-ordlistan med sin formella beskrivning av många ord. I det senare fallet behöver inte skiljelinjen mellan vad som är lexikaliserat eller inte (hur nu denna linje skall dras) bli avgörande för urvalet av ord.

Men om man använder frekvens och spridning i en korpus som kriterium på ordurval så hamnar man i ett annat dilemma: man riskerar att göra sig av med ord som har dålig spridning av det enkla skälet att de är ämnesbundna. Ett ord som *anställningsrutin* förekommer flera gånger men bara i en källa (Riksdagens snabbprotokoll 1978–79). Om korpusen är ämnesindlad bör denna typ av effekter bli ännu mer accentuerad: finns *avdragsgill* bara i texter om deklaration och skatt? Ändå skulle många anse ordet vara ett centralt ord i sitt ämne och en självklar kandidat till att ingå i en ordbok. Hur ser den procedur ut som automatiskt accepterar *avdragsgill* men avvisar *brödjobbsrutin*?

Det finns inget enkelt svar. Kanske kan man säga att alla ord som förekommer i fler än ett ämne eller en textkategori är starka kandidater

till urvalet i en ordbok. Men bland de ord som slås ut genom spridningskravet döljer sig väl lexikaliserade ord med ämnesspecifikt innehåll. Dessa ord kan inte lokaliseras på något enkelt automatiskt sätt. En väl planerad och stor korpus är därför i detta avseende ett nödvändigt men inte tillräckligt verktyg för urval till ordböcker.

(2) Korpusen är otillräcklig eller snedfördelad

Den andra invändningen handlar om att korpusen är otillräcklig: inte stor nog, inte balanserad nog. Den utgör därför ett otillräckligt underlag för ordurval men ger också otillräckliga besked på alla de punkter där en ordbok har att ge besked: stavning, ordböjning, ordbildning, syntaktiskt uppträdande, semantiska fält etc.

Kritik brukar riktas från lexikografiskt håll mot "små" standard-korpusar av typ Brown-korpusen och LOB-korpusen. De är visserligen användbara för en hel del syften och har betytt åtskilligt för korpusforskningen men de är helt otillräckliga när det gäller att få grepp om mindre vanliga ords språkliga uppträdande. Mot denna bakgrund skall man se uppbyggandet av dagens mycket stora textkorpusar på hundra miljoner löpande ord (*British National Corpus*) eller ännu mer (*Collins COBUILD*).

Det finns därför bara ett sätt att svara på den invändningen: man måste ha en stor och väl balanserad korpus för att överhuvudtaget med någon säkerhet uttala sig om ordförrådet och det språkliga bruket överhuvudtaget. Detta gäller oavsett om man är ute efter att beskriva bruket eller normera bruket: den språkliga verkligheten måste speglas i verkliga texter.

Här finns dessvärre en paradox: det man kanske mest vill ha svar på – t.ex. ovanliga ords böjning – ger korpusen ofta knapphändiga svar på. Så här kan det se ut om man letar i en korpus som Press 95 (som är under uppbyggnad men för närvarande består av ca 6 miljoner löpande ord). Målet är att skaffa fram autentiska exempel på böjningen av latinska ord på *-us*. I Tabell 3 (se nästa sida) ges också de mycket varierande böjningsklasser (här icke närmare redovisade) som knutits till varje ord i SAOL.

Lägg märke till att ord som för språkvetare är självklara (*genus, korpus, numerus, kasus*) inte kan beläggas böjningsmässigt. Detta är kanske inte så förvånande – för att få besked på dessa punkter krävs en korpus som också innehåller språkvetenskapliga texter.

Tabell 3. Böjningsuppgifter för vissa ord på -us i Press 95.

Ord	Data i Press 95
thymus	
dekanus	en dekanus
manus	-et -en
genus	
bonus	-en -ar
opus	ett opus
korpus	
numerus	
virus	-et -ar -en (pl.)
korus	i korus
tesaurus	
kasus	
lapsus	
passus	en passus, passusar

Men svaret är inte alltid fler och bättre fördelade texter. De latinska orden på *-us* är allmänt svårhanterliga och språkbrukare tenderar att undvika böjning. Detta är ett välkänt fenomen. Det kan gälla engelska ord av typ *scanner* där språkbrukaren ställs inför en försvenskad pluralis och en engelsk, det kan gälla pluralformer till vanliga svenska ord som *baby* och *taxi*. Men korpusen svarar inte bara på hur ord böjs (om den nu överhuvudtaget ger något svar), den säger också något om vilka former som faktiskt realiseras, något som konstruktören av en ordlista inte alltid är medveten om. Böjningen i korpusen kan slumpmässigt (?) saknas på någon punkt i böjningssystemet: *bok-boks-boken-bokens-böcker-böckers-böckerna-böckernas*. Skall man i så fall fylla ut systemet och säga att det är komplett? Hur gör man då om man hittar böjningen *kärlekar* (som i filmen *Tre kärlekar*)? Och vad skall man inte fylla ut utan lita på korpusen (t.ex. avsaknaden av **glädjor* av *glädje*)? Hur gör man med komparation hos adjektiven: skall man fylla ut där ett adjektiv kan ha komparation (även om inget exempel finns i korpusen) men inte fylla ut där man tycker att komparation är omöjlig (*död-dödare-dödast*)?

Bristen på sammanhang i semantiska fält är en vanligt förekommande kritik mot korpusbaserade ordböcker. Man kan lätt peka ut enskilda ord inom sådana fält som slumpmässigt kan saknas. Det kan röra sig om namn på månader, gradbeteckningar, verktyg eller överhuvudtaget sådant som kan utgöra en del av ett större begreppsält. Man kan med rätta fråga sig varför man skall ha med *april* men inte *maj*,

generalmajor men *integenerallöjtnant*, *lövsåg* men inte *fogsvans* etc. Samtidigt skall man komma ihåg att alla fält inte nödvändigtvis skall utfyllas: det heter *pojkklippt* men inte **flickklippt*, *poj kvalp* men inte **flickvalp*. osv.

(3) Korpusen innehåller fel

Den vanligaste invändningen mot att basera lexikaliska källor på korpusmaterial gäller dock inte vare sig tillfälliga bildningar eller korpusens otillräcklighet utan dess bristande tillförlitlighet. Den innehåller fel. Kritiken vänder sig framför allt mot två typer av fel. Den första typen är tämligen harmlös och består helt enkelt av sådana fel som alla är överens om är fel: korrekturfel, stavfel, omkastningar, konstiga avstavningar etc. och annat som visserligen kan vara irriterande för läsaren men som inte heller försvaras av någon utan på sin höjd urskuldats.

Hit hör kanske en typ av fel som är medvetna förvrängningar av författaren till en text. I Press 65 citeras ett programblad där publiken hälsas välkommen till teatern med enbart o-ljud (*Horrogod so rologt ott no kom*). Och i en annan text i samma korpus uppträder – i något semesterparadis – maträtten *kyskling med behäkning* (dvs. kyckling plus något som är helt obegripligt). Det behöver knappast sägas att dessa förvrängningar inte hör hemma i en ordbok.

Den andra typen av "fel" är viktigare: här gäller det tändningar i den språkliga normen som gärna betraktas som fel tills de slutligen accepteras. Typen finns på alla nivåer. På stavningens område kan vi se på sådan variation som *cigarrett:cigaret* där variantformen igår var fel men idag är accepterad, om än motvilligt. Det finns många liknande exempel som just gäller anpassningen av främmande ord. Ordet *juice* gavs tidigt en försvenskad variant *jös* som av korpusdata att döma inte har slagit igenom och på senare tid kommit att exemplifiera välmenande omsorg om språket som är dömd att misslyckas.

På uttalssidan finns många standarduttal som vän av ordning vill utesluta trots att en korpus visar att "felen" är rikligt företrädda. I fall som *acceptera* och *elektriker* (ofta uttalade /assepte'ra/ och /elektri'ker/) låtsar lexikaliska källor inte om uttalens existens. Samtidigt erbjuds språkbrukare alternativ som enligt korpusdata avvisas (*sambo* uttalas mot ordböckernas önskan med akut accent).

Fortsätter man till ordbildningen så accepterade länge ordböcker enbart bildningen *ortnamn* trots att korpusdata visade att *ortsnamn* var

det normala – men ordboksredaktörerna visste genom studier i nordiska språk att det skulle heta *ortnamn* och så blev det.

Ordböjningen uppvisar en liknande skillnad mellan korpusdata och ordboksrekommendationer. Standardexemplet är de relativt framgångsrika ansträngningarna från språkvårdshåll att undgå de negativa verkningarna i svensk ordböjning av engelskt plural-s (se Söderberg 1983). Resultatet blir att framför allt engelska lånord på *-er* ibland något hårdhänt tilldelas svensk plural på *-ar* (*containrar* etc). Korpusdata är i detta fallet mindre följsamma gentemot språkvårdarens rekommendationer.

Rena syntaktiska felaktigheter förekommer givetvis också i korpusen. En vanlig typ är brott mot valensregler (heter det *värna om x* eller bara *värna x* och finns det någon betydelskillnad?). Hit hör också prepositionsbruket där språkbrukare är osäkra och normen vacklar (heter det *intresserad av* eller *i* eller *för*?). Och hur är det med satsflätor (*ett brev som jag inte minns om jag har besvarat*) som tidigare ofta fungerade som språkvårdens måltavla? Och brott mot subjeksregeln (*efter att ha ätit en god lunch avgick tåget*)?

På den semantiska sidan är det inte heller svårt att finna korpusdata som bryter mot, eller åtminstone tänjer, normen. Ett av många exempel är den bildliga användningen av ordet *andas* (här exemplifierat med korpusen Riksdagsspråket 1978–79). Vad kan något (*ett svar, en motion, ett uttalande* etc) *andas*? Ett svar kan *andas optimism, misstro, missnöje, oro, tillförsikt* m.m. Men kan en proposition *andas* "ett mycket stort ansvar"? Eller kan ett svar *andas* "samma slags tongångar som när vi 1972 tog upp asbest"?

Slutligen kan nämnas en typ av korpusdata som visserligen är autentiska men behäftade med ett systematiskt fel: texterna består av översättningar som – det vet vi – alltid av påverkade av förlagan (Gellerstam 1996). Om det engelska *conversation* ("samtal") översätts till svenska med *konversation* ("artigt samtal (utan djupare innehåll)") så är översättningen inte lämplig att använda som korpusunderlag för en ordbok.

6. En korpusstyrd rättskrivningsordlista?

Det finns uppenbarligen en hel del invändningar mot att låta en korpus få direkt genomslag i en ordbok. Det finns fog för Sinclair's "thousands of decisions" innan verkligheten omsätts till en idealiserad lexikalisk beskrivning. Men kanske är svårigheterna större ju djupare beskrivningen är, syntaktiskt och semantiskt. Stannar man på den formella nivån där ord stavas, böjs och varieras – dvs. handlar det om en

rättskrivningordlista – så är goda korpusdata så nödvändiga att man undrar på vilka punkter man egentligen frångår dessa källor för att ta redaktionella beslut. Ett sådant avsteg är medvetna rekommendationer om "lämpligt" språkbruk i fall där korpusen kanske talar i annan riktning. Det rör sig främst om alternativformer där man från språksystemets synpunkt borde föredra en inhemsk form framför en mer frekvent men oönskad nyhet. Som exempel skulle kunna nämnas försök att introducera den svenska pluralformen i engelska lånord på *-er* (*scanner* etc.) trots att orden åtminstone i början är motspänstiga mot en sådan pluralform.

Om man bortser från sådana medvetna strävanden så finns ingen anledning att frånga de besked som ges i korpusen. Dock måste frågan om vad som är lexikaliserade ordbildningar och "tillfälliga" förmodligen hänskjutas till en lexikograf. I övrigt skall korpusen ha vitsord under förutsättning att följande krav uppfylls, liknande de som ställs för den holländska rättskrivningsordboken:

1. Korpusen skall vara stor, avspegla ett aktuellt språkbruk och vara väl fördelad över olika ämnen, genrer och stilarter.
2. Urvalet av ord skall baseras på korpusen – se dock inskränningen ovan – liksom uppgifter om vilka böjningsformer som aktualiseras i korpusen.
3. Informationen om ett ord (urval, stavning, böjning) skall stödjas av minst två källor.

Litteratur

- Allén, Sture 1971: *Nusvensk frekvensordbok*. 2. Stockholm: Almqvist & Wiksell.
- Allén, Sture & Loman, Bengt & Sigurd, Bengt 1986: *Svenska Akademiens och svenska språket*. Stockholm: Norstedts.
- Gellerstam, Martin 1978: Att välja sina ord. Om ordurval till enspråkiga ordböcker. *Rapporter från Språkdata*. 9. Göteborgs universitet, Institutionen för språklig databehandling.
- Gellerstam, Martin 1996: Translations as a source for cross-linguistic studies. I: Karin Aijmer & Bengt Altenberg & Mats Johansson (eds), *Languages in Contrast*. Lund: Lund University Press.
- Malmgren, Sven-Göran 1994: Språkprovets form och funktion i svenska betydelseordböcker från Östergrens Nusvensk ordbok till Svensk ordbok. I: *LexicoNordica* 1, 107–117.

- Collins COBUILD English Language Dictionary*. Editor in Chief: John Sinclair. London/Glasgow:Collins 1987.
- SAOL = *Svenska Akademiens ordlista över svenska språket*. 11 upplagan 1986 Stockholm: Norstedts.
- SOB = *Svensk ordbok*. Utarbetad vid Språkdata, Göteborgs universitet. 1986. Stockholm: Esselte Studium
- Söderberg, Barbro 1983: *Från Rytters och Cowboys tillTjuvstrykers. S-pluralen i svenskan*. Stockholm: Almqvist & Wiksell International.
- Teleman, Ulf 1979: *Språkrätt: om skolans språknormer och samhällets*. Lund: LiberLäromedel.
- The American Heritage School Dictionary*. 1972.
- The American Heritage Word Frequency Book*. 1971.
- Woordenlijst Nederlandse taal*. 1995. Haag: SDU.
- Zgusta, Ladislav 1996. The Lexicographer's Creativity. I. Martin Gellerstam et al (eds), *Euralex '96. Proceedings I–II*. Göteborg: Göteborg University, department of Swedish, 323–336.

Henrik Holmboe

Genbrug af korpora

The word *corpus* has at least two meanings of interest to us; an older meaning: "A body or complete collection of writings", and a more recent one: "A limited collection of written or spoken material upon which a linguistic analysis is based". In this paper we show that the older meaning is approximately 2000 years old, while the more recent meaning is hardly 50 years old. Corpus in the old sense and corpus in the recent sense of the word establish two quite different scientific scenarios as for use and reuse. We assume that the concept of a text bank as a repository of texts in machine readable form will gain importance; from here the user can select and compose corpora of his own for specific and focused linguistic purposes. Such texts being more widespread and more easily accessible will give rise to the demand for precise declarations of corpora and for standardisation of corpora and of the electronic software tools used for the linguistic analysis of text corpora.

Betydninger af ordet *korpus*

Ordet *korpus* havde som filologisk og sprogvidenskabeligt fagudtryk én betydning, førend først og fremmest den datamatiske lingvistik og forskellige andre datamatstøttede filologiske og sprogvidenskabelige studier medførte mindst én ny betydning. Det er i regelen klart ud fra tradition og situationel kontekst, hvilken af de mulige betydninger af korpus der er den på stedet aktuelle, men lad os ikke glemme, at der er flere betydninger. Den eller de nye betydninger af ordet har ikke fortrængt den ældre betydning, men eksisterer ved siden af den. Ligeledes er det sandt, at der er lingvister og filologer – og jeg regner mig selv heriblandt – der arbejder med forskellige opgaver, hvor snart den ældre betydning, snart en af de yngre betydninger er relevant.

Den ældre betydning

Hvis vi går til de store danske ordbøger, finder vi kun sparsom hjælp til en fastlæggelse af betydning og betydningsudvikling. Videnskabernes selskabs ordbog (VSO) har ikke ordet, hverken under **corpus** i bindet fra 1793 eller under **korpus** fra 1820. ODS s.v. corpus henviser til korpus, som vi finder i pågældende bind, bind XI, som er fra 1929: den fjerde af de anførte betydninger er med "kors", der står for "nu

ubrugeligt": † *samlingsværk, samlingsbind*, med følgende eksempel: "alle de Forordninger, som fra Tid til anden ere udkomne .. skal sankes tilsammen, og bringes i et Corpus." Eksemplet stammer fra Københavnske nye Tidender om lærde Sager fra 1742. ODS-supplementets seddelsamlinger rummer også et enkelt Holberg-citat, men ikke noget fra 19. årh. Ordet er naturligvis et lærd og filologisk ord, og når der ikke bringes flere belæg i ODS, og når danske ordbøger fra 19. århundrede ikke har denne brug med, kunne det hænge sammen med, at folk, der professionelt brugte ordet, udtrykte sig på latin eller i al fald ikke på dansk. Og dette fører mig til også at nævne den femte og sidste betydning i ODS, nemlig *korpus* som navn på en skriftstørrelse, således kaldet, fordi den først anvendtes ved trykningen af den romerske lovsamling *Corpus iuris civilis* fra 6. årh. i 1583.

På klassisk latin kender vi ordet *corpus* i bl. a. betydningen – og jeg henholder mig til Lewis & Short (1958) – "a whole composed of parts united", herunder om bøger; i sit skrift om grammatikere bruger Sueton (75–150 e.Kr.) *corpus* om en bestemt udgivelse af ni volumina og begrundet det udtrykkeligt med, at disse ni hørte sammen. Også Cicero (106–43 f.Kr.) og Livius (59 f.Kr.–17 e.Kr.) kan citeres for belæg, og her fra sammenhænge, der ikke er grammatiske traktater, man kunne næsten fristes til at sige sammenhænge fra det almindelige sprog, altså ikke-LSP. Da *corpus* på latin i den specielle betydning, vi efterlyser her, er et lærd ord, er det nærliggende at undersøge, om det skulle være et oversættelseslån fra græsk, kort sagt, kan det græske ord *svma* [så:ma] bruges i samme betydning? Liddell & Scott (1953) angiver "a body of writings" med kun ét belæg, nemlig et Cicerobrev (ad Att. II,1,3). Hvis dette virkelig er ældste belæg, må det betyde, at opfattelsen af en komplet samling af skrifter om et bestemt emne som et begreb er en romersk nydannelse; hvad har man kaldt det i Alexandria eller i Athen? måske ingenting; det ville være konsekvent, hvis begrebet først stammer fra romersk tid.

Den yngre betydning

Ser vi ordet **Korpus** efter i ordbøger for nogle moderne sprog, finder vi en ret god overensstemmelse fra sprog til sprog.

Begyndende med tysk konstaterer vi, at Trübner (1939–1957) slet ikke har ordet med; hos Klappenbach (1973) finder vi den latinske betydning: "Gesamtheit von Texten, Schriften: das K. der altdeutschen Urkunden." og tilsvarende i Dudens *Bedeutungswörterbuch* (Duden 1970), der i udgaven fra 1970 har: "vollständiges gedrucktes Werk, in

dem Urkunden, Gesetze o. ä. gesammelt sind (meist die Antike oder das Mittelalter betreffend)." Bemærkelsesværdigt er det, at Dudens Das große Wörterbuch der deutschen Sprache i 6 bind fra 1978 (Duden 1978) noterer: "Sammlung einer begrenzten Anzahl von Texten, Äußerungen o. ä. als Grundlage für sprachwissenschaftliche Untersuchungen." og anfører et citat fra 1974.

Opslag i Le Robert électronique, 1989 giver tilsvarende: "Recueil de pièces, de documents concernant une même discipline." og derefter som betydning to med et belæg fra 1961: "Ensemble limité des éléments (énoncés) sur lesquels se base l'étude d'un phénomène linguistique."

I den seneste udgave af The Oxford English Dictionary. Second Edition on Compact Disc 1992 finder vi også betydningen "A body or complete collection of writings or the like", og så tidligt som fra 1956 også betydningen: "The body of written or spoken material upon which a linguistic analysis is based."

	Ældre betydning	Yngre betydning	Første belæg
Klappenbach: Wörterbuch der deutschen Gegenwartssprache, 2. Aufl. 1973	Gesamtheit von Texten, Schriften		
Duden: Das große Wörterbuch der deutschen Sprache, 1978		Sammlung einer begrenzten Anzahl von Texten, Äußerungen o.ä. als Grundlage für sprachwissenschaftliche Untersuchungen	1974
Le Robert électronique, 1989	Recueil de pièces, de documents concernant une même discipline	Ensemble limité des éléments (énoncés), sur lesquels se base l'étude d'un phénomène linguistique	1961
The Oxford English Dictionary. Second Edition on Compact Disc 1992	A body or complete collection of writings or the like	The body of written or spoken material upon which a linguistic analysis is based	1956

Denne betydning af samling af et begrænset sprogligt materiale som grundlag for sprogvidenskabelige undersøgelser er den nye betydning af *korpus*, som i modsætning til den gamle ikke stiller forventning om totalitet; den gamle betydning ekspliciterer til gengæld intet om et fokuseret formål; formålet er blot at have alle dokumenter af en speciel kategori samlet på en minimalt systematisk måde. Belægget for den nye betydning kan som anført for tysks vedkommende dateres til begyndelsen af 70-erne, og noget tilsvarende – måske lidt senere – vil formentlig være muligt for dansks vedkommende. Riber Petersen

(1984) noterer dog ikke ordet. Dansk Sprognævns seddelsamlinger har først et belæg fra 1990, og det samme gælder Den danske Ordbogs materiale. Blinkenberg/Høybye (1975) har betydningen i 3. udg. af dansk-fransk ordbog fra 1975 både under **corpus** og **korpus** (Tekstudvalg for filologiske undersøgelser); men om dette beror på observationer af dansk eller på en tilbageoversættelse af det franske *corpus*, er nok usikkert; (jeg vil selv vove den påstand, at der er tale om en tilbageoversættelse.) Blinkenberg/Høybye (1984) har denne betydning med i fransk-danske ordbog fra 1984, 2. reviderede og forøgede udgave; men ikke i første udgave fra 1964. I norsk er betydningen registreret fra 1978, og i svensk fra 1986.

Brug og genbrug af korpora

Når vi taler om brug og genbrug af korpora, må vi gøre os klart, at der er tale om to forskellige ting, alt efter om vi mener korpora i den ældre eller i den nyere betydning af ordet.

Først om den ældre betydning: Her er korpus en totalitet af et materiale. Sammensætningen af et sådant korpus er dikteret af, at man er færdig med at sammensætte sit korpus, når der ikke er mere materiale. Eksempler herpå er velkendte. Det danske rune-korpus er alle danske runeindskrifter, Søren Kierkegaard-korpus eller H.C. Andersen-korpus er det samlede værk, og tilsvarende for udenlandske forhold: Shakespeare-korpus og Corpus inscriptionum latinarum er totaliteten. Dette giver en videnskabelig sikkerhed, idet indsamlingen af materiale og udnyttelsen af materialet i forskellige undersøgelser er principielt adskilte. Der er ikke tale om, at det pågældende korpus er sammensat med henblik på en eller flere på forhånd besluttede eller forudsete undersøgelser. Formålet med indsamlingen af et korpus kunne således være rent musealt. Dette betyder ikke, at korpusset ikke er etableret videnskabeligt eller uden axiomatik, men etableringsfasen er selvberørende og vil som sit fornemste formål have at leve op til forventningen om først totalitet og derefter autenticitet. Sekundære hensyn, bl.a. til anvendelighed og søgbarhed, kan tilsige, at man afviger fra autenticiteten. Man kan således vælge på denne måde at få et større antal tokens subsumeret under kun én type. Som konkret eksempel kan jeg anføre mit tidligere arbejde med de danske runeindskrifter. I en vis fase af korpusetableringsprocessen traf vi det valg at opretholde den autentiske forskel på 16-tegns og 24-tegns futharken, i en anden fase valgte vi at suspendere forskellen og anvende en form for standardrunenordisk, hvilket i denne fase betød, at tokens fra de to

skriftsystemer kunne sammenfattes under én og samme type. Et andet valg væk fra det autentiske var brugen af en standardlakune til gengivelse af en lakune af en hvilken som helst størrelse. Afgørende er det imidlertid, at alle senere brugere af sådanne korpora ikke er præjudicerede i deres undersøgelser og i deres undersøgelsesmetoders tilhørsforhold til den ene eller anden sprogvidenskabelige skole. Brugeren må selvfølgelig affinde sig med det autenticitetsniveau, der er lagt i korpus, men derefter står brugeren helt frit. Også til at foretage udvalg af tekster fra dette korpus for at etablere et udgangspunkt for fokuserede lingvistiske undersøgelser. Det er almindelig sprogbrug at tale om at bruge et sådant korpus, og at bruge det igen og igen; jeg er meget i tvivl om, hvorvidt man kan tale om at genbruge et sådant korpus.

Den nye betydning af ordet *korpus* etablerer en helt anden laboratoriesituation. Sigtet er her at foretage en lingvistisk undersøgelse for at nå et eller andet resultat. Udgangspunktet er ønsket om at opnå en eller anden indsigt, og med dette for øje tilrettelægger man etableringen af et korpus. Sigtet, hensigtsmæssigheden styrer, bestemmer næsten på forhånd, hvad der er støj, og hvad der er signal, og dette kan være såre godt og legitimt og videnskabeligt velbegrundet; man sørger for så at sige at rense præparatet, inden man undersøger det. Man kan sikre, at den samme overordnede hypotese er til stede både i etableringen af eksaminand og i eksaminationen. Men denne tæthed medfører også en vis lukkethed. Et sådant korpus kan naturligvis helt uproblematisk genbruges, hvis genbrugeren deler undersøgelsesformål og overordnet hypotese med det projekt, der oprindeligt tilvejebragte det pågældende korpus, men i det omfang, dette ikke er tilfældet, vil korpuset være mindre velegnet til det ændrede formål. Den nye bruger kan så vælge at revidere korpuset, tilføje nye tekster og udelade andre mindre relevante, og dermed i realiteten etablere et nyt korpus.

Korpora og tekstbanker

På denne måde kan allerede etablerede korpora, både i den ældre og den yngre betydning, komme til at fungere som tekstbanker, hvorfra brugere kan forsyne sig, når de sammensætter egne korpora til fokuserede formål. Jeg forestiller mig, at netop denne adfærd vil blive mere og mere udbredt, efterhånden som flere og flere enkelttekster bliver tilgængelige i digitaliseret form. Når vi i de forløbne 20–30 år har oplevet, at enkelte korpora i den nye betydning har fået en vis status, hænger det nok bl. a. sammen med, at elektroniske korpora var

nye, sjældne og dermed uprøvede; selve fremkomsten af dem trak disciple til; den eller de, der havde sammensat disse korpora, blev med det samme anset for også at være eksperter i at udnytte dem – og også i stand til at inspirere og lære andre kunsten; der groede videnskabelige miljøer ud af disse korpora under de rette omstændigheder. Naturligvis har jeg Sture Allén og hans miljø i Göteborg i taknemmelig erindring.

Men som før anført, adgangen til korpusmateriale er blevet meget lettere, adgang til forholdsvis simple, men effektive værktøjer til en vis analyse af store tekstmængder er tilsvarende blevet lettere, og mange flere benytter sig i dag af det i det daglige arbejde. I dag er efterslæbet ikke manglende adgang til maskinlæsbare tekster eller værktøjer, men derimod en manglende viden om teorier og metoder inden for kvantitativ lingvistik og sammenhængen mellem kvantitative og kvalitative problemstillinger.

Deklaration af korpustekster

Ikke mindst i dette perspektiv er det på sin plads at overveje, hvilke mindstekrav til varedeklaration af korpustekster man kunne ønske sig. Som regel rummer et korpus oplysning om, hvad det indeholder, og det er jo godt. Ofte savner man derimod oplysning om autenticitetsniveauet. Oplysninger om repræsentationsform, anvendte tegnsæt og koder kan man undertiden finde, og lang erfaring har belært mig om, at de sjældent står til troende. Og dette betyder ganske konkret, at mindre befarne kan komme til på grundlag af eksisterende elektroniske tekster fra forskellige kilder evt. suppleret med egne bidrag at etablere egne korpora, der ganske vist har en præcis indholdsoversigt, men som ikke har et ensartet autenticitetsniveau, og som rummer en blanding af forskellige repræsentationsformer. Det er let at påvise, at mange senere analyser vil være upræcise fra fødselen, hvis autenticitet og repræsentationsform er inegale. Jeg vil derfor gerne genopfriske nogle gamle filologiske dyder og formulere dem som tre gyldne regler for opstilling af korpora af enhver art, nemlig vedr.

- indhold
- autenticitet
- repræsentationsform

disse tre skal være veloplyste. Er de ikke det, vil det nedsætte værdien af derivede undersøgelsesresultater, også selv om disse ikke sættes ind i nogen sammenhæng uden for selve det aktuelle projekt.

Standarder for etablering af korpora

Et korpus rummer lingvistisk viden om sig selv, og denne viden kan vi analysere frem på forskellig måde. Denne korpusinterne viden skal kunne ekspliciteres på en reproducerbar måde, og alene derfor skal de gyldne regler gerne være overholdt. Imidlertid vil et korpus også kunne karakteriseres relativt i forhold til en række fikspunkter, der ligger uden for korpusset selv. Denne viden kommer kun frem, hvis vi kan referere til en veletableret viden uden for vort korpus, og en sådan reference vinder betragteligt i værdi, hvis mål og vægt i korpusset korresponderer med det, der er anvendt i de korpuseksterne fikspunkter. Som eksempel kan jeg nævne det velkendte, at en tekst består af en lille mængde meget højfrekvente ord og en stor mængde lavfrekvente ord. Detaljerne herom kan teksten selv meddele efter en hensigtsmæssig analyse. Det er også rigtigt, at de moderat frekvente og lavfrekvente ord kan anskues på en skala, der i midten har de stabile ord og på fløjene henholdsvis neologismer og obsolete ord, ord i knop og ord, der visner. Men ingen tekst kan i isolation meddele denne information om sig selv, ingen analysemetode kan levere dette.

Hensynet til ønsket om at kunne sammenligne analyseresultater fra forskellige korpora har formuleret et ønske om standarder, og helst internationale, for etablering af korpora. At argumentere imod standarder som sådanne lader sig næppe gøre og skal ejheller forsøges her. Men det må være på sin plads at kræve, at områdets standarder bliver detaljerede og veldokumenterede for at kunne tage hensyn til den mangfoldighed af fænomener, som vil vise sig i forskellige korpora. Og erfaring viser, at detaljerede standarder er svære at opfylde og overholde. Er standarderne ikke detaljerede, risikerer man, at en overholdelse af standarder medfører en lavere grad af autenticitet; og i så fald bliver standarder et tvivlsomt gode.

Standardisering af softwareværktøjer til korpusanalyse

En anden bestræbelse med samme overordnede sigte vil være lettere at gennemføre, nemlig ønsket om, at korpora skal analyseres med værktøjer, der opfylder visse standarder. På dette felt er man allerede godt igang. Ønsket om standardiserede informationsteknologiske værktøjer er nævnt i den danske regerings Handlingsplan som udtrykt i forskningsministerens Redegørelse til Folketinget om "Info-samfundet år 2000" og IT-politisk handlingsplan 1995. (Fra vision til handling). Tilsvarende er der i EU's telematikprogram under Language Engi-

neering reserveret ressourcer til udvikling af og udbredelse af softwareværktøjer til lingvistiske formål, herunder også behandling af korpora. Parallelt hermed igangsættes et opklaringsarbejde og en evalueringssproces netop med det formål at identificere eventuelle de facto standarder og om muligt få sammensat et repertoire af værktøjer, som kan anbefales som standardværktøjer og måske få endelig status af at være officielle standarder på området. Standardværktøjerne skal opfylde en række specifikationer, og der er hermed mulighed for, at den enkelte bruger til specielle formål kan udvikle egne værktøjer, der som udgangspunkt opfylder disse specifikationer og dermed kan være certificerede inden for et vist område samtidig med, at de opfylder en række specielle formål, som måske endnu ikke er omfattet af nogen standard.

I forbindelse med standarder for og standardisering af korpora har jeg udtrykt et vist forbehold med hensyn til store forventninger på kort sigt. Med andre ord: På dette punkt er der meget lang vej igen. Jeg har slet ikke et sådant forbehold med hensyn til standardisering af elektroniske analyseværktøjer. Årsagen er ikke bare, at vi her kan imødesee en succes på kort sigt, men en grundlæggende opfattelse af, at det er videnskabeligt korrekt at standardisere måleredskaberne snarere end patienten. En del arbejde er allerede i gang, alle er tilsyneladende vel tilfredse med konceptet om at tilvejebringe standardiserede elektroniske analyseværktøjer, og argumenterne imod er næppe til at få øje på. Arbejdet er ikke let eller trivielt, men jeg må anse det for at være et overordentligt vigtigt sprogvidenskabeligt anliggende, at analyserne af den stadigt stigende mængde af elektronisk tilgængelige tekster resulterer ikke bare i output, men i viden, ny viden om sprogene og om sproget.

Litteraturliste

- Andersen, Ingeborg/Henrik Holmboe 1983: *Konkordans over de danske runeindskrifter – Transskription*. Sprog Og Mennesker 5. Århus.
- Andersen, Ingeborg/Henrik Holmboe 1983: *Konkordans over de danske runeindskrifter – Translitteration*. Sprog Og Mennesker 6. Århus.
- Blinkenberg, A/Høybye, P. 1975: *Dansk-fransk ordbog*. 3. rev. og forøg. udg. København.
- Blinkenberg, A/Høybye, P. 1984: *Fransk -dansk ordbog*. 2. rev. og forøg. udg. København.

- Duden 1970: *Bedeutungswörterbuch*. Mannheim.
- Duden 1978: *Das große Wörterbuch der deutschen Sprache*. Mannheim.
- Fra vision til handling. Info-samfundet år 2000*. Forskningsministerens Redegørelse til Folketinget om "Info-samfundet år 2000" og IT-politisk handlingsplan 1995. København.
- Klappenbach, Ruth/Steinitz, W. 1973: *Wörterbuch der deutschen Gegenwartssprache*. 2. durchges. Aufl. Berlin.
- Lewis, Ch. T. & Short, Ch. 1958: *A Latin Dictionary*. Oxford.
- Le Robert électronique 1989= *Le Robert électronique*. Paris.
- Liddell, H. G./Scott, R. 1953: *A Greek-English Lexicon*. 9. Ed. Oxford.
- ODS = Dahlerup, V.: *Ordbog over det danske Sprog*. København 1919–1954.
- The Oxford English Dictionary*. Second Edition on Compact Disc. Oxford 1992.
- Riber Petersen, Pia 1984: *Nye ord i dansk 1955–75*. København.
- Trübner 1939–1957: *Deutsches Wörterbuch*. Berlin.
- VSO = *Dansk Ordbog udgivet under Videnskabernes Selskabs Bestyrelse*. København 1793–1905.

Tove Jacobsen

Øversettelseskorpora og tospråklige ordbøker

Translation corpora are fairly new to bilingual lexicography. After a brief presentation of some of the characteristics of translation corpora, the author shows how even a small translation corpus can improve the entries of a recent Norwegian-French dictionary.

Behov for forbedring av tospråklige ordbøker

Tospråklige leksikografer ble i 80-årene sterkt kritisert for å lage ordbøker som ikke gjenspeilte språket slik det virkelig ble brukt. Denne holdningen finner man også igjen i f.eks. i tittelen på et innlegg av Mary Snell Hornby på Euralex-konferansen i Leeds 1983: "The Bilingual Dictionary – help or hindrance?"

Leksikografene kan vel være enige i at denne kritikken har vært berettiget. En av grunnene til denne svakheten er at materialet til tospråklige ordbøker vesentlig er blitt hentet fra andre ordbøker (spesielt enspråklige) og supplert med egen intuisjon, informanter og spredte belegg fra aviser, romaner, radio, TV o.l. Et middel til å lage bedre ordbøker er utvilsomt bruken av korpora.

Enspråklige korpora og tospråklige ordbøker

Enspråklige korpora ble naturlig nok først brukt til redigering av enspråklige ordbøker, men er også blitt et viktig hjelpemiddel ved redigering av tospråklige ordbøker. Til utarbeidelsen av Oxford-Hachettes og Collins-Roberts nye engelsk-franske/fransk-engelske ordbøker har redaktørene hatt til rådighet store enspråklige korpora både for engelsk og for fransk. Selve øversettelsesarbeidet har blitt gjort på tradisjonell måte, men øversettelsene ble også sjekket mot korpus. På den måten kunne man se om den foreslåtte øversettelsen ble brukt i samme type kontekst som oppslagsordet i utgangsspråket, og om det hørte til samme språknivå. Likeledes fikk man en omfattende liste over kollokasjoner på begge språk (Atkins 1994). Fremdeles er det lite erfaring med bruken av tospråklige (øversettelses)korpora ved redigering av ordbøker.

Elektronisk lagrede oversettelseskorpora i Norge

På det nåværende tidspunkt eksisterer det ett prosjekt som omfatter både et engelsk-norsk (ferdig) og et fransk-norsk oversettelseskorpus (i startfasen) i Oslo-Bergen.

Disse korpora blir ikke laget spesielt for leksikografi, men for oversettelsesteori og kontrastive studier i sin helhet. Materialet er hentet fra skriftspråket. Hvert korpus vil bestå av ca. 2 millioner fortløpende ord. (Johansson/Ebeling under trykking, Johansson/Hofland/Ebeling 1996)

Kriterier og problemer for oversettelseskorpora

I tillegg til de kriterier som gjelder for enspråklige korpora, vil det være spesielle kriterier for oversettelseskorpora. Disse gjelder spesielt oversetterne. Mange forskjellige oversettere bør være representert i korpus. (Baker 1995 setter opp en liste over kriterier for oversettere hvor det i tillegg kommer inn momenter som går på om oversetterne er amatører eller profesjonelle, innfødte eller ikke, free-lance eller ikke.) Kriteriet med flest mulig forfattere og flest mulig oversettere er vanskelig å oppfylle for små språk. For oversettelser fra norsk til fransk f.eks. er det bare et lite antall bøker som blir oversatt, og antallet oversettere er også begrenset. Oversetterne har i tillegg tendens til å spesialisere seg på få forfattere, slik at heller flere bøker av én forfatter blir oversatt enn én bok av flere forfattere. For å få representert flest mulig forfattere og flest mulig oversettere blir det i det engelsk-norske og fransk-norske korpuset bare tatt med tekstutdrag og ikke hele bøker. Hvert tekstutdrag starter fra begynnelsen av boka, og man kutter etter ca. 10.000–15.000 ord, helst ved slutten av et kapittel.

Imidlertid må korpora hvor man skal lete etter leksikalske ord være mye større enn disse. Man må derfor ikke ha for store forventninger til å finne eksempler på forekomster og oversettelser av alle oppslagsord i en ordbok. Det vil si at man kanskje ikke finner så mange oversettelser av kulturbundne ord og såkalte "faux amis" som man hadde håpet på. Det betyr imidlertid ikke at et slikt korpus er ubrukelig i ordboksøyemed. For grammatiske ord og andre frekvente ord som ofte kan være vanskelige å oversette, er det til stor hjelp å se hvordan oversettere har løst oversettelsesproblemer i kontekst.

For at data skal være lett tilgjengelige, bør man ha et effektivt automatisk lenkingsprogram som gjør at man kan søke på ett eller flere ord og få frem forekomstene både i setningen på originalspråket og den tilsvarende oversatte setningen samtidig. Med det automatiske lenkingsprogrammet som brukes i det norske oversettelseskorpuset, er feilmarginen på 2%. (Hofland 1995, Johansson/Hofland/Ebeling 1996.)

Forventninger til bruken av oversettelseskorpora i tospråklig leksikografi

I litteraturen rundt leksikografi er bruken av oversettelseskorpora nevnt som et viktig redskap for at ordbøkene i mye høyere grad enn i dag kan vise hvordan oppslagsord skal oversettes i kontekst. Salkie peker på hvor få oversettelsesmuligheter ordbøkene tar med sammenlignet med de mangfoldige oversettelsene han har funnet i et engelsk-fransk oversettelseskorpus (Salkie 1995). Salkie mener at ordbøker bør gjenspeile de strategier oversettere bruker i kontekst. Hartmann sier det samme når han snakker om at ordboksredaktørene må ha en "awareness of the techniques that can be used to bring about such translation equivalence." (Hartmann 1995:292f.).

Man bør imidlertid være litt skeptisk til det å forsøke å gjengi strategier og teknikker i ordbøker. En strategi som oversettere bruker – spesielt under tidspress – er utelatelsen. En tekst kan godt fungere utmerket i den situasjonen den er beregnet på, uten at alle ordene er oversatt. Denne strategien fungerer derimot ikke for ordboksredaktøren.

Innvendinger mot bruken av oversettelseskorpora i tospråklig leksikografi

Oversatte tekster har visse trekk til felles som ikke finnes i originaltekster, og disse skal være universelle. I tillegg kommer interferenser fra utgangsspråket. (Baker 1995, Gellerstam 1989.) Et slikt trekk er at den oversatte teksten gjøres mer forklarende enn originalteksten. Et eksempel på det finnes i følgende oversettelse fra engelsk til fransk:

"And the car's in the westbound lane?" I enquired, for Spindrift, North Carolina, is three and a half hours east of Richmond. (Cornwell, *All that remains*, 3)

Et la voiture est sur l'autoroute direction *ouest*? m'étonnai-je.

Spindrift, petite ville de Caroline du Nord, est en effet à trois heures et demie de route à l'est de Richmond. (6)

Denne tendensen til å forklare er nok universell, men det som er individuelt er at oversetterne ikke føler dette behovet på samme tid. I den norske oversettelsen av den engelske originalteksten er *Spindrift* ikke forklart:

"Mens bilen er i *vestgående* kjørebane?" spurte jeg, for *Spindrift* i North Carolina er tre og en halv times biltur østenfor Richmond. (6)

Et annet typisk trekk for oversettelser er at visse fenomen kan opptre mye hyppigere der (eller mindre hyppig) enn de ville gjøre i originaltekster.

Disse særtrekkene er ikke noe hinder for at et oversettelseskorpus kan brukes i leksikografien. Men det er viktig å følge med i oversettelsesteoretikernes arbeid med å kartlegge typer av slike universelle trekk for oversatt språk. Like viktig er det å være godt skolert i kontrastiv lingvistikk, så man lett ser når oversetteren ikke har klart å riste av seg utgangsspråket. (En oversikt over forskjeller mellom engelsk og fransk finnes bl.a. i Vinay/Darbelnet 1958.)

Det har vært vanlig å hevde at såkalte parallelle tekster, dvs. tekster som omhandler samme emne på begge originalspråkene, er mer verdifulle for tospråklig leksikografi enn oversettelseskorpora. Dette kan være riktig hvis man kan gå ut fra at slike tekster er fullstendig upåvirket av et fremmedspråk. Men hvis man ser på avistekster f.eks., så baserer journalister mye av sitt stoff på nyheter fra utenlandske kilder, og de kan lett bli påvirket av språket der. Dette gjelder i enda høyere grad utenlandskorrespondenter som etter en tid i et fremmed land er blitt påvirket av landets språkbruk. Skal man være sikker på at språket er ekte, burde man bare ta hensyn til artikler som behandler hendinger fra ens eget land hvor man ikke kan influeres av fremmedspråk. Ett eksempel på det kan være artiklene om OL på Lillehammer i 1994.

Et konkret forsøk på å utnytte et mindre oversettelseskorpus til å forbedre en ordboksartikkel

Som tidligere nevnt er det nesten bare frekvente ord man kan regne med å finne i et norsk-fransk oversettelseskorpus. (Man kan også finne ufrekvente ord, men i et nokså tilfeldig utvalg.) Jeg har her valgt

adverbet *akkurat* for å se om bruken av et oversettelseskorpus vil kunne forandre ordboksartikkelen med dette ordet som oppslag vesentlig i Stor norsk-fransk ordbok. Den opprinnelige artikkelen står nedenfor. Man vil se at oversettelsekvivalentene ikke har noen forklaring med hensyn til hvilken kontekst de kan brukes i. Derimot er det et forholdsvis stort antall eksempler som skal vise fordelingen av de tre oversettelsekvivalentene

Eksemplene fra Stor norsk-fransk ordbok er valgt ut etter konsultasjon av norske enspråklige ordbøker, og oversettelsene er laget ved hjelp av informanter eller eksempler man har funnet i franske tekster og ordbøker:

akkurat adv - (1) exactement, précisément, juste: **a-da** à ce moment précis; **han er akkurat kommet hjem** il vient juste de rentrer; **det koster a- 20 kroner** ça coûte exactement 20 couronnes; **hodet mitt rakk ham a- til skulderen** ma tête lui venait juste à l'épaule; **det varte a- en uke** ça a duré juste une semaine; **det er a- de samme folkene** ce sont exactement les mêmes personnes; **det er a- det samme for meg** ça m'est complètement égal; **han er ikke a- noe geni** il n'est vraiment pas génial; **det er a- like så godt** c'est exactement aussi bon (2) interj: **var det slik det var? - akkurat!** cela s'est vraiment passé comme ça? - exactement! (*el* parfaitement!, absolument!)

En gjennomgåelse av bare et lite oversettelseskorpus gir et mye rikere bilde av *akkurat* og de uttrykk det går inn i. Man kan dele dem inn i grupper:

1. Uttrykk for tid

Jeg gikk tilbake til, nåja, stuen, skrudde på radioen akkurat tidsnok til å unngå den første morgenandakten og prøvde å få et aldri så lite overblikk over situasjonen.

Je retournai dans, disons, le salon, ouïr la radio, juste à temps pour éviter la première méditation religieuse du matin, et essayai de faire le point de la situation.

Hun hadde svingt bilen helt over tvert og var akkurat i ferd med å kjøre inn gjennom et portrom.

Elle avait traversé la rue et s'engageait sous une porte cochère.

Jeg skulle akkurat til å gå inn, da jeg hørte, på et nokså gebrokkent spansk:

J'allais juste y entrer quand j'ai entendu, dans un espagnol quelque peu hasardeux:

"Jeg skal akkurat til å begynne.

J'allais justement m'y mettre.

Riktignok har jeg fått et slik forhold til posten at jeg kan la det gå dager da jeg ikke åpner et brev, men akkurat nå skulle jeg ha mitt noenlunde på det tørre: Lønnsoverføringen gikk inn på konto den 27., vanlige forfall kommer ikke før i mai.

Certes, j'ai maintenant avec le courrier des rapports tels que je puis fort bien attendre plusieurs jours avant d'ouvrir une lettre, mais, en ce moment précis, je n'attends rien: mon salaire a été viré le 27, et les échéances habituelles ne tombent pas avant le mois de mai.

Tora prøvde å forestille seg hvordan det var å være hoppe akkurat nå.

Tora essaya de se mettre à la place de la jument en ce moment précis.

Det skulle nok la seg ordne. Men akkurat nå var det utrygt.

On arrangerait cela, mais, pour l'instant, ce n'était guère prudent.

Jeg antar han ikke oppfatter seg selv som særlig attraktiv eller vellykket akkurat nå.

J'imagine qu'en ce moment précis il ne se sent ni très séduisant, ni particulièrement réussi.

Det fortelles blant annet at han målte høyden på en pyramide i Egypt ved å måle pyramidens skygge akkurat når hans egen skygge var like lang som ham selv.

On raconte entre autres qu'il calcula la hauteur d'une pyramide en Égypte en mesurant l'ombre de celle-ci au moment précis où l'ombre de son propre corps coïncida avec sa taille réelle.

Segnet om akkurat idet han skiftet gir for å rygge inn i garasjen.

Il s'était effondré au moment précis où il avait change de vitesse pour reculer dans le garage.

2. *Uttrykk for sted*

Hver gang noen kom eller gikk i døra, kjente hun trekken komme sigende og finne frem akkurat der buksekanten glapp og huden ble bar fordi hun hadde vokst så i sommer at strømpene var blitt for korte.

Chaque fois que la porte s'ouvrait, elle sentait le courant d'air, juste à l'endroit où la peau était nu, entre la culotte et des bas qui ne montaient plus jusqu'en haut parce qu'elle avait beaucoup grandi au cours de l'été.

3. *Sammenligning*

Til vanlig var mammas øyne store og grønne med et tynt, matt gardin for, akkurat som sommergardinene hos tante Rakel.

D'habitude, maman avait de grands yeux vert clair, qui se voilaient d'un mince rideau terne, tout à fait semblable aux rideaux d'été de tante Rakel.

Tora visste plutselig i et blink at alle sto rundt gjerdet og stirret på at hingsten pumpet ut og inn av hoppa, og at de kjente det rare, hemmelige suget i underlivet akkurat som hun.

Tora se rendit subitement compte qu'ils étaient tous là autour de l'enclos à observer les mouvements du cheval sur la jument et que, tout comme elle, ils ressentaient dans le bas du ventre une secrète et étrange aspiration.

Hun ble en del av hverdagen deres, akkurat som dyrene som trengte et minimum av stell, men ellers greide seg selv.

Elle fit partie du quotidien, exactement comme le bétail, qui demandait un minimum de soins, mais pour le reste était livré à lui même.

Den gule konvolutten var lagt direkte i postkassen akkurat som de to hvite konvoluttene.

L'enveloppe jaune avait été déposée directement dans la boîte aux lettres tout comme les deux enveloppes blanches.

"Det er ikke akkurat slik som det var lenger, Mikael."

"Les choses ne sont plus tout à fait comme avant, Mikael."

Han er på tokt hos fienden og gjør akkurat som han vil.

Il est en campagne chez l'ennemi et n' en fait qu' à sa tête.

Uansett hva Sofie gjorde, gjorde hun akkurat det samme.

Sophie avait beau faire, l'autre faisait exactement pareil.

Denne boken er da skrevet akkurat slik.

C'est exactement la façon dont ce livre est écrit.

Tora nølte akkurat så mye at mora snudde seg og viste det ansiktet som Tora minst av alt ville se i kveld.

Tora hésita tout juste assez pour que sa mère se retourne et exhibe le visage qu'elle avait le moins envie de lui voir ce soir-là.

I dette korpuset er det mange regelmessigheter som kan overføres til vår ordboksartikkel **akkurat**. Artikkelen vil både bli fyldigere og mer konsekvent. Det er først ved hjelp av et korpus at man ser hvor lite man har klart å tenke i stereo. Selv når man senere ikke har noe korpus til rådighet, kan man kanskje lære seg å tenke som om man hadde det. Spørsmålet man kan stille seg er om resultatet hadde blitt det samme hvis man hadde hatt et enspråklig korpus fra starten av. Det tror jeg imidlertid ikke. Den norske delen ville også da kunnet ha blitt mer konsekvent, men man ville ikke øyeblikkelig se hva man burde ta med pga. oversettelsesproblemer.

Korpuset kan også fremvise eksempler hvor oversetteren rett og slett har hoppet over den delen av setningen hvor oppslagsordet befinner seg:

En mann med et kontor, en leilighet, to telefoner og en bunke regninger som iallfall akkurat nå ble litt mindre for hver dag som gikk.
Un homme avec un bureau, un appartement, deux téléphones et un tas de factures qui s'amoncelaient au fil des jours...

Slike eksempler må man bare se bort fra.

Ved andre anledninger er det omskrevet for mye til at resultatet er klart nok til å brukes i en ordboksartikkel. Hvis brukeren skulle få noe utbytte av disse oversettelsesforslagene måtte det være i en form for konkordans hvor man kunne finne de oversettelsene som ligger utenfor det regelmessige.

"Det er jo ikke akkurat så mye å høre fra dere bortsett fra når dere liksom skal forberede meg på et eller annet nytt sprell."
"On ne peut pas dire que vous vous manifestiez tellement souvent. Sauf lorsqu'il s'agit de me préparer a quelque nouvelle calamité."

Vi er mer opptatt av hvordan de tenkte enn av akkurat hva de tenkte.
Il s'agit donc de nous concentrer davantage sur leur manière de penser que sur le contenu exact de leur pensée.

Nei takk, alt er utmerket, og hun går, mens jeg langsomt blir klar over at det jeg egentlig ønsker akkurat da, er at hun skal bli stående der og se på meg spise.
Non merci, c'est très bien comme ça. Sur quoi, elle s'en va, cependant que je me mets à regretter qu'elle ne reste pas là pour me regarder manger.

De vanskeligste tilfellene er når oversettelsen er veldig lik den som vi regner som regelmessig, men for en som har utgangsspråket for øye,

mangler det noe. Da må man undersøke om dette er typisk for visse oversettere. Har de kanskje ikke riktig forstått betydningen av det originale uttrykket? Eller er det slik at for en franskmann er det ikke noen forskjell om man sier *en ce moment* istedenfor *en ce moment précis*? Er det siste en overkonkretisering som ikke er fransk? (Jacobsen 1992:536). For å undersøke dette nærmere kan man konsultere et enspråklig fransk korpus for å se hva som egentlig blir brukt: *en ce moment précis* eller *en ce moment* osv.

"Du, hvor er du akkurat nå?"

"Dis donc, ou est-ce que tu es en ce moment?"

Nei, jeg kan ikke fjerne meg, gi opp alt dette akkurat nå.

Non, je ne peux pas m'en aller, abandonner tout ça maintenant.

Musikk har sjel, akkurat som folk.

La musique a une âme, comme les gens.

Ingenting kan bli til noe annet enn akkurat det det er.

Rien ne pouvait devenir autre chose que ce qu'il est.

Sofie visste ikke om hun hadde noen særlig tro på at det var akkurat jord, luft, ild og vann som alt var laget av.

Sophie ne savait pas au juste si elle devait vraiment croire que la terre, l'air, le feu et l'eau étaient à l'origine de tout ce qui est créé dans la nature.

Etter å ha gått igjennom korpus og plukket ut de tingene man vil putte inn i ordboksartikkelen, ser det ved første øyekast ut til at artikkelen må bli lengre. Heldigvis vil korpus også ofte gjøre det mulig å fjerne opplysninger som viser seg å være irrelevante. I ingen av eksemplene som er funnet i korpus har adverbet *précisément* blitt brukt, altså kan det fjernes fra artikkelen. (Derimot er adjektivet *précis* brukt, men det er her knyttet til *moment* i oversettelsene for *akkurat da/nå/når*, altså kan det ikke stå som selvstendig oversettelsesekvivalent.) Sannsynligvis bør et større korpus brukes for at man kan være sikker på hva som kan fjernes.

akkurat adv - (1) exactement, (særlig om tid) juste:

a- da/nå à/en ce moment précis; **a-når/idet** au moment précis où; **a- tidsnok til å** juste à temps pour; **han er akkurat kommet hjem** il vient juste de rentrer; **jeg skulle a- til å gå inn** j'allais juste entrer; **det varte a- en uke** ça a duré juste une semaine; **det koster a- 20 kroner** ça coûte

exactement 20 couronnes; **hodet mitt rekker ham a- til skulderen** ma tête lui venait juste à l'épaule; **det er a- de samme folkene** ce sont exactement les mêmes personnes; **det er a- det samme for meg** ça m'est complètement égal; **a- som** tout (*el* exactement) comme; **hun gjør a- som hun vil** elle n'en fait qu'à sa tête; **han er ikke a- noe geni** il n'est vraiment pas génial (2) interj: **var det slik det var?** – **akkurat!** cela s'est vraiment passé comme ça? – exactement! (*el* parfaitement!, absolument!)

Den delen av artikkelen som ubetinget er blitt bedre er den som skal beskrive bruken av *akkurat* i tidsuttrykk. For andre typer uttrykk er det enda ikke klart når man skal bruke de forskjellige oversettelses-ekvivalentene. Med et større tospråklig korpus ville det forhåpentligvis være lettere å finne ut mer om bruken av *juste* og *exactement*. Det er langt fra klart om man kan generalisere ut fra eksemplene som nå finnes i artikkelen. Her ville nok et enspråklig fransk korpus være til mer nytte for å se hvilke ord *juste* og *exactement* kollokerer med.

Litteratur

Sitert skjønnlitteratur

- Cornwell, Patricia 1993: *All that remains*. London: Warner Books
 Cornwell, Patricia 1993: *Et il ne restera que poussière ...* (overs. av Gilles Berton). Paris: Éditions du Masque
 Cornwell, Patricia 1995: *Det som blir igjen*. (overs. av Arne-Carsten Carlsen) Oslo: Den svarte serie, Gyldendal Norsk Forlag
 Faldbakken, Knut 1988: *Bad Boy*. Oslo: Gyldendal Norsk Forlag
 Faldbakken, Knut 1990: *Le Monarque*. (overs. av Éric Eydoux) Paris: Presses de la Renaissance
 Gaarder, Jostein 1991: *Sofies verden*. Oslo: Aschehoug
 Gaarder, Jostein 1995: *Le Monde de Sophie*. (overs. av Hélène Hervieu/Martine Laffon) Paris: Seuil
 Staalesen, Gunnar 1982: *Bukken til havresekken*. Oslo: Gyldendal Norsk Forlag
 Staalesen, Gunnar 1994: *Le loup dans la bergerie*. (overs. av Olivier Gouchet) Monaco: Éditions du Rocher
 Steen, Thorvald 1993: *Don Carlos*. Oslo: Tiden Norsk Forlag
 Steen, Thorvald 1994: *Don Carlos*. (overs. av Alain Gnaedig) Paris: Calmann-Lévy

- Wassmo, Herbjørg 1981: *Huset med en blinde glassveranda*. Oslo: Gyldendal Norsk Forlag
- Wassmo, Herbjørg 1987: *La véranda aveugle*. (overs. av Eric og Elisabeth Eydoux) Paris: Actes Sud
- Wassmo, Herbjørg 1989: *Dinas Bok*. Oslo: Gyldendal Norsk Forlag
- Wassmo, Herbjørg 1994: *Le Livre de Dina*. (overs. av Luce Hinsch) Paris: Gaïa Editions

Ordbøker

- Atkins 1994 = Corréard, Marie-Hélène/Grundy, Valerie (red.): *The Oxford-Hachette French Dictionary*. Oxford: Hachette/Oxford University Press
- Duval, Alan/Vivian Marr (red.) 1995: *Collins-Robert Comprehensive English-French French-English Dictionary*. HarperCollins Publishers og Dictionnaires le Robert
- Grundt, Lars Otto 1994: *Stor norsk-fransk ordbok*. 2. utg. Oslo: Universitetsforlaget

Faglitteratur

- Baker, Mona 1995: Corpora in Translation studies. An Overview and Some Suggestions for Future Research. I: *Target* 7:2, 223–243
- Gellerstam, M. 1986: Translationese in Swedish novels translated from English. I: L. Wollin & H. Lindquist (red.): *Translation studies in Scandinavia*. Lund: CWK Gleerup, 88–95
- Hartmann, R.R.K. 1994: The Use of Parallel Text Corpora in the Generation of Translation Equivalents for Bilingual Lexicography. I: *Euralex 1994 Proceedings*. Amsterdam, 291–297
- Hofland, Knut 1995: A Program for Aligning English and Norwegian Sentences. Innlegg fra ALLC/ACH '95, Santa Barbara, 11.–15. juli 1995
- Jacobsen, Tove 1992: En type "bevegelsesverb" i norsk, og hvordan de er behandlet i noen norsk-franske ordbøker. I: Ruth Vadtvedt Fjeld (red.): *Nordiske studier i leksikografi. Rapport fra Konferanse om leksikografi i Norden 28.–31. mai 1991*. Oslo: Nordisk forening for leksikografi, 532–539
- Johansson, Stig/Hofland, Knut/Ebeling, Jarle 1996: Coding and aligning the English-Norwegian parallel corpus. I: Karin Aijmer, Bengt Altenberg, Mats Johansson (red.): *Languages in contrast. Papers from a Symposium on Text-based Cross-linguistic Studies, Lund, 4–5 March 1994*. Lund: Lund University Press, 87–112

- Johansson, Stig/Ebeling, Jarle: The English-Norwegian Parallel Corpus: Introduction and Applications. I: *Proceedings from XXVIII International Conference on Cross-language Studies and Contrastive Linguistics*.
- Salkie, Raphael 1995: Parallel Corpora, Translation equivalence and contrastive Linguistics. (Abstract for ACH/ALLC '95, 11–15. juli 1995 Santa Barbara
- Snell-Hornby, Mary 1984: Towards a Learners Bilingual Dictionary. I: Anthony Cowie (ed.): *The Dictionary and the Language Learner. Papers from the EURALEX seminar at the University of Leeds, 1–3 April 1983*. Tübingen: Niemeyer, 159–170
- Varantola, Krista 1994: The Dictionary User as Decision Maker. I: *Euralex 1994 Proceedings*. Amsterdam, 606–611
- Vinay J.P./Darbelnet J. 1958: *Stylistique comparée du français et de l'anglais*. Paris: Didier

Lars G. Johnsen & Torbjørn Nordgård

Korpusbasert leksikonbygging

The article outlines the motivation for two research projects relevant for computational lexicography applied to Norwegian: a text corpus and a computational dictionary. The corpus must be composed with special attention to the rather special status for written Norwegian because Norwegian has two written standards (Bokmål and Nynorsk). It is argued that these standards contain substandards which are not described systematically. We argue that such substandards can be discovered by careful use of corpora, and the substandards can be represented as different electronic dictionaries to be used in for instance proof reading systems.

1. Innledning og sammendrag

I denne artikkelen blir det skissert hvordan korpus kan brukes i konstruksjon og bygging av et datamaskinelt leksikon innenfor det språkteknologiske samarbeidet *Norsk Infrastruktur For Språkteknologi* (NIFST). Dette samarbeidet startet våren 1995 mellom de språkteknologiske miljøene ved universitetene i Bergen (UiB), Oslo (UiO) og Trondheim (NTNU), og resulterte i følgende ønskeliste for ressurser og verktøy for norsk språkteknologi:

- i. Et norsk tekstkorpus
- ii. Et komputasjonelt ordleksikon
- iii. En tagger for morfologisk analyse og merking av ordene i en vilkårlig tekst
- iv. En grammatikk for automatisk syntaktisk analyse og generering av tekst

Norges Forskningsråd har bevilget støtte til utvikling av tagger¹ ved UiO og konstruksjon av komputasjonelt leksikon (NorKompLeks) ved NTNU. Gjennom NIFST sikres disse prosjektene en koordinert utvikling. Sentralt i koordineringsarbeidet er å utvikle et system for klassi-

¹ En tagger er et program som merker ord mht. ordklassetilhørighet, morfosyntaktiske egenskaper etc.

fisering av ord og fraser som brukes av både tagger, grammatikk og leksikon.²

I det følgende skal vi gå inn på korpus og leksikon. Koblingen mellom de to blir illustrert ved å se på hvordan hypoteser om stilformer kan valideres eller genereres fra korpuset.

2. Norsk Tekstkorpus

Humanistisk datasenter ved UiB har ansvaret for å gjennomføre oppbyggingen av et norsk tekstkorpus. Fra tidligere finnes ikke noe offisielt korpus til bruk i studier av norsk, og prosjektets formål er å gi forskere en felles empirisk ressurs for støtte til filologiske eller lingvistiske studier, spesielt med tanke på å danne et grunnlag for språkteknologisk utvikling. Prosjektets navn er Allmennspråklig Korpus (ASK), og det fokuserer på norsk skriftspråk i alle dets varianter med vekt på bokmål og nynorsk og delvarianter av disse.

Korpusets formål dikterer dets struktur og størrelse, forstått som det utvalg og den kvantitet av tekster som inngår. Gitt koblingen med det komputasjonelle leksikonet vil informasjonen om norsk som er beskrevet der, legge bestemte føringer på hvilke tekster eller teksttyper som skal inngå. Leksikonet vil, foruten kvantitativ informasjon, først og fremst inneholde kvalitativ informasjon om et ord og dets bøyingsformer. Et korpus balansert med hensyn til ytre faktorer vil ikke i seg selv gi noen garanti for at den ønskede informasjon er til stede.

Det er fire grammatiske moduler korpuset skal bevitne:

- (1) Leksikonet som en liste av ord, stammer og affikser, idiomer og kollokasjoner
- (2) De leksikalske regler som beskriver hvordan ord bygges opp til nye ord, og hvordan egenskaper til ord endrer seg i denne prosessen
- (3) De syntaktiske regler som beskriver hvordan ord og fraser kan settes sammen til nye fraser
- (4) Stilvalg som antyder hvilke systematiske delstandarder av språket som er i bruk

² Denne klassifiseringen er ikke klargjort ennå, men taggerprosjektet og leksikonprosjektet samarbeider om distinksjoner og grammatisk terminologi. Det må bl.a. avgjøres om den grammatiske beskrivelsen skal være i form av trekkstrukturer eller lister av kategorier.

Av disse modulene vil (1–2) og (4) kodes i NorKompLeks. Men det er spesielt (2) og (4) som legger en særlig føring på korpusets innhold. Med hensyn til (1) ville strategien for å sikre god dekning ganske enkelt være å samle flest mulig tekster med størst mulig variasjon. Problemet med å realisere (1), særlig for enkelte germanske språk og spesielt norsk, er at de har forholdsvis frie regler for ordsammensetninger slik at det vil være umulig å sikre at alle ord som finnes i språket, skal være representert i korpuset. Dette, sammen med de økonomiske og administrative føringene som ligger på selve størrelsen av korpuset, betyr at utvalget i større grad må bestemmes av representativiteten av de språklige egenskaper som *kan* kontrolleres.

I forbindelse med (2) kan vi se på de regler som lager substantiv av verb og som oppretter systematiske forbindelser mellom verbets syntaktiske kontekst og konteksten for det resulterende substantiv. For eksempel, regelen som legger affikset *ing* til enkelte verb, vil systematisk ta med seg noen av verbets kontekster, men ikke alle. For *spise* gjelder det at *spisingen av eplet* er helt i orden, mens *spisingen av opp eplet* ikke er mulig. Slike og andre eksempler er beskrevet i detalj i Johnsen *et.al.* (1989).

Dersom et korpus skal kunne bevitne slike leksikalske sammenhenger, må de ganske enkelt forekomme i korpuset, og tekstene må velges slik at det er en sjanse for å finne dem. Vi skulle forvente *a priori* at et utvalg av tekster som hadde høy hyppighet av enkelte ordstammer, vil ha større sjanse til å bevitne slike sammenhenger enn tekster der variasjonen i ordstammer er stor gitt samme størrelse på utvalget. Studiet av delspråk beskrevet i McEnery & Wilson (1996) støtter opp om dette. Dersom en begrenser tekstutvalget til å gjelde tekster som omhandler spesifikke områder³, vil et bestemt leksikalsk element opptre i flere varianter enn i et generelt utvalg av tekster. For våre behov betyr det at i korpuset bør en substansiell del være viet et delspråk som inneholder forholdsvis få ordstammer.

For å kunne dekke stilarter må man ta hensyn til hvilke uoffisielle "stilnivå" tekstutvalgene er basert på. Stilnivå og stilarter vil vi ta for oss i seksjon 4, men for å kunne representere dem fordres det hypoteser om hvor man kan vente å finne variasjon i stil.

3. NorKompLeks

³ I dette tilfelle IBM-manualer versus Canadian Hansard Corpus.

Prosjektet *Norsk komputasjonelt leksikon* (NorKompLeks) har som målsetting å realisere et språkteknologisk leksikon for norsk. Norsk blir forstått som de to skriftspråksstandardene bokmål og nynorsk. I tillegg skal det utarbeides et omfattende fonologisk leksikon for norsk, basert på leksemene i bokmål og nynorsk.

3.1. Leksikalsk informasjon i NorKompLeks

I NorKompLeks vil oppslagsordene ha følgende typer informasjon knyttet til seg: *kategori*, *grafemisk form*, *fonologisk form*, *morfologiske paradigmer*, *morfosyntaktiske egenskaper*, *syntaktiske egenskaper* og *semantiske egenskaper*. Informasjonen er organisert som rekursive trekkstrukturer, jf. Shieber (1986), noe som gir et deklarativt og oversiktlig representasjonsformat som det er enkelt å modifisere, det være seg med mer informasjon, mindre informasjon eller annen informasjon. Et eksempel er angitt i figur 1, der leksemets grafemiske form er *bil*, ordklassekategorien er *N* (nomen), identifikasjonskoden er 5868, den fonologiske beskrivelsen er */bi:l/*, det morfologiske bøyingsparadigmet er *m1*, den morfosyntaktiske egenskapen *form* har verdien ubestemt, den morfosyntaktiske egenskapen *kjønn* har verdien hankjønn, den morfosyntaktiske egenskapen *tall* er entall, og den morfosyntaktiske egenskapen *type* er appellativ. Videre er de semantiske egenskapene *masseterm* og *animat* markert som negative.

ID:	5868								
Kategori :	N								
Grafemisk form :	"bil"								
Fonologisk form :	/bi:l/								
Morfologiske paradigmer :	{ m1 }								
Morfosyntaktiske egenskaper	<table border="1"> <tr> <td>Form :</td> <td>Ubestemt</td> </tr> <tr> <td>Kjønn :</td> <td>Hankjønn</td> </tr> <tr> <td>Tall :</td> <td>Entall</td> </tr> <tr> <td>Type :</td> <td>Appellativ</td> </tr> </table>	Form :	Ubestemt	Kjønn :	Hankjønn	Tall :	Entall	Type :	Appellativ
Form :	Ubestemt								
Kjønn :	Hankjønn								
Tall :	Entall								
Type :	Appellativ								
Semantiske egenskaper	<table border="1"> <tr> <td>Masseterm :</td> <td>nei</td> </tr> <tr> <td>Animat :</td> <td>nei</td> </tr> </table>	Masseterm :	nei	Animat :	nei				
Masseterm :	nei								
Animat :	nei								

Figur 1: Leksikonoppslaget **bil**

Vi skal i det følgende konsentrere oss om den morfologiske informasjonen.

3.2. Morfologiske beskrivelser i NorKompLeks

Morfologikoden *m1* skal fortolkes som en instruksjon om hvordan bøyingsparadigmene til leksemets grunnform skal produseres. I dette tilfellet er det enkelt: Ubestemt form entall er lik grunnformen, bestemt form entall dannes ved å legge til bokstavene *en*, ubestemt form flertall dannes ved tillegg av *er*, mens *ene* i tillegg til grunnformen gir bestemt form flertall. Andre instruksjoner kan erstatte en eller flere bokstaver i grunnformen med andre bokstaver, noe som er typisk for omlydsbøyninger. I NorKompLeks-prosjektet er det definert og implementert et system som ekspanderer morfologiske koder til fullstendige paradigmer; for mer detaljerte spesifikasjoner, se Nordgård (1996).

Norsk er spesielt mht. morfologisk valgfrihet og liberal, eller kanskje mer presist, ufullstendig normering. Dette har å gjøre med den norske språksituasjonen, noe vi kommer tilbake til nedenfor. I figur 2 finner man et utsnitt av beskrivelsen av verbet *stryke*.⁴ Legg spesielt merke til at det finnes to sidestilte preteritumsvarianter: *strøk* og *strauk*. Det spesielle her er at vi finner to likestilte morfologiske realisasjoner av samme morfologiske kategori, i dette tilfellet preteritum av verbet *stryke*.

Imperativ : "stryk"

Infinitiv : "stryke"

Presens : "stryker"

Preteritum : { "strauk", "strøk" }

Perfektum partisipp "strøket"

Presens partisipp : "strykende"

Perifrastisk passiv "strøket"

Inherent passiv : "strykes"

Figur 2: Bøyingsparadigmet til *stryke*

⁴ Vi ser her bort fra de deriverte adjektivformene *strøket*, *strøkne* og *strykende*. Disse er med i den implementerte versjonen av leksikonet.

I bokmålsnormen er det faktisk slik at mer enn halvparten av alle verb har alternative bøyingsmønstre. Omkring 500 verb kan bøyes etter tre alternasjoner, f.eks. *anklage* som kan ha preteritumsformene *anklaget*, *anklaga* eller *anklagde*. Enkelte verb kan bøyes på fire måter, f.eks. *love*, der man kan bruke preteritumsformene *lovet*, *lova*, *lovde* og *lovte*.

Bokmålsnormen er altså liberal mht. hva som er tillatt i skriftlig norsk. Men det finnes en del stilnivåer som ikke bør brytes. Eksempelvis bør man ikke bruke formen *strauk* (av verbet *stryke*) og formen *anklaget* (av verbet *anklage*) i samme tekst; *strauk* og *anklaga* hører mer naturlig sammen, men det er ingen som kan referere til mer eller mindre offisielle normer i slike tilfeller. Før vi går videre i gjennomgangen av NorKompLeks og forholdet til korpusmateriale, er det nødvendig med en kort gjennomgang av noen aspekter ved den norske språksituasjonen.

3.3 Den norske språksituasjonen

Det er som kjent to skriftspråkstandarder for norsk. *Bokmål* har en historie som går tilbake til dansk, mens nynorsk ble etablert i forrige århundre på grunnlag av norske dialekter og gammelnorsk (norrønt). De to skriftspråkene er offisielt likeverdige i den forstand at man kan velge hvilket man vil bruke.

Det har alltid vært tendenser til polarisering i begge språkleirene. Den mest konservative bokmålssiden assosieres gjerne med *riksmål*. Forenklet kan vi si det slik at riksmålsforkjemperne helst unngår alle rettskrivningsreformer som får bokmål til å ligne nynorsk. Den radikale bokmålsfløyen vil derimot akseptere ordforråd og bøyingsformer som ligger nært opp til nynorsk. De fleste bokmålsbrukerne kan lokaliseres til et sted mellom riksmål og radikalt bokmål, men det er i dag ikke mulig å karakterisere flertallsnormen(e) på empirisk basis. I nynorskleiren finner vi også en konservativ fløy. Den er bl.a. kjennetegnet ved at den vil bevare et nynorsk som ligner mest mulig på norrønt og Ivar Aasens opprinnelige forslag, innenfor gjeldende valgfrihet. Den andre fløyen vil ha et nynorsk som både ligger nærmere opp til radikalt bokmål og observert språkbruk i dialektene.⁵ Det finnes et slags kompromiss mellom (radikalt) bokmål og (moderat) nynorsk, og det kalles gjerne *samnorsk*. Samnorsk kan karakteriseres som unionen

⁵ Denne leiren benevnes ofte som "moderat nynorsk", men "moderat" nynorsk må ikke forveksles med det vi her kaller konservativt nynorsk.

av (radikalt) bokmål og (moderat) nynorsk, gjerne med så lite normering som mulig.

På denne bakgrunn kan vi systematisere skriftlig norsk som seks ulike varianter som vi skal kalle *stilnivåer*: riksmål, moderat bokmål, radikalt bokmål, samnorsk, moderat nynorsk og konservativt nynorsk. Dette er en grovklassifisering, og man kan sikkert argumentere for andre inndelinger. Likevel vil vi i det følgende holde oss til denne taksonomien.

3.4 Morfologiske stilnivåer

Stilnivåene bør ikke rotes sammen. Følgende eksempel blander sammen egenskaper ved alle stilnivåene:

(1)

Enhver offentlig tenestemann har høve til å leite efter sproglige inkonsistensar som me finn i sume tekster.

Vi vil nok aldri finne et så ekstremt tilfelle av stilsammenblanding som i eksempel (1). De aller fleste skrivekyndige nordmenn har en ubevisst eller innlært stilsans. Selv om ingen kan forby noen å sette sammen ord fra ulike skriftspråksvarianter som vist ovenfor, vil sammenblandinger fortone seg som komiske, og det er lett å la seg irritere av slikt. Gjeldende normering av bokmål tillater ikke ord som *sume*, *me* og *tenestemann*. Derimot er det tillatt å formulere seg slik i henhold til gjeldende bokmålsnorm:

(2)

Enhver offentlig tjenestemann har anledning til å leite etter språklige inkonsistenser som vi finner i somme tekster.

Det er grunn til å anta at nordmenn flest vil reagere på ordvalget i en slik setning. Ord som *enhver* og *anledning* passer dårlig sammen med *leite* og *somme*. Men vi kan ikke referere til noen definerte normer som er brutt, bortsett fra at ord som *leite* og *somme* ikke er aksepterte riksmålsord.

3.5 Stavemåter og stilnivåer

Det kan også identifiseres systematiske valg av stavemåter som kan knyttes til de ovenfor nevnte stilnivåene. Verbene *støype* og *støpe* er sidestilte former i bokmål i den forstand at de er sidestilte stoveformer av samme leksem. Videre er stavemåten *røyke* en hovedform, mens *røke* er sideform.⁶ Dersom man velger å bruke alternativet *røke*, bør man også velge *støpe*, og man bør også holde seg til preteritumsformen *strøk* og ikke *strauk*, og man bør unngå *anklaga* og *lova*.

3.6 Leksemer og stilnivåer

Et leksem som substantivet *beriktigelse* bør ikke kobles sammen med radikalt bokmål, mens verbet *skeise* (å gå på skøyter) ikke hører naturlig sammen med riksmål og moderat bokmål.

4. Korpora og deduksjon av skriftstandarder

Vi har sett at bøyingsformer, stavemåter og leksemvalg kan, og etter vårt syn *bør*, systemiseres. Gitt den norske språksituasjonen kan man ved hjelp av korpusmateriale være interessert i å gjøre følgende:

- Beregne frekvensen til godkjente leksemer
- Beregne hvilke stavemåter som foretrekkes dersom det er valgfrihet (*støpe* eller *støype*)
- Beregne forholdet mellom hovedformer og sideformer (*røyke* eller *røke*)
- Identifisere nye leksemer
- Finne ut hvilke bøyingsalternasjoner som er mest frekvente der det er valgfrihet (*strøk/strauk*, *lovet/lova/lovde/lovte*)

Korpusmaterialet kan gjøre det mulig å etablere *empiriske* kriterier for hva som mest naturlig skal være leksikalske og morfologiske hovedformer. Resultatene vil naturligvis være avhengig av hvilket korpusmateriale som legges til grunn. Derfor vil spørsmål angående representativiteten til korpusmaterialet være meget viktige.

4.1 Frekvensen av tillatte leksemer

⁶ At to former eller leksemer er sidestilte, også kalt jamstilte, betyr at begge er fullt ut aksepterte. En sideform er derimot marginalisert og skal ikke brukes i godkjenningspliktige skolebøker eller i offentlig forvaltning.

Ved hjelp av store taggede korpora er det mulig å beregne frekvensen til leksemer som er med i ordlisten. Dette forutsetter at taggeren tar med leksemkodene i merkingen av ordene i teksten. I NorKompLeks er alle lemmaer utstyrt med numeriske identifikasjonskoder, og det skal derfor være mulig å etablere en akseptabel frekvensordliste for norsk basert på leksemenes frekvens (til forskjell fra "naive" frekvensordlister som simpelthen teller opp frekvensen til bøyde ord i løpende tekst), naturligvis forutsatt at taggeren benytter NorKompLeks-ordlistene.

4.2 Frekvensen til sidestilte bøyingsformer

En mer avansert utnyttelse av korpuset er å foreta morfologiske analyser av de bøyde formene for på den måten å identifisere hvilke bøyingsalternativer som benyttes. Dermed kan man på *empirisk* grunnlag undersøke bruken av sidestilte (jamstilte) bøyingsalternativer, f.eks. alternasjonen *strøk/strauk*. Merk at man kan både beregne den totale bruken av en bestemt bøyingsklasse, og bruken av en bestemt klasse i relasjon til en annen i et bestemt leksem.

4.3 Sideformer og hovedformer

Gitt at man ved hjelp av korpus kan beregne frekvensen til sidestilte former, kan man også finne ut i hvilken grad en hovedform virkelig er mer utbredt enn sideformene.

4.4 Leksemer

Det er liten tvil om at bokmål tillater mange leksemer som hovedformer selv om de i praksis sjelden eller aldri er i bruk. Korpusmateriale setter oss i stand til å identifisere slike leksemer, ganske enkelt ved å inspisere frekvensopptellingen på leksemnivå, jf. 5.1 ovenfor.

4.5 Empirisk motiverte skriftstandarder

Kombinasjonen av korpora og maskinleselige ordbøker setter oss i stand til å identifisere skriftstandarder på empirisk grunnlag. Troverdigheten av slike standarder er direkte avhengig av tekstutvalget som bygger opp korpusmaterialet. Man kan se for seg et par hovedstrategier. Den ene er å samle sammen alt tilgjengelig elektronisk tekst-

materiale for bokmål og beregne frekvenser og foreta opptellinger på dette materialet. En annen strategi kan være å lage flere tekstsamlinger som tar hensyn til omtrentlige standarder slik de er definert ovenfor, eventuelt kombinasjoner av dem. Den første strategien er risikabel i den forstand at tekster som favoriserer en skriftstandard, kan bli overrepresentert. Store og balanserte korpora er et bedre alternativ, men det vil ta tid å lage slike. Strategi nummer to vil få problemer med å bestemme hvor ulike tekster skal kategoriseres.

4.6 Empirisk motiverte skriftstandarder

Vi har så vidt vært inne på at leksemvalg, valg av bøyingsvarianter, sideformer og hovedformer er med på å bestemme det vi har kalt stilnivåer. Det vil være svært vanskelig å definere slike stilnivåer på en måte som skaper tilstrekkelig oppslutning til at de tas i bruk. Derimot vil balanserte korpora kunne utnyttes til å utlede slike stilnivåer. Dette stiller naturligvis strenge krav til sammensetningen av tekstene, men som metodisk prinsipp er en slik empirisk tilnærming meget interessant.

5. Konklusjoner

Korpusbasert identifikasjon av skriftspråksstandarder vil være et nytt redskap for norsk skriftspråksnormering, forutsatt at det finnes et språkpolitisk klima som aksepterer at slik aktivitet. Resultatene vil gjøre det relativt enkelt å lage elektroniske ordlister for de stilnivåene man finner. Slike ordlister kan i sin tur inkorporeres i tekstbehandlingssystemer, og vi vil kunne få programmer for stavelseskontroll som er følsomme for ulike stilnivåer. Dette vil utvilsomt være et kjærkomment hjelpemiddel for folk som har problemer med å beherske skriftlig norsk, f.eks. utlendinger og folk med lese- og skriveproblemer.

Når det gjelder norsk leksikografisk arbeid, vil slike nye muligheter sette oss i stand til å produsere ordlister med nyttig informasjon til brukerne, f.eks. ved at stilistiske standarder tas med i fremtidige utgivelser av norske ordbøker, samt at frekvensinformasjon kan inkluderes.

Referanser

- Johnsen, Lars/Anneliese Pitz/Lars Hellan 1989: *TROLL (The Trondheim Linguistic Lexicon Project)*. Lingvistisk institutt, NTNU, Trondheim.
- McEnery, Tony/Andrew Wilson 1996: *Corpus Linguistics*. Edinburgh Textbooks in Empirical Linguistics, Edinburgh University Press, Edinburgh.
- Nordgård, Torbjørn 1996: *NorKompLeks 1.1*. Manuskript, Lingvistisk institutt, NTNU, Trondheim
- Shieber, Stuart 1986: *An Introduction to Unification-Based Approaches to Grammar*. CSLI Lecture Notes nr. 4, Stanford University.

Göran Kjellmer

Idiomen, kollokationerna och lexikonet

Against the background of the extensive use of language corpora in the dictionary-making industry of today, the paper stresses the importance of including set phrases in future lexicographic undertakings. A distinction is made between idioms and collocations, and it is argued that while the significance of the former category has long been recognised, the latter has only recently come to attract the attention of lexicographers. To gauge the relative importance of the two categories in modern English, a few experiments were undertaken, the results of which indicated clearly that collocations are greatly superior to idioms in frequency and can be said to make up the backbone of language. Consequently a plea is finally made for the inclusion of larger quantities of collocations in dictionaries of the future.

1. Revolution inom lexikografin

Den snabba utvecklingen inom datatekniken har gjort det möjligt att skapa ordsamlingar av tidigare oanade och otänkbara dimensioner. Genom att en handfull hängivna lingvister tidigt insåg deras möjligheter och gjorde dem användbara för vetenskapligt bruk har villkoren för lexikonframställning radikalt förändrats på mindre än ett decennium. Från att inte alls ha använts har korpusar kommit att bli en nödvändig förutsättning för ordböcker skrivna i dag. Här, liksom då och då i fortsättningen, ber jag att få anlägga ett anglistiskt perspektiv. För att demonstrera betydelsen av de nya metoderna inom lexikografin kan man då nämna följande ganska omfattande nya engelska lexika, som samtliga är korpusbaserade:

- COBUILD 1987–, Cobuild Corpus, 20 (7)m 1995 Bank of English, 200m+ ord
- Collins English Dictionary*³ 1991 Bank of English, 200m+ ord
- Longman Language Activator* 1993 ("the World's First Production Dictionary")
Longman Lancaster Corpus, 30m ord
- BBC English Dictionary* 1993 "Based on an analysis of over seventy million words of broadcasting output from 1988 to the present day"
- Cambridge International Dictionary of English (CIDE)* 1995 Cambridge Language Survey, skriven och talad: 100m ord
- Longman Dictionary of Contemporary English (LDOCE)* 1995 British National Corpus (BNC) (100m ord) + Longman Lancaster Corpus (30m ord) + Longman Learner's Corpus
- Oxford Advanced Learner's Dictionary* 1995 BNC 100m ord + Oxford American English Corpus (40m ord)

Concise Oxford Dictionary 1995 BNC 100m ord + dataläsbara citatsamlingar + material från *Oxford English Dictionary* (OED)

Fig. 1. Nya engelska korpusbaserade lexika

2. Fraseologins representation i lexikonen

Språket är till stor del uppbyggt av formelartade uttryck. En mycket viktig komponent i en talares kompetens i ett språk utgöres av hans förmåga att använda sådana fraser i tal och skrift:

Stretches of formulaic and semi-formulaic speech pervade the conversations and monologs of everyday life in all speech communities: when people greet, introduce or take leave of one another; when they buy and sell, congratulate, commiserate, ask, refuse, apologize, bless, curse, pray, say magical spells, or chant skipping rhymes, and when they tell jokes or folk tales, cheer for their team, or call play-by-play commentaries on a game of football. (Andrew Pawley s.v. "Formulaic speech" i Bright 1992: 2/23)

Det är då väsentligt att lexikonen i största möjliga utsträckning redovisar sådana fraser. När nu lexikonmaterialet i så hög grad är hämtat från levande, sammanhängande språk, skrivet och talat, har lexikograferna enastående möjligheter att återge inte bara språkets enskilda byggstenar, orden, utan också de sammanhang de förekommer i, de stående kombinationerna eller fraserna. När man framhåller önskvärdheten av att lexikograferna utnyttjar denna möjlighet, händer det att det invänds att lexikonen i så fall skulle svälla över alla bräddar, då ju mycket stora delar av ordförrådet naturligen förekommer i fasta sammanhang. En sådan invändning påminner om den som opponerar mot att man i lexikonet inkluderar substantiven, som ju också är talrika. Om fraser å ena sidan utgöres av en samling icke förutsägbara kombinationer och å den andra sidan är en integrerad del av språkets lexikala sida kan de, menar jag, göra anspråk på en självklar och framträdande plats i våra ordböcker. I själva verket har dagens lexikografer i betydligt högre grad än sina föregångare kommit till insikt om fraseologins stora betydelse för språkbrukaren. Detta framgår tydligt om man jämför behandlingen av samma ord i tidigare och senare upplagor av samma lexikon, t.ex. behandlingen av *earn* och *job* i *Concise Oxford Dictionary* 1964 och 1995.

Fig. 2. Ur Concise Oxford Dictionary 1964 (t.h.), 1995 (t.v.)

Lexikonets uppgift är att tillhandahålla information om väsentliga och icke förutsägbara delar av språkets ordförråd, och dit hör de (mer eller mindre) fasta uttrycken/fraserna. Då inget lexikon, inte ens de största (t.ex. *Ordbog over det danske Sprog*, *Ordbok över svenska språket utgiven av Svenska Akademien*, *The Oxford English Dictionary*), av rent principiella skäl kan täcka hela ordförrådet (stora delar av ordförrådet existerar som potentiella bildningar, det finns öppna kategorier som räkneorden etc.), måste därför någon form av värderings- och

urvalsprincip etableras i varje typ av lexikografi. Detta gäller således vare sig lexikonet är av den gamla, huvudsakligen ordförtecknande eller den nyare ord- och frasförtecknande typen. Jag skall här något uppehålla mig vid frågan om principerna för införande av fasta förbindelser i lexikonerna och stanna vid två, delvis överlappande, varianter av sådana förbindelser: idiom (de mest kända) och kollokationer. Frågan jag vill diskutera är denna: Tar korpusbaserade lexika hänsyn till idiomerna och kollokationerna i tillräcklig utsträckning? För att kunna besvara den måste vi först begrunda tre andra frågor: (a) Vad är idiom? (b) Vad är kollokationer? (c) Vad är "tillräcklig utsträckning"?

3. Vad är idiom?

"Idiom" definieras vanligen som ett uttryck vars betydelse inte framgår av betydelsen av de i uttrycket ingående orden. Säger man att någon är "i hatten" kan man inte av sin kännedom om betydelsen av *i* och *hatten* sluta sig till betydelsen 'berusad'. *I hatten* är alltså ett idiom. Karakteristiskt för idiomerna är att de har begränsade variationsmöjligheter; om *i hatten* varieras till *i en hatt* eller *i hattarna* försvinner den idiomatiska betydelsen. Total invariabilitet är dock inte ett krav för att ett uttryck skall vara idiomatiskt. *Att slå runt* betyder 'att vara ute och festa', en betydelse som inte är härledbar ur betydelsen av *slå* och *runt* och som alltså är idiomatisk. Samma betydelse finns i grammatiska variationer av uttrycket: *slår runt*, *slog runt*, *slagit runt*.¹ Variationsmöjligheten är trots detta begränsad. Uttrycket kan inte med bibehållen idiomatisk betydelse förekomma i kluvna satser (**Det var runt som han slog*, **Vad han slog var runt*) eller i en topikaliserad sats (**Runt slog han*).

Det semantiska kravet på ett idiom kan förefalla absolut men är i själva verket ganska elastiskt. Detta hänger samman med polysemins oklara gränser hos många ord. För vissa språkbrukare är *penna* 'vingfjäder' och *penna* 'skrivverktyg' ett ord med något olika betydelser, för andra är de två homonyma ord. Om ett ord alltså kan ha olika användningar vilkas inbördes samband inte alltid står klart för språkbrukaren, gäller detta förstås också ordkombinationer. Uttryck som av somliga betraktas som en naturlig utvidgning av de ingående ordens användningsområde kan då av andra ses som metaforiska eller idiomatiska. Skall det svenska fras verbet *se út* 'välja, utvälja', som i "Hon såg ut en lämplig kavaljer åt sin dotter", anses ha en oförutsebar betydelse,

¹ Se Moon 1996 för en exemplifiering av sådana variationsfenomen i engelskan.

även om man känner betydelse hos *se* och *ut*, och skall verbet alltså betraktas som ett idiom, eller kan man inse dess betydelse utifrån delarnas betydelse och därmed bedöma det som icke-idiomatiskt? Olika språkbrukare lämnar säkert olika svar på en sådan fråga. Detta visar, menar jag, att det finns en gråzon mellan idiom och andra mer eller mindre fasta uttryck. Det finns en livlig diskussion om olika uttrycks idiomatiska status, där en titel som "The grading of idiomaticity as a presupposition for a taxonomy of idioms" (Gläser 1988) kan ses som symptomatisk.

4. Vad är kollokationer?

Termen **kollokation** har definierats olika av olika forskare, men det gemensamma för dem är att kollokationer är två eller flera ord som ofta förekommer i varandras närhet. Något krav på idiomatisk betydelse föreligger inte. Exempel på svenska kollokationer är nominalfraser som *en ljus framtid*, *goda förhoppningar*, *helvetets portar*, prepositionsfraser som *av egen kraft*, *i största hast*, *med stor möda*, verbfraser som *hugga ved*, *tappa humöret*, *äta middag*, *borsta tänderna*, och t.o.m. satser som *som du vet*, *när allt kommer omkring*, *om Gud vill* o.s.v. Också här kan man observera att variationsmöjligheterna är begränsade, jfr **när mycket kommer omkring*, **om Gud har velat*. Om man släpper kravet på komponenternas omedelbara närhet, "kontiguitet", kan man också som kollokationer räkna ordgrupper som *ko: kalv: beta*, *gräs: grön*, *skriva: läsa: brev* o.s.v. (se t.ex. Halliday & Hasan 1976:287).

5. Glidande skala idiom: andra kollokationer

Det är då uppenbart att gränsdragningen mellan idiom och kollokationer är svår att uppehålla. Det är naturligt att se idiomerna som en ganska speciell sorts kollokationer med – ofta – klara semantiska egenheter, men var gränsen mellan idiomatiska och icke-idiomatiska kollokationer går är vanskligt att säga. När idiomerna diskuteras brukar man begränsa sig till sådana som otvetydigt uppfyller det semantiska kravet, d.v.s. att betydelsen av helheten inte skall vara härledbar ur betydelsen av delarna. När jag i fortsättningen använder termen är det denna grupp som avses.

6. Idiom mer uppmärksammade än kollokationer

Idiomen är färgstarka inslag i språket, och det är naturligt att deras förekomst sedan länge har uppmärksammats. På engelskt område har under flera decennier pågått en livlig debatt om idiomens natur, där deras analys, förståelse och lexikografiska behandling har diskuterats. Förutom de artiklar där debatten ägt rum har en rad idiomordböcker och -samlingar² publicerats. Skolundervisningen, i varje fall i Sverige, har vederbörligen tagit intryck, och hos skolbarn inpräntas vikten av att känna till uttryck av typen *raining cats and dogs*, *skating on thin ice*, *spilling the beans*. Intresset för de mindre färgstarka och på sätt och vis mera anonyma övriga kollokationerna har varit mindre påfallande under denna tid. Inom lexikografien har som nämnts uppmärksamheten först under senare år kommit att riktas mot hithörande frågor i någon större utsträckning. Om man vill försöka bedöma i vilken utsträckning det är rimligt att ta hänsyn till idiom och övriga kollokationer i en ordbok kan det vara av intresse att se i vilken utsträckning de faktiskt förekommer i allmänspråket. En liten sådan undersökning skall här redovisas.

7. Undersökning av faktisk förekomst av idiom i engelska korpora.

Idiomen i en (kort) engelsk idiomordbok (Collins 1958), som säger sig innehålla de allra vanligaste idiomerna³ jämfördes med sin eventuella förekomst i den amerikanska Brown-korpusen, så som de förtecknats i *A Dictionary of English Collocations* (Kjellmer 1994). Materialet i de båda samlingarna härrör från ungefär samma tid. Antalet idiom som förekom på båda ställena kunde då ses som ett mått på idiomens vanlighet i engelskan. En kontroll fick göras att frasen i kollokationsordboken faktiskt användes som idiom i korpusen. T.ex. fanns *on the shelf* med men inte i den idiomatiska betydelsen 'på överblivna kartan' utan i den ordagranna betydelsen 'på hyllan'. Sådana fraser fick utgå.

Några reservationer måste göras. Collins 1958 är brittisk medan Brown-korpusen är amerikansk. Detta spelar dock mindre roll när det gäller de allra vanligaste idiomerna, som till allra största delen är gemensamma för de två varianterna av engelska. En annan reservation gäller *A Dictionary of English Collocations*, som förtecknar de fraser

² T.ex. Allwood 1965, Curry 1985, Hargevik & Ljung 1989, Hübinette & Odenstedt 1988, Laycock and Allwood 1959, Makkai 1972, Orr 1962, Schenk 1994.

³ "[O]ut of the thousands [of idioms] in the language, the natural course has been to choose those that are today in most frequent use", p. ix.

som förekommer minst två gånger. Engångsförekomster är alltså utslutna. Detta bör då påverka statistiken något.

Collins 1958 innehåller 888 olika idiom "of the thousands ... that are today in most frequent use".⁴ Frågan är då hur många av dessa som återfinns i den amerikanska miljonordskorpusen. En inte alltför vild gissning skulle väl kunna vara c. 50%.

No. i Collins	"Idiom" i Collins belagt i Brown	Idiomatiskt bruk i DEC
4	in the air	ja
10	all over	nej
25	as it were	ja
26	asking for it	ja
27	at that	ja
71	blue ribbon	ja
74	out of the blue	ja
100	to bring home	ja
114	by and large	ja
122	carry on	ja
169	come to	nej
180	cup of tea	ja
195	cut out	ja
208	(for) a rainy day	ja
226	pass away	ja
227	passing	ja
228	to the death	ja
269	on earth	ja
304	in the face of	ja
308	on the face of (it)	ja
327	a far cry from	ja
376	friend at court	?
418	toe the line	ja
459	against the grain	ja
505	turn one's head	nej
542	knew all the answers	ja
567	in the long run	ja
582	make it	ja

⁴ För den aktuella undersökningen har de alla accepterats såsom "idiom", trots att deras idiomstatus emellanåt är diskutabel.

603	soft		nej
640	okay (O.K.)		ja
677	a pot		nej
742	rule of thumb		ja
759	by the same token		ja
794	on the shelf		nej
796	in his shoes		ja
803	a showdown		ja
808	shut (slam) the door		nej
828	so long		nej
832	something in the wind		nej
868	square with		nej
911	take to task		ja
960	in touch		ja
967	turn down		ja
975	up to him		ja
Summa		44	10 nej, 33 ja, 1 ?

Fig. 3. Idiom i Collins 1958 som återfinns i *A Dictionary of English Collocations*

8. Resultat

Vissa potentiella idiom visar sig vara icke-idiomatiska kollokationer.⁵ Men av "those that are today in most frequent use" förekommer endast (44 orensade av 888 = 4.95% och) 33 rensade av 888 = 3.72%. Idiomen är alltså minst sagt dåligt representerade i korpusen och fjärran från den oinformerade gissningens 50%. Inte ens de nästan legendariska *an axe to grind*, *carry coals to Newcastle*, *drop a brick*, *kick the bucket* eller *rain cats and dogs* uppträder i korpusen.

9. Vad är "tillräcklig utsträckning"?

⁵ Jfr Bar-Hillel: "When a certain phrase occurs in the idiom dictionary, this does not necessarily mean that *all* occurrences of this phrase within any conceivable context have to be treated as idioms and rendered accordingly. Even our *red herring* might occasionally, when used in describing a painting by Marc Chagall, for example, have its so-called 'literal' meaning ..." (Citerat från Sinclair, Jones & Daley 1970:94.)

John Sinclair har beskrivit hur han bestämde sig för att undersöka förekomsten av det typiska idiomet *red herring* 'villospår' i engelskan.⁶ I den textkorpus han undersökte fann han inte ett enda exempel härpå. Detta är inte särskilt förvånande. Idiomerna är närmast "lyxartiklar" i förhållande till vanliga kollokationer. För den aktive språkbrukaren – talaren eller skribenten – är kollokationskännedom en absolut nödvändighet, medan idiomkännedom har mera kosmetiskt värde och intresse; han kan producera ett fullgott språk utan att använda några idiom så länge han behärskar sina kollokationer. För den passive språkbrukaren – lyssnaren eller läsaren – har idiomerna ett större, men ändå begränsat intresse: större, eftersom han bör förstå dem när de uppträder, men begränsat eftersom de är sällsynta i förhållande till kollokationerna. En infödd talare är väl ofta mindre intresserad av att i ett lexikon söka de högfrekventa kollokationer han redan behärskar och använder utan att närmare reflektera över dem; han brukar vara mer intresserad av de färgstarka idiom som fångar hans uppmärksamhet. En utlänning å andra sidan är i störst behov av riklig kollokationsrepresentation. Lexikonets karaktär bör då variera beroende på vilken typ av användare utgivaren tänker sig. Många ordböcker vänder sig till båda kategorierna, men i varje fall inom det engelska området fokuseras intresset främst på användare för vilka engelskan är ett främmande språk. Med tanke på dessa förefaller det alltså som om man bör ställa krav på ett modernt lexikon att det skall ha en rimlig idiomrepresentation och en mycket god kollokationsrepresentation. Var sedan gränsen dras är en fråga om ekonomiska resurser, avsedd marknad och andra sådana icke-språkliga hänsyn.

10. Illustration

För att konkretisera dessa synpunkter skall jag presentera idiom och kollokationer i ett kort stycke skönlitterär text, början av Graham Greenes *Travels with My Aunt*. För att identifiera kollokationerna har jag återigen använt *A Dictionary of English Collocations*. I texten har kollokationerna kursiverats och idiomerna gjorts fetstilta. I överlappande kollokationer har det överlappande elementet understrukits.⁷

Graham Greene, *Travels with My Aunt*

⁶ Sinclair, J. McH., Jones, S., & Daley R. 1970:94.

⁷ Skrivningen *For those reasons* avser att markera att även om *for those reasons* inte återfinns som kollokation, flera andra snarlika ändå gör det: *for many/several/social/two reasons*.

I met my Aunt Augusta for the first time in more than half a century at my mother's funeral. My mother was approaching eighty-six when she died, and my aunt was some eleven or twelve years younger. I had retired from the bank two years before with an adequate pension and a silver handshake. There had been a take-over by the Westminster and my branch was considered redundant. Everyone thought me lucky, but I found it difficult to occupy my time. I have never married, I have always lived quietly, and, apart from my interest in dahlias, I have no hobby. For those reasons I found myself agreeably excited by my mother's funeral.

My father had been dead for more than forty years. He was a building contractor of a lethargic disposition who used to take afternoon naps in all sorts of curious places. This irritated my mother, who was an energetic woman, and she used to seek him out to disturb him.

I met my Aunt Augusta for the first time in more than *half* a century at my mother's funeral. My mother was approaching eighty-six when she died, and my aunt was some eleven or twelve *years* younger. I had retired from *the bank two years* before with an adequate pension and a **silver handshake**. There *had been* a take-over by the Westminster and my branch *was considered* redundant. Everyone thought me lucky, but *I found it difficult to occupy* my time. I have never married, I have always lived quietly, and, *apart from my interest in* dahlias, I have no hobby. For *those* reasons I found myself agreeably excited by my mother's funeral.

My father had been dead for more than forty years. He was a building contractor of a lethargic disposition who *used to take* afternoon naps in *all sorts of* curious places. This irritated *my mother*, who was an energetic woman, and she *used to seek him out to disturb* him.

Sammanställningen visar att texten innehåller få idiom men många kollokationer. Det är givet att åskilliga av de kollokationer som markerats i texten kan produceras av en talare utan tidigare kunskap om dem, liksom att de på samma sätt inte erbjuder några svårigheter för läsaren. Men andra är både mindre förutsägbara och mycket centrala: *for the first time in, found it difficult to, apart from, my interest in, all sorts of, seek him out*. Eftersom kollokationerna är baserade på förekomster i en relativt begränsad korpus (Brown) är resultatet inte heltäckande – en större korpus skulle säkert dessutom ha registrerat t.ex. *retired from, an adequate pension, agreeably excited, of a lethargic disposition* och *afternoon naps* som kollokationer. Sådana fasta förbindelser är nödvändiga för språkbrukaren att behärska aktivt. De är långt mera vanliga och centrala än de idiom som här representeras av *a silver handshake*.

Som en andra illustration av den centrala roll kollokationerna spelar vill jag hänvisa till en egenhet som de som är något bekanta med kommunikation på Internet säkert lagt märke till. I varje fall i engelsk-

språkiga sammanhang är det mycket vanligt att man använder förkortningar, i första hand kanske i tekniska diskussioner, men också ofta, och det är här det viktiga, i vardagligt icke-tekniskt tankeutbyte. Förkortningarna är inte de traditionella utan bygger på förekomsten av stående fraser i allmänspråket. En lista på några av de vanligare ser ut så här:

AFAIK	as far as I know
AKA	also known as
ASAP	as soon as possible
BTW	by the way
BRB	be right back
FAQ	frequently asked questions
FYI	for your information
FWIW	for what it's worth
FUBAR	f**d up beyond all recognition
IDK	I don't know
IMHO	in my humble opinion
IMO	in my opinion
IYKWIM	if you know what I mean
LOL	laughing out loud
OTOH	on the other hand
TIA	thanks in advance
TPTB	the powers that be
TTFN	ta ta for now
RTFM	read the f**n manual
WRT	with respect to
WYSIWYG	what you see is what you get
YWIA	you're welcome in advance

Fig. 4. Vanliga Internetförkortningar

Sådana förkortningar – och en hel del andra – förutsätts vara omedelbart begripliga och förklaras nästan aldrig. De används alltså i den trygga förvissningen att kollokationerna de bygger på är en lika självklar del av mottagarens lexikon som de enskilda orden.

11. Slutord

Jag ställde tidigare frågan: Tar korpusbaserade lexika hänsyn till idiomerna och kollokationerna i tillräcklig utsträckning? Svaret kan knappast uttryckas i absoluta termer, eftersom det finns stora skillnader mellan olika lexika. Men tendensen är ändå klar: idiomerna förefaller ganska väl tillgodosedda i förhållande till sin användbarhet, medan övriga kollokationer är styvmoderligt behandlade, återigen i förhållande till sin användbarhet. Slutsatsen för lexikografens del måste då bli denna:

Basera för all del lexikonerna på korpora, men bli inte alltför orolig om idiomerna är dåligt representerade. Det viktigaste är att (de övriga) kollokationerna är väl företrädna.

Referenser

- Allwood, Martin S. (i samarbete med Elsie Postlethwaite, Monica Nilsson och G.K. Laycock) 1965: *Engelska vokablar och idiom med viktigare amerikanska varianter*. 2. uppl. Stockholm Liber.
- Bright, William (ed) 1992: *International Encyclopedia of Linguistics*. New York & Oxford: Oxford University Press.
- Collins, V.H. 1958: *A Book of English Idioms*. London: Longmans.
- Curry, Dean 1985: *Idiomatic expressions in context : with exercises : advanced level / Dean Curry*. Washington, D.C. United States Information Agency.
- Gellerstam, Martin *et al* (eds) 1996: *Euralex '96. Proceedings*. I–II. Göteborg University, Department of Swedish.
- Gläser, R. 1988.: The grading of idiomaticity as a presupposition for a taxonomy of idioms. In W. Hülsen and R. Schutze (eds) *Understanding the Lexicon: Meaning, Sense and World Knowledge in Lexical Semantics*. Tübingen: Max Niemeyer.
- Halliday, M.A.K./Ruqaiya Hasan 1976: *Cohesion in English*. London: Longman.
- Hargevik, Stieg/Magnus Ljung 1989: *Say when och 1284 andra engelska idiom*. Esselte studium Stockholm.
- Hübinette, Lars och Bengt Odenstedt 1988: *Ord och inga visor: 2000 svenska idiom i engelsk översättning*. Lund Studentlitteratur.
- Kjellmer, Göran 1994: *A Dictionary of English Collocations*. I–III. Oxford: Clarendon.
- Laycock, G. Kenneth and Martin S. Allwood 1959: *Idiomatic English sentences with Swedish equivalents* 4. ed. Stockholm 1959
- Makkai, A. 1972: *Idiom Structure in English*. Janua Linguarum Series Major 48. The Hague: Mouton.
- Moon, Rosamund 1996: "Data, Description, and Idioms in Corpus Lexicography." I Gellerstam *et al.* 1996: 245–56.
- OED* = Simpson, J. A./Weiner, E. S. C. (eds.) 1989: *The Oxford English Dictionary*, 2nd ed. Oxford: Clarendon.
- Ordbog over det danske Sprog*. Grundlagt af Verner Dahlerup; udgivet af Det Danske Sprog- og Litteraturselskab; redigeret af H. Juul-Jensen ... København: Gyldendal 1919–1956.
- Orr, John 1962: *Old French and modern English idiom*. Oxford

- SAOB = *Ordbok över svenska språket utgiven av Svenska Akademien*.
1– . Lund 1898– .
- Schenk, André Yvon 1994: *Idioms and collocations in compositional grammars* = Idiomen en collocaties in compositionele grammatica's.
Utrecht.
- Sinclair, J. McH., Jones, S., & Daley R. 1970: *English Lexical Studies*.
Report to OSTI on Project C/LP08. Birmingham: Department of
English, The University.

Carl-Erik Lundblad

Värdering av SAOB:s korpus och språkprovssamling

The Swedish Academy Dictionary bases its description of the Swedish language from 1521 till today on a collection of about 7.5 million slips containing a full sentence illustrating the use of a word or an expression. These are excerpts made from about 20,000 sources. Texts that have been systematically excerpted, thus contributing about three words per page, number about 11,000.

The collected material gives a good coverage of written standard Swedish of various styles. The quotations have mainly been made by selected and properly instructed persons. This is one reason for the reasonably high quality and usefulness of the material. The enormous amount of concerted work carried out on word selection and lemmatization is another. Concordances based on mainly 20th-century texts are used to fill in the material and to check the analysis based on the slips. Concordances of older texts are not yet available in great enough number to be of any importance.

Svenska Akademiens ordbok baseras på en korpus i bemärkelsen: textsamling sammanställd för visst ändamål. T.ex. Aitken (1978:30) använder också termen "korpus" så, men skulle lika gärna ha kunnat tala om sin ordboks källor. Det är inte den numera mest gångbara användningen av "korpus". De stora historiska ordböckerna bygger i allmänhet ordbeskrivningen på en korpus i denna äldre bemärkelse och inte i den moderna: textsamling för visst ändamål, där enskilda element kan (datamaskinellt) undersökas i hela samlingen. I denna uppsats har "korpus" mest den äldre betydelsen. SAOB:s korpus, dvs källor, och den därav gjorda materialsamlingen ska granskas.

SAOB:s källor kan med viss rätt men något överdrivet, vilket framgår nedan, uppges till ett antal av ca 20.000 från år 1521 och framåt (se Svensson 1991:372). Ur källorna har det excerperats ca 7,5 miljoner språkprov (se Hast 1983:160). Är detta material en tillräcklig bas vid ordboksredigeringen, och kan det brukas av andra än ordboksredaktionen? Genom att hänvisa till de redan utgivna 31 banden av SAOB vore det lätt att konstatera att materialet fungerar åtminstone någorlunda när man skriver en historisk ordbok över svenskan. Eftersom excerpten, utskrivna på lappar med stickord, kontext och källangivelse, ofta anlitas av forskare utanför redaktionen vore det lätt

att ge ett positivt svar även på frågan om materialet kan komma till användning av andra. Men svaren blir ointressanta utan precisering.

Storlek och kvalitet

SAOB:s källor och material har utförligast beskrivits i Hast (1983: 159–198). Här ska endast erinras om vilka principer som gällt och gäller vid urval av källor och excerpt och därmed sätter sin prägel på ordbokens grundmaterial.

Korpusen har klar avgränsning bakåt i tiden. Ingen källa är äldre än från 1521. Åt andra hållet finns ingen gräns, utan vid redigerings-tillfället nyskriven litteratur kan användas. Härefter skiljer sig SAOB från den nederländska ordboken *Groot Woordenboek der Nederlandsche Taal* (WNT), som inte beaktar yngre källor än från 1921 (Sterkenburg 1984:86).

SAOB:s korpus är en expanderande sådan även i det avseendet att också äldre källor kan bli fler. Det kan alltså mycket väl tas språkprov t.ex. från en tidigare inte excerperad 1600-talskälla. Också i detta avviker SAOB numera från den nederländska ordboken, som har slutat att excerpera och alltså redan skulle kunna trycka en slutgiltig källförteckning.

Den grundläggande excerperingen gjordes under 1800-talets sista decennier, men excerperingen fortgår alltjämt. Att korpusen växer på detta sätt innebär naturligtvis att ord som hör till senare delen av alfabetet beskrivs med större material än andra. Tillväxten kan medföra den risk som ordbokschefen Esaias Tegnér d.y. hade i tankarna när han talade om att ordboken i stället för att närma sig slutet kom allt längre bort därifrån. Fullt så illa är det inte, men expansionen medför en ojämnhet i materialet och därmed en skillnad mellan de tidigare och senare banden. Allmänt sett är underlaget för beskrivningen magrare i de tidigare banden, och idag nya ord och ordanvändningar beskrivs bara i den mån de faller på den senare delen av alfabetet. Med tanke på ordbokens långa utgivningstid – efter drygt 100 år är 5/6 av ordförrådet avverkat – hade emellertid ett slutårstal för excerperingen varit ett sämre alternativ. Tack vare den kontinuerliga materialökningen har varje band skrivits med så fylligt material som möjligt. Senare betydelseutvecklingar beskrivs i sitt etymologiska sammanhang med äldre betydelser, och utvecklingen kan därmed förklaras tydligare. Det är inte otänkbart att sena utvecklingar ibland upptäcks eller att deras riktiga förklaring uppenbaras tack vare att sena språkprov behandlas till-

sammans med äldre. Att materialet är bättre i de senare delarna av alfabetet bör alltså ge ordboken ett extra värde.

Den kontinuerliga materialtillväxten har sannolikt inte spelat någon avgörande roll för ordbokens långsamma utgivning. Ordboken skulle ha kunnat skrivas betydligt snabbare trots att inget slutdatum för excerperingen har satts. Tidsutdräkten måste i första hand tillskrivas en självpåtagen och obefogad noggrannhetsiver och överdriven detalj-skräpa i redovisning av ordens ibland rent tillfälliga användningar. Dessutom beror förseningen på den ineffektivitet som med närmast naturlags förutsägbarhet drabbar vilken mänsklig organisation som helst där man nonchalerar riskerna och inte vidtar sådana motåtgärder som i konkurrensutsatta projekt lättare framstår som nödvändiga.

Korpusens täckning av svenskt skriftspråk

Man har försökt att vid excerperingen täcka svenskt skriftspråk i olika genrer och ämnesområden. Svensson (1991:373) har kompletterat Hast (1983: 174) och gett en översikt av källornas fördelning på genrer. Svensson (1991:374) har vidare meddelat intryck i fråga om hur SAOB:s källor representerar svensk bokproduktion. Praktiskt taget allt 1500-talstryck har excerperats och det mesta av vikt från 1600-talet. Av den betydligt mer omfattande 1700-talslitteraturen har bara en del excerperats men med god spridning över genrer. Hur väl 1800- och 1900-talets svenska är representerad är svårt att ha en mening om, eftersom skrifterna från den perioden är många och därför inte har blivit lika bra översiktligt beskrivna som tidigare perioders vad avser relativ storlek och fördelning på genrer. Viss tillförsikt ges dock av den uppräknig i Svensson (1991: 376) av genrer eller ämnesområden utöver skönlitteratur som är väl representerade i SAOB:s material: biografi, botanik, ekonomi, etnografi, författning, förordning, geografi, historia, juridik, litteraturvetenskap, ordbok, periodica, samhällsvetenskap, sport, språkvetenskap, teknologi, teologi, tidningar, tidskrifter och undervisning.

Att exakt och med siffror ange hur väl svenskt skriftspråk är representerat i SAOB:s material är svårt dels av det skäl som redan nämnts, nämligen att vi inte har tillräcklig kunskap om de senaste seklernas svenska skriftspråk, dels därför att SAOB-källorna ofta har använts för bara något enstaka excerpt.

Antalet källor

SAOB:s källförteckningar upptar 20.000 titlar. Siffran är imponerande och kan göras något högre. Man kan nämligen peka på det förhållandet att vissa källor omfattar flera volymer men har räknats såsom en enda, t.ex. Gustav I:s registratur, Publique Handlingar, Oxenstiernas skrifter och brevväxling. Men som mått på omfånget av SAOB:s korpus minskar siffran betydligt när man betänker att i källförteckningens supplement I resp. II alla citerade källor har medtagits, även om bara ett enda belägg skulle ha hämtats därur. Sådan komplettering med enstaka citat vid redigeringen är inte ovanlig. Enligt Hast (1983:173) torde bara 20 procent av källorna i Supplement I ha excerperats, dvs finnas representerade med åtminstone ett belägg per sida. Jfr Svensson (1991:372). Enligt stickprov som jag gjort i det senare supplementet (II) är det lika få källor där som har excerperats verket igenom. Fortsätter man beräkningarna i Hast (1983:174) och inkluderar källförteckningens supplement II torde antalet helt igenom excerperade källor inte uppgå till mer än ca 11.600.

Siffran 11.600 kan nog även den vara i överkant, eftersom ingen modifiering av antalet verk i den grundläggande källförteckningen har gjorts. Visserligen är det allra mesta där regelrätt excerperat, men undantag finns i form av källor som bara har citerats någon enstaka gång. Förteckningen upptar även viss språkvetenskaplig referenslitteratur som inte alls har excerperats, t.ex. Axel Koch: Umlaut und Brechung im Altschwedischen. Men sådana verk är inte särskilt många.

Språkprovssamlingen

Den språkprovssamling på uppskattningsvis 7,5 miljoner lappar (ett belägg i kontext på varje) som nu är SAOB:s material har växt fram under ett drygt sekel. Under redaktionens första decennier, närmare bestämt från 1884 till 1903, var tillväxten som intensivast. Den avtog markant från 1930-talet. Från 1966 till 1979 har den genomsnittliga årliga tillväxten varit knappt 15.000 excerpt. Se Hast (1983:169, 172). En överslagsberäkning som jag gjort visar att tillväxten efter 1979 är ännu lägre.

Ur språkprovssamlingen hämtas de språkprov som redaktörerna går igenom och beaktar och efter utförd analys väljer exempel ur för att illustrera betydelsebeskrivningen och belägga uppgifter om böjning och stavning. Alla belägg anges med noggrann uppgift om källa.

Ordböcker som likt SAOB baserar sin beskrivning på en excerpt-samling torde vara de enda som kan uppge en minimisumma för antalet

beaktade språkprov. Men det används alltid ännu fler. De språkprov som excerptister utanför redaktionen bidrar med kompletteras nämligen, främst med material från andra lexikon, samlingen till Östergrens Nusvenska ordbok, ordförteckningar från Tekniska nomenklaturcentralen, excerpt som ordboken får från Svenska språknämnden samt ord hämtade från floror, riksdagsprotokoll och några andra verk med alfabetiska ordförteckningar. Allt ger den stomme med vars hjälp redigeringen görs. Därtill kompletteras vid behov med datorlagrade konkordanser på Institutionen för svenska språket vid Göteborgs universitet och en finlandssvensk textkorpus (FISC) på Institutionen för nordiska språk vid Helsingfors universitet. Den excerptering som anlätade och per excerpt avlönade personer utanför redaktionen utför och har utfört ger emellertid den viktigaste delen av ordbokens material.

Excerpteringen

Materialet som ordbeskrivningen ska grundas på är redan från första stund valt med tanke på vad det ska användas till. Hur pass väl man lyckats uppnå representativitet i urvalet av källor är som nämnts ovan svårt att ha en bestämd uppfattning om, men mängden representerade genrer ger ett gott intryck och enligt Hast (1983:187) skulle ca 12.000 verk vara ett stort urval i jämförelse med andra historiska ordböcker.

Excerpteringen innebär som regel att 3 ord per sida har valts. Detta val har ibland lite slarvigt och felaktigt beskrivits i termer av slumpmässighet (jfr Allén 1979:128 och Ekbo & Loman 1965:84). I stället för att vara slumpmässig innebär excerpteringen den första av många absolut nödvändiga selektioner som måste göras efter källurvalet.

De av redaktionen anlätade excerptisterna är språkintresserade, läsvana personer, ibland och i synnerhet under redaktionens första årtionden språkvetenskapliga experter. Förr instruerades excerptisterna mycket noggrant och resultatet av deras arbete kontrollerades ingående (jfr Hast 1983:164ff, där sådana anvisningar återges i en bilaga). Numera inskränks anvisningarna till två skrivna A4-sidor och en kortare muntlig genomgång.

Excerptisternas allmänna uppgift är att välja sådana språkprov som ordboksredaktionen kan tänkas ha nytta av. Variation kan förekomma efter typ av litteratur. Excerptering av facklitteratur t.ex. är ofta av det slaget att man vill ha belägg på typiska, ämnesspecifika men ändå i allmänspråket påträffade ord och uttryck. Sådan excerptering görs inte sällan av sakkunniga inom det ämnesområde som boken behandlar.

Facksakkunniga excerptister har uppmanats att på lappen om nödvändigt och möjligt ge en kort förklaring av ordets betydelse om den skulle vara speciell och inte direkt begriplig. I övrigt ska alla språkprov alltid innehålla så mycket kontext att ordets konstruktion och användning blir belyst. Sådana språkprov som ger en definition eller förklaring eller på annat sätt är ägnade att kasta ljus över betydelsen är redaktionen tacksam för.

Excerptister med uppgift att gå igenom äldre källor är ofta vana läsare av historiska texter. De bör ha god förmåga både att välja ord och uttryck som avviker från senare eller tidigare språkbruk och att välja det idiomatiska och typiska.

De allmänna anvisningar som excerptister av moderna texter numera får handlar om att finna belägg på nya ord eller nya användningar av ord och uttryck. Excerptisterna uppmanas att undvika dels tillfälliga och självklara sammansättningar, dels uppenbara felaktigheter. De uppmanas också att undvika mycket vanliga och till betydelsen helt klara ord såsom *dörr* och *näsa*, om de inte ingår i speciella uttryck (*ha skinn på näsan*). I äldre, mer detaljerade anvisningar uppmanades emellertid excerptisten att ta med ett eller annat belysande språkprov med sådant vanligt ord. Se vidare Hast (1983:187 ff).

En selektiv, traditionell excerptering har sina nackdelar, men fördelarna torde överväga. Benägenheten att förbigå det typiska och lägga vikt vid det speciella kan ge för stor uppmärksamhet åt perifera ordanvändningar och en beklaglig försummelse av de vanligaste orden. Såväl högfrekvent, vardagsspråklig användning av betydelseord som typiskt bruk av formord riskerar att förbigås. Men medvetenheten därom och anvisningarna till excerptisterna liksom övervakningen av deras arbete under den intensivaste excerpteringsperioden motsvarar stränga krav på excerptering (jfr Reichmann 1990:1603 ff) och kan förmodas ha gett SAOB ett förhållandevis gott grundmaterial. För modernt språk har redaktionen numera god möjlighet att fylla luckor i det excerpterade materialet genom att anlita datorlagrade konkordanser när misstanke har väckts eller som rutinkontroll. För äldre språk erbjuds ännu inte samma möjlighet. Alltför få äldre texter från den av SAOB beskrivna perioden föreligger nämligen i datorlagrad form.

Det är en stor fördel vid redigering av en historisk ordbok att grundmaterialet samlas genom traditionell excerptering, eftersom excerptisterna på sitt stadium kan delta i den nödvändiga texttolkningen. De kan nämligen när de utför arbetet ta hänsyn till kontexten, tillkomsttiden, författaren, ev. översättning, genren, ordens frekvens, lexikaliseringsgrad, fraseologi, betydelse, speciell användning eller idiomatik. De kan vidare avgöra ordets grundform, normalisera stavningen av

stickordet och ange ordklass. Visserligen är det möjligt att använda dator både för lemmatisering och för sådan excerpering som modifieras t.ex. med hänsyn till högfrekventa ord, varigenom risken att drunkna i en uppsjö av de allra vanligaste orden elimineras, men de texter som en historisk ordbok med stor korpus har att beakta bjuder oftast så rik och oförutsedd variation i ordens böjning och stavning att automatiken blir omöjlig. Det arbete som därför måste utföras efter en automatisk excerpering blir tidsödande, dyrt och ineffektivt i jämförelse med arbetet under och efter en traditionell excerpering. En excerpist, som t. ex. kan beakta såväl den omedelbara kontexten som periodens språk i allmänhet, kan utföra en väsentlig begränsning snabbare och smidigare i samband med excerperingen än vad redaktören kan göra vid en genomgång av ett ordmaterial som har plockats maskinellt från många källor och flera århundraden. När en manuell selektiv excerpering ombesörjs av så många excerpister som fallet är i SAOB (enligt Hast 1983:161 har 300–400 hittills varit i verksamhet) minimeras dessutom risken för att individuell preferens präglar materialet.

Den manuell-cerebrala excerperingen är den enda som kan användas för handskrivna källor. SAOB har haft många skickliga handskriftsläsare som excerpister, främst Ludvig Larsson (se Hast 1983: 186).

Ett material måste begränsas på något sätt för att effektivt kunna utnyttjas vid ordboksredigeringen. Den per år bearbetningsbara delen av en ordbokskorpus är enligt Aitken (1978:31) ca 15.000 språkprov, om arbetet ska bli grundligt gjort. Med hänsyn till avbrott och tillfälliga förseningar räknar Aitken med i runda tal 10.000 språkprov per år. SAOB har 11.000 språkprov som årspensum för redaktörerna, den nederländska ordboken (WNT) ca 8.000 (jfr Sterkenburg 1984: 77).¹

Komplettering av excerperat material

Som nämnts ovan använder redaktionen datorbaserade konkordanser av 1900-talstexter, mest romaner och tidningar, men även t.ex. några Strindbergstexter från 1800-talet. Som komplement till det excerperade materialet är dessa konkordanser av stort värde, t.ex. när man vet eller misstänker att en ordanvändning förekommer fastän den saknas i excerptmaterialet. På så sätt kan man ofta lätt hitta ett belägg på t.ex. *en*

¹ Skillnaden mellan ordböckerna säger förmodligen ingenting om kraven på redaktörerna eftersom det t.ex. kan variera hur mycket materialet behöver kompletteras, kontrolleras i källan och hur noggrant beskrivningarna utförs.

säker etta i fråga om fotbollstips o.d. Konkordanserna är också smidiga när det gäller att få bekräftelse på eller stöd för en på excerperat material utförd betydelseanalys. Det är emellertid anslående att konkordanser över moderna texter förhållandevis sällan tillför det traditionellt excerperade SAOB-materialet något överraskande nytt och praktiskt taget aldrig uppvisar språkprov som tvingar till nya eller andra betydelsemoment än dem som baseras på språkprovslapparna. Det kan naturligtvis bero på redaktörens motvilja att ändra, men sannolikt är det traditionella materialet så grundligt att nytillkommet i stort sett bara kan bidra med tydligare exempel och sena belägg.

Det hade naturligtvis varit av stort värde med tillgång till även äldre texter i datorbaserade konkordanser, i synnerhet om de är lika lätthanterliga som eller ännu lätthanterligare än de som Institutionen för svenska språket i Göteborg tillhandahåller. De äldre texter som redan finns inlästa och sålunda kan fås i konkordansform är på grund av ringa antal och omfattning närmast försumbara vid ordboksredigeringen.

Lexikografiska institutionen vid Islands universitet har rapporterat om användning av en datorläsbar textbank för att fylla ut luckor i ett excerperat material (Pink et al. 1993:1–18). Något liknande vore önskvärt för SAOB. Ett antal av ordbokens excerperade källor från varje århundrade borde vara tillgängliga i datorläsbar form och som konkordanser. Förslagsvis borde 1 miljon ord från 1500-, 1600- resp. 1700-talet läsas eller skrivas in. Då skulle framförallt det excerperade materialets representativitet kunna kontrolleras, betydelseanalysen kunna bekräftas eller fördjupas och språkprovssamlingen kunna kompletteras med äldre eller bättre belägg. Om SAOB-projektet skulle ha börjat idag med förhållandevis lika mycket resurser som har använts hittills borde alla källor innan excerperingen sattes igång ha gjorts tillgängliga i en korpus i den moderna bemärkelsen. Att få denna korpus lemmatiserad hade naturligtvis gjort den oerhört värdefull. Det arbete som därvid krävs är egentligen detsamma som föregår normal ordboksredigering. Tillgång till konkordanser över ordförrådet i alla källor hade inneburit värdefulla möjligheter liksom detta att få uppgift om absoluta frekvenser och förekomst i vissa genrer o.d. samt att lätt få fram idiom och kollokationer genom att ordna materialet efter ordens högfrekventa samförekomster. Dessutom skulle hapax-ord och lexikonarv lättare kunna avslöjas.

Materialets tillgänglighet

Så länge en ordboks språkprovssamling inte finns lagrad som datortext i sökbar form är förmodligen de traditionella lapparna med ett språkprov på varje den bästa lagringsformen och kanske är de också alltid att föredra som arbetsmaterial. De är behagliga att läsa och lätta att placera i grupper och ordning allt efter vad den språkliga analysen bjuder. Naturligtvis hade emellertid en inläsning för datorlagring av dem som nu finns inte bara inneburit en säkerhet mot brand eller annan skada utan också gett ökade sök- och bearbetningsmöjligheter. Om all text i språkproven hade varit inläst hade det varit allra bäst, men tidsåtgången för att skriva av alla de handskrivna lapparna skulle vara enorm även med SAOB-mått. Enbart en inskanning av språkproven som bild med komplettering på så sätt att stickordet och källans utgivningsår skrevs in och blev sökbart skulle kosta mycket tid och pengar. Säkerhetsvinsten vore emellertid fortfarande stor och även användbarheten, trots begränsningar på grund av att man inte kunde söka i textcitaten. Dessutom skulle man uppnå att materialet blev mer tillgängligt för forskare utanför ordboksredaktinen.

Förväntningar på språkprovssamling och ordbok

Alltsomoftast ställs orimliga krav eller förväntningar på SAOB:s språkprovssamling, även från ordbokens egna led. T.ex. ville en av dess första redaktörer, den junggrammatiskt skolade Gustaf Cederschiöld, att ordboken skulle beskriva orden i alla dess tänkbara sammanhang. Sålunda skulle ett ord inte bara beskrivas med avseende på dess stilvärde och varierande bruk i olika samhällsgrupper (något som ordboken naturligtvis försöker informera om ifall det är möjligt) utan orden skulle helst också beskrivas i fråga om variation med andra ord. Även sådan information ger ordboken, men i blygsam utsträckning. Över huvud taget ger ordboken förhållandevis sällan direkt uppgift om synonymi- och antonymirelationer eller annan opposition och (hierarkisk) ordning inom ordförrådet. Cederschiöld ville emellertid ha upplysningar om t. ex. alla synonymer som ordet ersatts av eller ersatt, när det skett och hur förhållandet mellan synonymer varit under olika tider (se Cederschiöld 1894). Sådant kräver ingående undersökningar som inte utan besvär kan utföras ens med hjälp av en fullbordad SAOB. Då behövs en väldefinierad korpus med framför allt i fråga om genre representativa texter. Dessutom borde SAOB inte bara vara klar utan så beskaffad att ordens betydelsebeskrivning enkelt kan knytas till förekomst i korpus. Korpusen skulle alltså vara helt betydelseanalyserad. Det torde vara oerhört svårt att göra en så noggrann ordbeskrivning som Cederschiöld efterlyste, och den blir inte mycket lättare att utföra med hjälp av en klar SAOB.

Ordboken beskriver orden vart och ett. Den kan knappast läsas på annat sätt än så som den är skriven, ord för ord. Därmed inte sagt att ordboken med sitt material inte skulle vara en värdefull om inte oumbärlig källa vid överlexematiska studier av svenskans ordförråd. Men om man t.ex. vill ta reda på hur *snille*, *geni* och *begåvning* förhåller sig till varandra eller jämföra användningen av *civilisation*, *samhälle*, *stat*, *rike* och *nation*, så krävs eget forskningsarbete.

Den historiska ordbokens uppgift är att beskriva ordens historia, framför allt deras betydelse och betydelseutveckling. Det är inte ordbokens uppgift att svara på frågor om hela eller ens delar av ordförrådet. Vid sådan forskning är en historisk ordbok visserligen en hjälp, men den utgör en begränsad faktabank med upplysningar om orden var för sig och bara lite om deras förhållande till varandra. Det är i stort sett bara indirekt genom de distinktiva definitionerna som SAOB ger information om ordens (lexemens) betydelseförhållande till ett överordnat begrepp och talar om vilka specifika särdrag som utmärker orden och därmed skiljer dem från andra.

En av SAOB:s brister är dess långa tillkomsttid, som medför att t.ex. upplysning om ordens bruklighet måste tas emot med beaktande av när den aktuella ordboksartikeln skrevs. Ett "numera föga br." kan vara ett omdöme om ordet i en av dess betydelser för 100 år sedan. Nu kanske ordet inte längre ingår i vårt ordförråd och borde ha dödförklarats. Men det kunde ha varit ännu sämre i detta avseende, nämligen om ordboken försetts med ett eller flera integrerade supplement på samma sätt som t.ex. OED (Oxford English Dictionary). Där kan t.ex. bruklighetsbeteckningen "now rare" referera antingen till tidpunkten för redigeringen av det ursprungliga verket eller till endera av de båda supplementens tillkomsttid. Se Duekilde (1995).

Möjligheter som ordboken och dess samlade material kan erbjuda upptäcks ibland under inflytande av sådan entusiasm som nya hjälpmedel väcker. Den datorlagrade bibliografi över SAOB:s källor som beskrivs i Svensson (1993:376 f) är mycket värdefull, framför allt som lättanvänd källförteckning. Om man dessutom kan göra undersökningar på tvärs i ordboken med dess hjälp återstår att se. Det beror bl.a. på hur representativt och vältäckande inte bara excerperat material utan även det därav valda och i ordboken citerade materialet är för en bestämd tids ordförråd. Är det möjligt att t.ex. via bibliografin spåra texter från ett visst fackområde, t.ex. juridik, och sedan söka källorna i databasen OSA (en inläsning av SAOB:s tryckta delar) för att gå vidare till uppslagsorden och därefter upprätta en förteckning över ordförrådet i juridiskt fackspråk från t.ex. 1700-talet (jfr Svensson 1991:377)? Sannolikt är just det inte möjligt, bland annat eftersom andra ord än

fackord kan ha excerperats ur fackspråkliga källor. Vi vet inte heller hur mycket av det totala (fack)ordförrådet som uppmärksammas i ordboken, men det är i alla fall bara ett urval, om än stort. Dessutom återstår det att fullborda kodningen och göra OSA ännu mer användbar än nu. Fulltextsökning är än så länge visserligen inte den enda men den viktigaste funktionen. Men prova gärna! Det väcks ofta nya förväntningar ur misslyckade försök och alla kommer inte på skam.

Varken ordboken eller dess språkprovssamling kan alltså ge direkta svar på frågor om ordförrådet i dess helhet. Även om språkprovssamlingen hade förelegat i lättillgänglig form skulle den ha svikit. Materialinsamlingen har gjorts med tanke på att ordboken ska beskriva enskilda ord, och excerperingen är oerhört gles; antalet språkprov är väldigt litet i förhållande till det enorma totala ordantal som korpusen omfattar. Vi erbjuds ingen säker information om ordens förhållande till andra ord. Tyvärr räcker alltså varken SAOB:s materialsamling eller den skrivna ordboken till om man vill ha direkta upplysningar om relativa ordfrekvenser, ordförrådets växlingar under årtionden eller sekler, växlingar i olika genrer o.d. Det kan bara ges antydningar om sådant. Däremot tillhandahåller ordboken mycket upplysning om enskilda ords betydelse och användning under olika tider och med olika stilvärde. Ordbokens korpus och materialsamling är en nödvändig och bra grund att stå på vid redigeringen av denna historiska faktabank. Men det är rättmätigt att önska mer, och för den uppfinningsrike och tålmodige finns alltid mer information att få. Som hjälp står språkprovssamlingarna till allmänt förfogande. Än så länge krävs det emellertid att man tar sig till redaktionen i Lund.

Traditionell språkprovssamling kontra modern lexikografisk korpus

Det har ovan pekats på några fördelar hos den traditionella typ av excerptsamling som utgör det förnämsta materialet för de historiska ordböckerna. Det har förhoppningsvis framgått att datorbearbetad och bearbetningsbar korpus är till stor fördel som komplettering av en excerptsamling. Att få till stånd en korpus i den moderna bemärkelsen med vilken man som enda hjälp skulle kunna skriva en historisk ordbok torde emellertid vara oerhört svårt och tidskrävande. Eiríkur Rögnvaldsson (1995) presenterar en värdefull konkordans över de isländska sagorna och nämner fördelar med korpus till skillnad från excerptsamling vid redigering av ordbok över sagalitteraturen. Huruvida hans bedömningar i fråga om en sagaordbok är korrekta får andra

avgöra. Men det som gäller de isländska sagorna behöver naturligtvis inte vara direkt allmänt tillämbart. Det ska t.ex. betänkas att sagalitteraturen är en tämligen enhetlig genre med litet ordförråd jämfört med t.ex. den skrivna svenskan från de senaste seklerna. I nysvenskan är såväl orden, böjningarna, formerna som betydelseerna många fler. Dessutom ska man ha i minne att de historiska ordboksprojekten gör en etymologiskt baserad betydelsebeskrivning, vilket en periodordbok över sagalitteraturen också kan välja att göra. För den uppgiften är emellertid frekvensundersökningar inte fundamentala, varför t.ex. en beskrivning av *heimsókn* hos Fritzner (1954) inte behöver vara mindre lyckad av det skäl allena att en korpusundersökning eller flera skulle visa att den betydelse som Fritzner angett först som nummer 3 är den vanligaste (jfr Eiríkur Rögnvaldsson 1995:130f).

Den höga status som korpusbaserad lexikografi numera har kan leda till ett ointresse eller i vissa fall förakt för områden där den inte är fullt eller ensamt tillämbart, t.ex. historisk lexikografi. De historiska ordboksprojekten förtjänar emellertid att bedömas för vad de faktiskt åstadkommer, inte förutfattat för det sätt på vilket de har samlat och organiserat sitt material. Det vore värt en ingående studie att få veta hur de historiska ordböckerna bäst når målet att göra en etymologisk betydelsebeskrivning. Det är också intressant att få de korpusbaserade ordböckerna noggrant granskade med tanke på vilken effekt korpushantering har, framför allt på betydelseanalysen och -beskrivningen. Beskrivs orden verkligen med beaktande av betydelseernas frekvens eller görs andra överväganden? Utnyttjas möjligheten till automatiska, statistiska undersökningar av ordens samförekomst för att få exempel på vanliga kollokationer (*ha ont i magen, kulmage, ha mage att, mat i magen, ligga på magen; borsta tänderna, borsta bort dammet, borsta håret*) eller används detta för att spåra grundläggande betydelser? Det senare vore en tvivelaktig metod. En kritisk granskning av såväl historisk som nuspråklig lexikografi efterlyses alltså, men tyvärr finns för närvarande ett ointresse för historisk, etymologiskt baserad lexikografi.

Litteratur

- Aitken, A. J. 1980. On Some Deficiencies in our Scottish Dictionaries. I: W. Pijnenburg & F. de Tollenaere (red.): *Proceedings of the Second International Round Table Conference on Historical Lexicography*, 33–56. Foris Publications, Dordrecht.

- Aitken, A.J. 1978. Historical Dictionaries, Word Frequency, Distributions and the Computer. I: *Cahiers de lexicologie. Revue internationale de lexicologie et de lexicographie*. Vol. XXXII nr 1, 28–47. Besançon.
- Allén, Sture 1979. Hur djupt går SAOB? I: *Nysvenska studier* 59, 119–129.
- Cederschiöld, Gustaf 1894. *Om de senast framställda fordringarna på en historisk ordbok*. Göteborg. Omtryckt i: *Om ordlekar och andra uppsatser i språkliga och historiska ämnen*, 83–120. Stockholm 1910.
- Duekilde, Anne 1995. Hovedværk og supplement. I: *Nordiske studier i leksikografi 3. Rapport fra Konferanse om leksikografi i Norden 7.–10. juni 1995. Skrifter utgitt av Nordisk forening for leksikografi. Skrift nr 3*, sid 83–92. Reykjavík.
- Eiríkur Rögnvaldsson 1995. A Concordance to Old Icelandic Texts and its Lexicographic Value. I: *Nordiske studier i leksikografi 3. Rapport fra Konferanse om leksikografi i Norden 7.–10. juni 1995. Skrifter utgitt av Nordisk forening for leksikografi. Skrift nr 3*, 123–135. Reykjavík.
- Ekbo, Sven & Loman, Bengt 1965. *Vägledning till Svenska Akademiens ordbok*. Skrifter utgivna av Nämnden för svensk språkvård 2. Läromedelsförlagen, Stockholm.
- Fritzner, Johan 1954. *Ordbog over Det gamle norske Sprog*. Nyt uforandret opptrykk av 2. utgave (1883–1896) med et bind tillegg og rettelser redigert av Didrik Arup Seip og Trygve Knudsen. Oslo: Tryggve Juul Møller.
- Hast, Sture 1983. Om SAOB:s material. I: *Nysvenska studier* 63, 159–192.
- Ordbok över svenska språket, utgiven av Svenska Akademien. Källförteckning*. 1939. Lund.
- Supplement I. 1975. Lund.
 - Supplement II. 1990. Opublicerad. Svenska Akademiens ordboksredaktion. Lund.
- Pind, Jörgen & Bjarnadóttir, Kristín & Jónsson, Jón Hilmar & Kvaran, Gudrun & Magnússon, Fridrik & Svavarsdóttir, Ásta 1993. Using a Computer Corpus to Supplement a Citation Collection for a Historical Dictionary. I: *International Journal of Lexicography*. Vol. 6 Nr 1. Oxford.
- Reichmann, Oskar 1990. Formen und Probleme der Datenerhebung I. Synchronische und diachronische Wörterbücher. I: *Wörterbücher*.

- Ein Internationales Handbuch zur Lexikographie.* Utgiven av Franz Josef Hausmann, Oskar Reichmann, Herbert Ernst Wiegand & Ladislav Zgusta. (Handbücher zur Sprach- und Kommunikationswissenschaft. Band 5.2) Walter de Gruyter. Berlin, New York.
- Sterkenburg, P. G. J. van 1984. *Van woordenlijst tot woordenboek. Inleiding tot de geschiedenis van woordenboeken van het Nederlands.* E. J. Brill, Leiden.
- Svensson, Lars 1991. Om SAOB:s material och en bibliografisk databas. I: *Nordiske studier i leksikografi I. Rapport fra Konference om leksikografi i Norden 28.–31. mai 1991. Skrifter udgivet af Nordisk Forening for Leksikografi. Skrift nr. 2.* Oslo.

Förkortningar

- DOST = *Dictionary of the Older Scottish Tongue from the Twelfth Century to the End of the Seventeenth.* 1931–. Chicago.
- FISC = Finland Swedish Corpus. (Finlandssvensk textkorpus. Se <http://www.ling.helsinki.fi/dept/nord/pro.html>)
- OED = *Oxford English Dictionary.* Second edition. 1989. Clarendon Press, Oxford.
- OSA = Om svar anhålles. (SAOB som databank. Se <http://logos.svenska.gu.se/saob/osa.html>)
- SAOB = *Ordbok över svenska språket, utgiven av Svenska Akademien.* (Svenska Akademiens ordbok). 1898–. Lund.
- WNT = *Groot Woordenboek der Nederlandsche taal.* 1882–. 's Gravenhaage och Leiden.

Nina Martola

Finlandssvenska, finsk-svensk tvåspråkslexikografi och korpusar

There is a need for a reasonably large corpus as a tool for the research into the Swedish language as it is used in Finland and as a basis of dictionary compiling. Actually, a corpus of the Swedish used in Finland already exists. It was created by the Department of Scandinavian Languages and Literature of the University of Helsinki and it consists of different types of texts from the early nineties, in other words quite modern language. However, the corpus is far too small, only about two and a half million words, which means that it needs to be enlarged. Such a project will be carried out by the Swedish Department of the Research Institute for the Languages of Finland, if possible in collaboration with the Department of Scandinavian languages and literature of the University of Helsinki. Unfortunately the resources for the project are quite limited, but as it is possible to use partly automatic tagging, an aim of twenty million words is not unrealistic. As for the types of texts, the outlines for the European PAROLE-project will be followed. Other text types will be included in a later phase.

Another corpus project to be carried out is the creation of a parallel corpus with the same texts in Swedish (preferably Standard Swedish) and Finnish. It should consist of Finnish texts with translations into Swedish, Swedish texts with translations into Finnish and, finally, translations from other languages into both Finnish and Swedish. This corpus will be used for finding Swedish correspondences of certain types of Finnish words and constructions that are problematic for the lexicographer.

Bakgrund

Under åren 1990–95 utarbetades manuskriptet till en ny finsk-svensk storordbok som ett samarbetsprojekt mellan Forskningscentralen för de inhemska språken och förlaget WSOY. Boken utkommer enligt planerna hösten 1996. När arbetet inleddes hade redaktionen inte tillgång till en enda korpus. I slutskedet kunde tre olika korpussamlingar utnyttjas; Språkbanken vid Göteborgs universitet och de finska och de finlandssvenska korpusarna vid Helsingfors universitet. En avsevärd utveckling på tre år, alltså. Men trots det kan man knappast kalla arbetet med finsk-svensk ordbok **korpusbaserad lexikografi**, snarare då **korpusstödd lexikografi**.

De närmaste åren kommer forskningscentralen att satsa mycket på korpusarbete och i det följande presenteras denna satsning. En betydligt

mera omfattande finsk korpus än den nu existerande kommer att sammanställas i samarbete mellan Helsingfors universitet och Forskningscentralen för de inhemska språken.

På svenska avdelningen vid Forskningscentralen för de inhemska språken kommer det att satsas dels på en utvidgad finlandssvensk korpus, dels på en parallellkorpus med samma texter på finska och svenska. Det är de två senare projekten som i första hand kommer att behandlas i denna artikel. Det finska korpusprojektet presenteras dock i korthet, eftersom tanken är att samma program och samma kodnings- och söksystem skall utnyttjas i arbetet med den finlandssvenska korpusen.

En nufinsk korpus

För det finska korpusprojektet har projektmedel redan beviljats, och planerna är längre framskridna än för den utvidgade korpusen över finlandssvenska. Forskningscentralen har inlett samarbete med Institutionen för allmän språkvetenskap vid Helsingfors universitet och man kommer att delta i det europeiska PAROLE-projektet. Den nya finska korpusen kommer i första fasen att bestå av 20 miljoner ord och att den kodas i enlighet med TEI-normen enligt samma kriterier som i de övriga deltagarländerna. PAROLE-korpusarna skall enligt överenskommelse bestå av (Le Parole. Technical Annex 1995:25)

58–72% tidningstext
16–22% boktext
4–10% tidskriftsmaterial
8–12% diverse

Som bok räknas allt som har ett ISBN-nummer. Åtminstone i den finska korpusen kommer dock skönlitterär text och icke-fiktion att hållas isär. Småningom kommer korpusen förmodligen att utvidgas till 30–50 miljoner ord. Den skall sammanställas av hela texter av så varierande slag som möjligt. För att ge en heltäckande bild av språket borde korpusen också innehålla texter som inte har underkastats språkgranskning. I något skede vore det viktigt att få med även talat språk.

Den finska korpusen skall kunna utnyttjas för flera olika syften, men de mest omfattande textmassorna behövs för uppdateringen av Suomen kielen perussanakirja ('Finsk basordbok') och för andra nuspråkliga lexikografiska projekt. Korpusarna kommer också att utnyttjas bl.a. av

språkvården, det nufinska grammatikprojektet och av textforskare. (Lehtinen et. al. 1995:29ff.)

Materialet måste antingen kodas morfologiskt i så hög grad som resurserna medger eller också måste man ha tillgång till sökprogram som tillåter sökning på ordstam – detta för att kunna hantera finskans rika böjningssystem (Lehtinen et. al. 1995:69 f.)

Finlandssvenska

Det existerar såsom nämndes i inledningen redan en finlandssvensk korpussamling, men i sitt nuvarande omfång är den alldeles för liten för att ge tillfredsställande sökresultat. Så här ser det ut:

FINLAND SWEDISH CORPUS (FISC)

Department of Scandinavian Languages and Literature, University of Helsinki

# for	Extent (tokens)	Description
1. Daily newspapers (from 1991)	895.000	Hufvudstadsbladet (20 iss.) Vasabladet (7 issues)
2. Literary prose (from 1990–92)	665.000	12 novels or collections of short stories.
3. Non-fiction (from 1990–93)	455.000	Various texts of ordinary factual prose
4. (Texts by) Authorities (from 1990–94)	440.000	Legal texts from the Statute Book of Finland. Various texts produced by public services etc.
5. Spoken language	80.000	Transcriptions of speech
6. Monitor	undefined	Miscellaneous texts

Grupp 5 innehåller utskrifter av sex radiointervjuer och fyra radiodiskussioner. I grupp 6 ingår bl.a. ytterligare tidningstext; Hufvudstadsbladet ca 120.000 ord.

Ingen av de enskilda korpusarna innehåller alltså ens en miljon löpord, vilket betyder att man får alldeles för få belägg också på relativt vanliga ord och uttryck för att kunna dra några egentliga slutsatser. Som ett exempel kan nämnas att en sökning på uttrycket *av misstag* i Språkdatas DN-konkordans på 4,1 miljoner löpord gav 33 belägg medan en sökning på *av misstag* och *i misstag* i det finlandssvenska pressmaterialet gav sammanlagt sju belägg; fem där prepositionen var *av* och två där den var *i*. (Uttrycket *i misstag* är en s.k. finlandism.)

En finlandssvensk baskorpus

Den finlandssvenska korpusen måste alltså utvidgas för att ge mera tillfredsställande sökresultat. Tanken är att Svenska avdelningen vid Forskningscentralen för de inhemska språken skall försöka skaffa projekt pengar och att ett samarbete skall inledas med Institutionen för Nordiska språk vid Helsingfors universitet, som hittills har upprätthållit korpusen.

Det första steget blir att helt enkelt bygga ut den existerande korpusen genom att fylla på med mera texter i varje typ, d.v.s.:

- 1) Mer pressmaterial; dels fler nummer av de nu ingående tidningarna, dels fler olika regionala tidningar. Att pressmaterialet utgör en stor andel är befogat. Dels utgör ju tidningar en mycket stor andel av de texter som trycks per år i ett land, dels hör tidningstexter till de texter som läses allra mest.
- 2) Fler romaner
- 3) Mer faktatext av olika slag

Det första steget blir att sammanställa en traditionell korpus enligt samma principer som den finska systerkorpusen, alltså enligt PAROLE-projektet (se ovan). Den existerande korpusen är i ganska hög grad uppbyggd enligt samma principer som Språkbanken i Göteborg, som ju också kommer att ingå i PAROLE-projektet. Om korpusen skall följa samma principer som den finska bör den TEI-kodas.

Förutom tidningstext är broschyrer från olika statliga och kommunala verk en texttyp som produceras på svenska för finlandssvenskar och som läses rätt allmänt. Problemet med dessa är att de för det mesta är översatta. Endast i kommuner med svenska som majoritetsspråk produceras sådant material primärt på svenska. Skillnaderna på de texter som hade producerats direkt på svenska och de som hade översatts från finska i en och samma kommun visade sig, enligt uppgift från forskaren Eivor Sommardahl, vara stora i ett kommunalt klar-språksprojekt som har genomförts i samarbete mellan svenska avdelningen vid Forskningscentralen för de inhemska språken och Finlands kommunförbund. Det är ju i och för sig ingen överraskning. Mera om problemet med översatta texter längre fram.

Det vore naturligtvis önskvärt att också den finlandssvenska korpusen kunde utvidgas till 20 miljoner ord, men det är långt ifrån säkert att det lyckas inom en överskådlig framtid, eftersom det knappast kommer att kunna avsättas lika mycket resurser för den svenska

korpusen som för den finska. Med de redskap som har utvecklats på Institutionen för allmän språkvetenskap vid Helsingfors universitet finns det dock möjlighet till relativt långt driven automatisk kodning, vilket gör att det ändå kan finnas rimliga chanser att också med små personresurser bygga upp en korpus av hyggligt omfång.

Också den finlandssvenska korpusen skall kunna användas för olika syften. För vissa ändamål, t.ex. för syntaktiska studier, borde ett lämpligt urval av texter koda mer detaljerat, men för de lexikografiska behoven (och i hög grad också för språkvårdens behov) är det bättre att få med en så stor mängd text som möjligt och nöja sig med en grov kodning. Det viktigaste är oftast att helt enkelt få belägg på att ett ord eller uttryck existerar, eller att få ett tillräckligt antal belägg för att det skall gå att dra slutsatser om betydelse(nyanser). (Jfr Sinclair 1992:4).

En kompletterande korpus

Texter producerade av skolelever

Det har skrivits spaltmetervis om finlandssvenskan, men trots det stötte vi i arbetet med den finsk-svenska ordboken ständigt på problemet att finlandssvenskan är synnerligen otillräckligt kartlagd vad gäller regional variation. Det finns en stor excerptsamling över s.k. finlandismer, d.v.s. ord och uttryck som är okända (eller ibland bara föråldrade) i rikssvenskan. Men i regel har de påträffade språkdragen kontrasterats enbart mot rikssvenskan. Någon jämförelse mellan olika regioner i Finland har inte gjorts, och mycket av det som anförs som finlandismer kan vara typiskt för någon mindre region och helt okända i andra regioner. Ju ledigare och mera talspråksnära språket är desto större är naturligtvis variationen. Det skulle behövas "mera kött på benen" därvidlag och för det ändamålet vore en korpus med regionalt förankrade texter ett intressant experiment.

Hur kunde en sådan korpus tänkas se ut? En språkligt relativt "oförstörd" grupp i det här hänseendet, och en grupp som ändå skriver en hel del, är grundskolelever. Ett sätt att komma åt regionalismer vore att utnyttja datoriseringen i skolorna och be svensklärarna (eller varför inte andra intresserade lärare) i ett antal skolor fördelade på olika regioner sända in orättade elevtexter på diskett. Eleverna måste självklart bes om lov och de skall naturligtvis få vara anonyma. Texterna bör koda dels med avseende på skribentens hemort, dels med avseende på språklig bakgrund – standardspråk, dialekt samt grad av

tvåspråkighet (t.ex. ena föräldern finsk, båda finska, finska kamrater eller släktingar).

Enligt uppgifter från lärarhåll är det stor skillnad på grundskole- och gymnasienivå. Gymnasieeleverna har en tydlig strävan mot och en relativt stark medvetenhet om den skriftspråkliga normen, medan grundskoleleverna skriver ett talspråksnära språk, vilket betyder att de regionala variationerna borde komma bra fram i deras uppsatser. Tar man med uppsatser från bägge nivåerna får man ett gott jämförelse-material och kanske därmed en uppfattning om etableringsgrad för olika regionalismer.

Problemet med ett projekt av det här slaget är att materialet är ytterst arbetsamt att samla in och att förenhetliga. Om man räknar med att målet i första fasen vore en korpus på ca 2 miljoner ord innebär det att man måste bearbeta ca 4.000 elevuppsatser à ca 500 ord. Räknar man med nio regioner skulle det betyda 445 uppsatser per region. Fördelar man det på flera skolor är det i och för sig kanske inte en omöjlighet att få in ett sådant antal texter i elektronisk form inom rimlig tid. Men det betyder med andra ord att varje region representeras av 222.500 löpord, vilket är alldeles för litet. För att se om ett sådant här företag över huvud taget är möjligt och försöka beräkna vilka resurser som krävs är det säkert skäl att först genomföra ett mindre pilotprojekt.

Talspråk

Som framgick av översikten ovan ingår det i den finlandssvenska korpussamlingen en korpus bestående av 80.000 ord transkriberat tal. Radiospråket representerar dock en relativt formell nivå, och för att undersöka regionala skillnader borde man få till stånd en korpus också med informellt tal. Tillsammans med elevuppsatskorpusen kunde en sådan talspråkskorpus ge en bättre grund för beskrivningen av informell finlandssvenska och regionala skillnader. Att sammanställa en talspråkskorpus är dock tidskrävande, så det kommer att bli en senare uppgift.

Texter riktade till barn och ungdom

Intressanta texter i framtidsperspektiv är sådana texter som riktar sig till barn i skolåldern. Det språket spelar säkert en inte så obetydlig roll när det gäller att forma det finlandssvenska "standardspråket" som det kommer att se ut när dagens skolelever vuxit upp – citationstecken

eftersom någonting sådant som ett finlandssvenskt standardspråk egentligen inte existerar. Främst skulle skolböcker kunna komma i fråga; det produceras en hel del läromedel på svenska i Finland, även om mycket också är översatt från finska.

Det skrivs också en del barn- och ungdomsböcker på svenska i Finland. Vad gäller konsumerad litteratur spelar de en relativt liten roll, eftersom det sverigesvenska utbudet är många gånger större, men rent språkligt är det ett intressant material. Många barn- och ungdomsförfattare strävar efter att skriva ett målgruppsanpassat språk, så att läsarna skall uppleva språket som sitt. Det här betyder att man antagligen kan träffa på en hel språkdrag som författarna uppfattar som typiskt finlandssvenska i texterna. Här får man dock vara noggrann med urvalet, om det är just finlandssvenska särdrag man vill komma åt. En del författare har i stället en strävan att anpassa sitt språk efter sverigesvenska för att "bibringa barnen ett bra språk".

Översatta texter

Jag har redan varit inne på problemet med att så mycket är översatt i Finland. Översatta texter skiljer sig ju språkligt från orginaltexter på ett språk, och man kan därför argumentera att översättningar från finska till svenska inte hör hemma i en finlandssvensk korpus, att de egentligen inte speglar finlandssvenskan. Å andra sidan vet vi att språkdrag kan sprida sig från språk A till språk B via frekvent översättning och småningom börja uppträda också i texter skrivna direkt på språk B (Gellerstam 1986:92f.). Översatta texter – av bättre eller sämre kvalitet – utgör en avsevärd del av de svenska texter man som finlandssvensk konfronteras med. De kan därför bidra till att språkdrag som är frekventa i just översatta texter småningom befästs i finlandssvenskan. Att översatta texter och icke-översatta texter måste hållas isär i korpusar säger sig självt.

Användningen av den finlandssvenska korpusen

Den finlandssvenska korpusen är tänkt att tjäna flera syften. Inom forskningscentralen kommer språkvårdarna att bli en mycket viktig användargrupp, men korpusen skall också kunna utnyttjas för forskning i finlandssvenska. Dessutom skall den utnyttjas av två lexikografiska projekt.

Finlandssvenskan i finsk-svensk ordbok

Korpusen kommer att vara en tillgång vid framtida uppdateringar och bearbetningar av den finsk-svenska storordboken. Under arbetet med det manuskript som nu är under tryckning var det ett ofta återkommande problem att det inte fanns tillräckligt med finlandssvenska texter i maskinläsbar form för att man skulle kunna räkna med att avsaknaden av belägg faktiskt indikerade att det sökta ordet inte användes. När det gäller den finlandssvenska som används i mera officiella sammanhang går det ibland att hitta belägg på olika ord och uttryck i andra källor. Men långt ifrån alla källor har ju index över viktiga ord och termer, och dessutom är ju inte alla ord man söker terminologiska. Att försöka plöja igenom texter på måfå för att leta efter belägg på något visst ord är ju som att leta efter en nål i en höstack.

Ger man ut en finsk-svensk ordbok i Finland är det alldeles klart att finländska särdrag skall tas med. Sådana finns på alla språkliga nivåer och det är besvärligt att dra gränser för hur de skall redovisas. Nedan några exempel, först ett par officiella termartade benämningar (förkortningarna *Suom.* och *Ruots.* med versaler för 'i Finland' resp. 'i Sverige'), sedan ett kulturbundet fenomen och sist ett talspråkligt uttryck (förkortningarna *suomr.* och *ruotsr.* för 'finlandssvenska' resp. 'rikssvenska').

kantopiiri *Suom.* utbärningsdistrikt -et =; **Ruots.**

utdelningsområde -t -n *myös suomr. yleisk.*

diplomi-insinööri *Suom.* diplomingenjör -en -er; **Ruots.** civilingenjör

talko|ot¹⁷ *suomr.* talko -t -n (-n -r); **talkootyö** talkoarbeta -t

-n; *ruotsr.* arbetsgille -t -n; grannhjälp -en; *järjestimme ~*

varaston siivoamiseksi vi samlade ihop frivilliga för att få

lagret städat; vi hjälptes åt för att få lagret städat; **suomr.** vi

ordnade talko för att få lagret städat; *oja kaivettiin -illa* vi

hjälpptes åt att gräva diket; diket grävdes med gemensamma

krafter; **suomr.** diket grävdes på talko (som talkoarbeta)

simah|aa^{53*F} **väsähtää** slockna¹; **mennä rikki:** vars. *ruotsr.*

*paja*1; *suomr.* *laitteista* säga⁴ upp kontraktet; *hän -i*

kesken matkan han slocknade på halva vägen; *-anut juhlija* en

festare som slocknat

Klart är att termartade uttryck som förekommer i officiellt språkbruk skall redovisas. Att också sådana ord och uttryck som används över hela det finlandssvenska språkområdet skall redovisas rådde det i princip enighet om inom redaktionen. Däremot var redaktionen inte

alltid enig beträffande vad som är "allfinländskt". Ju mer man rör sig i riktning mot informellt språkbruk, desto större blir de regionala skillnaderna och desto svårare blir det också att hitta belägg på de ord man söker. Här skulle alltså korpusen med skolelevstexterna och talspråkskorpusen, kunna ge indikationer på hur allmän spridning ord och uttryck har, detta förutsatt att de kan byggas ut tillräckligt.

Ett stort lexikografiskt problem kvarstår dock, oavsett alla korpusar, och det är vad man skall göra i de fall när det inte existerar ett gemensamt talspråkligt ord på finlandssvenska på samma sätt som det många gånger gör i finskan och i rikssvenskan. Skall man redovisa olika alternativ med regionala markeringar av något slag eller skall man gå in för en viss regions talspråk, t.ex. Helsingforsregionens, eller skall man helt och hållet utesluta de finländska varianterna i sådana fall. Ingen av lösningarna är särskilt tillfredsställande.

En ordbok över finlandssvenskan

För några år sedan började en ordbok över finlandssvenskan utarbetas på svenska avdelningen vid Forskningscentralen för de inhemska språken. Det är egentligen inte en ren ordbok, utan det kommer att vara en handbok bestående dels av en textdel med resonerande text, dels en ordboksdel med ca 2.000 uppslagsord. Orden utgörs dels av särfinlandssvenska ord, dels av allmänsvenska ord som helt eller delvis har fått annan betydelse eller användning i finlandssvenskan. Ordboksartikeln i övrigt innehåller antingen en förklaring eller en uppgift om den allmänsvenska motsvarigheten och eventuella uppgifter om bakgrunden till ordet eller användningen. Så här ser det ut i en preliminär version:

Fig. 1. Ur en preliminär version av den finlandssvenska ordboken

Uppslagsorden är hämtade ur ett omfattande register över s.k. finlandismer, d.v.s. ord och uttryck som antingen används bara i finlandssvenskan eller som används i en annan betydelse i finlandssvenskan än i sverigesvenskan. Registret omfattar för närvarande ca 5000 poster och det har tillkommit dels genom att litteratur *om* finlandssvenska har exciperats, dels genom att finlandssvenskt textmaterial exciperats. Excerpteringen har gjorts av många olika personer och det excerpterade materialet har hållit varierande standard, så det finns upptaget en hel del sådant som verkar vara ytterst tillfälliga bildningar eller direkta språkfel. I allmänhet finns heller ingenting noterat om frekvens eller om regional utbredning. Det vore alltså en stor fördel att ha en tillräckligt omfattande korpus att systematiskt kontrollera registret mot. Samtidigt skulle säkert en sådan genomgång tillföra registret en hel del nya poster.

Registret kunde sedan utgöra utgångspunkten när lemmalistan till den finlandssvenska ordboken utarbetas. Tilltänkta uppslagsordskandidater kontrolleras mot korpusen och ord och användningar av ord som finns väl belagda tas med. På så vis får man en adekvat kärnuppsättning uppslagsord.

Å andra sidan råkar man naturligtvis alltid ut för att det finns bara något enstaka belägg i en korpus eller inga alls, hur stor korpusen än är. De fallen måste man bedöma utifrån andra premisser. En korpus kan knappast vara det allena saliggörande vid ordboksarbete. Den är ju alltid bara ett litet urval av språket, hur omfattande den än är.

För närvarande ligger redigeringsarbetet på is, eftersom inga personella resurser kan reserveras för det. Sannolikt kommer dock den första versionen av den finlandssvenska ordboken att utarbetas utgående från det preliminära manus som nu föreligger, men för framtida upplagor av ordboken kommer korpusen att bli en viktig bas.

Korpusanvändning inom tvåspråkslexikografin

Som tvåspråkslexikograf ställer man sig ibland litet undrande inför debatten om huruvida bara autentiska exempel får användas i en ordbok eller om även redaktionella exempel kan tillåtas. Det är en debatt som bara kan gälla enspråkslexikografin. Tvåspråkslexikografen har inget val. Hur enorma korpusar man än får till sitt förfogande kan man aldrig räkna med att för varje källspråksexempel man väljer ut hitta ett målspråksexempel som är en exakt motsvarighet och som oredigerat kan fungera som översättning. Man blir helt enkelt tvungen att bearbeta exemplen redaktionellt, ibland både käll- och målspråksexemplet, ibland åtminstone endera. Avskalade exempel fungerar dessutom oftast bättre i en tvåspråkig ordbok än obearbetade. Dels ger de generellare information, dels ger de inte upphov till onödiga översättningsproblem som är ovidkommande för det språkliga fenomen man vill belysa och som ibland rent av kan skymma det man ville klargöra.

Korpusarna för käll- och målspråk utnyttjas för olika ändamål. Om man börjar från noll när man utarbetar en ny tvåspråkig ordbok används källspråkskorpusen på i stort sett samma sätt som när man utarbetar en ny enspråkig ordbok. Det här är väl dock relativt ovanligt. Det vanligaste är väl att man utnyttjar ett existerande material, antingen så att en existerande ordbok revideras eller så att man utnyttjar en enspråkig ordbok eller en ordbok till ett annat målspråk som bas. Källspråkskorpusen använder man då i första hand för kompletterande uppgifter; för att få belägg på ordbetydelser eller användningar som man upptäcker fattas eller för att få fler exempel.

Målspråkskorpusen letar man i stor utsträckning i på källspråkets villkor, d.v.s. man letar efter ord, uttryck och konstruktioner som i så hög grad som möjligt motsvarar ett källspråksexempel, och man letar efter belägg på att ett ord som man tänker sig som ekvivalent faktiskt existerar.

En gemensam bas för olika tvåspråkiga ordböcker

En tanke som nu har framkastats på forskningscentralen är att utarbeta en gemensam bas för tvåspråkiga storordböcker från finska till andra språk. Andra målspråk än svenska kommer att bli aktuella framöver, och då vore det viktigt att ta vara på erfarenheterna från det finsk-svenska ordboksprojektet.

Det går naturligtvis inte utan vidare att överta en enspråkig ordbok och "översätta" materialet i den. På det viset uppstår inte en adekvat tvåspråkig ordbok. En tvåspråkig ordbok från språk A till språk B är inte nödvändigtvis heller en bra bas för en ordbok från språk A till språk C, eftersom ett annat målspråk aktualiserar delvis andra kontrastiva aspekter. En gedigen bas kunde däremot åstadkommas genom en syntes av en enspråkig och en tvåspråkig. För varje ny tvåspråkig ordbok som görs kan materialet kompletteras med nya kontrastiva aspekter. På det här viset skulle man upprätthålla en bas som är betydligt mer omfattande än vad de färdiga ordböckerna är tänkta att bli, och som utnyttjas till de delar som det aktuella målspråket förutsätter.

Det första steget skulle alltså vara en syntes av Suomen kielen perussanakirja och Finsk-svensk ordbok. Det senare materialet borde analyseras systematiskt utifrån kontrastiva aspekter, för i sitt nuvarande skick är det behäftat med vissa brister som beror på att en enspråkig ordbok legat till grund. Även om gallringar och tillägg har gjorts under redigeringsarbetets gång är nog materialet i vissa avseenden för "enspråkigt". En jämförelse med andra tvåspråksordböcker skulle säkert ge en hel del matnyttigt i all synnerhet vad gäller exempelaterialet.

Såsom tidigare nämndes kommer Suomen kielen perussanakirja att uppdateras. Kompletteringarna skall göras utgående från den nya nufinska korpusen. De uppdateringar och kompletteringar som görs för den enspråkiga ordboken bör naturligtvis utnyttjas också för den tvåspråkiga basen, men dessutom bör nog en separat genomgång av korpusarna göras för denna, eftersom bl.a. exempelvalet delvis utfaller annorlunda i så fall, än om man letar efter exempel för en enspråkig ordbok.

För att åstadkomma bra produktionsordböcker vore antagligen en "baklängesgenomgång" av korpusaterialet givande, d.v.s. att utgående från typiska uttryck på målspråket leta efter lämpliga motsvarande uttryck i källspråket. Att ta med typiska vändningar på målspråket och översätta till källspråket är i regel ingen bra lexikografisk metod, eftersom det lätt leder till otypiska ingångar, som ingen modersmåls-talande användare kommer på tanken att slå upp. Men om man söker motsvarigheter till de typiska vändningarna i ett autentiskt material kan metoden ge en hel del matnyttigt.

Möjlighet till morfologiska sökningar i de svenska korpusarna

I svenska korpusar når man hyggliga resultat vid fritextsökning, eftersom orden med stamförändringar är få. Men ibland vore möjligheten till sökning på morfologisk grund önskvärd, särskilt när man arbetar kontrastivt med svenska och ett annat språk. Sökningar som finska språket aktualiserar och som förhoppningsvis så småningom blir möjliga att genomföra i Språkbankens korpusar är t.ex.:

- Att kunna hålla isär adverb och t-former av adjektiv och particip så att man vid en sökning får listat bara de belägg där -t-formen faktiskt är adverb.
- Att vid finalalfabetiska sökningar kunna söka på (efterleds)stam och inte som nu vara tvungen att göra en separat sökning för varje form av ordet, t.ex. *-flicka, -flickan, -flickas, -flickans, -flickor, -flickorna, -flickors, -flickornas*.

En annan sökning vi på den finsk-svenska ordboksredaktionen ofta saknat i Språkdatas korpusar är möjligheten att leta efter ordkombinationer där de två orden skiljs åt av ett eller flera andra ord. Med en sådan möjlighet kunde man

- söka efter prepositionsfraser och valensuppgifter i större omfattning än vad som nu är möjligt
- kontrollera semantiska selektionsregler – t.ex. hurdana typer av subjekt respektive objekt ett visst verb kan ta. Det är ju ofta en synnerligen viktig information i en tvåspråkig produktionsordbok.

Sådana sökmöjligheter finns inte åtminstone tills vidare i den version som finns tillgänglig över nätet och som vi på forskningscentralen utnyttjar.

En tvåspråkig korpus

Att skapa en omfattande korpus där samma texter ingår på båda språken är ett stort projekt. Dels är det naturligtvis dubbelt jobb att få fram samma texter på båda språken (dubbelt fler rättigheter att skaffa fram t.ex., eller rent av tre gånger fler om man också tar med översättningar från tredje språk). För det andra är det ju en dubbel uppsättning texter som skall koda på lämpligt sätt.

Åtminstone i den första fasen kunde man därför skapa en mindre omfattande korpus som i första hand skulle utnyttjas för att undersöka

motsvarigheterna till vissa ordtyper, inte för enskilda ord. I den mån korpusen kunde utnyttjas för att kontrollera också översättningar av enstaka ord är allt gott och väl, men det skall inte vara det primära målet.

Tanken på en parallellkorpus för finska och svenska uppstod under arbetet med att redigera Finsk-svensk ordbok, eftersom vi under hela projektet gick en ojämn kamp med vissa oöversättbara ordtyper, ord som antingen helt och hållet saknar svenska ekvivalenter eller vars ekvivalenter inte kan ingå i samma typer av konstruktioner. Som lösning på problemet bestämde vi oss för att utarbeta syntaktiska modeller för hur orden kan hanteras vid översättning till svenska. (LexicoNordica 2:89 ff.). För att få fram exempelmaterial använde vi dels Helsingfors universitets finska korpus, dels ett par tre pocketböcker i en finsk och en svensk version parallellt. Den här parallellläsningen visade sig vara nyttig i all sin enkelhet och en mer omfattande korpus skulle säkert ge betydligt mer material.

Idén med modellartiklar har vi tänkt vidareutveckla i framtida upplagor. Dels kan de befintliga modellerna kanske ytterligare bearbetas utifrån korpusmaterialet, dels finns det fler ordtyper som vi gärna skulle åskådliggöra i modeller. Efter det att artikeln för LexicoNordica 2 skrevs har vi utarbetat en modell också för kausativa/faktiva verb. Den var vanskelig att göra och skulle säkert vinna på en genomgång av korpusmaterial.

En typ vi gärna skulle komma åt är paren punktuella – iterativa/durativa verb som finskan är så rik på. Det punktuella verbet motsvaras ibland men långt ifrån alltid av partikelverb med *till* på svenska. Den iterativa typen är ännu mer problematisk. Finska iterativa verb används dels för att ange att det plötsliga skeendet sker gång på gång, dels tillsammans med plurala subjekt. Ibland kan svenskan upprätthålla en liknande skillnad med hjälp av ett adverbial av något slag i satsen, men ofta verkar det som den här aspekten inte skulle uttryckas på svenska. Hur den problematiken skall lösas i en ordbok är långt ifrån klart, men en parallellkorpus kunde ge ett bättre underlag för kartläggningen av problemet.

En annan typ av verb som kunde behandlas i en modellartikel är verbpar som består av ett intransitivt verb på *-ua/-yä* och ett motsvarande transitivt (t.ex. *muuttua* 'förändras' – *muuttaa* '(för)ändra'). Par av denna typ är mycket vanligare i finskan än vad par med *-s*-verb och transitivt verb är på svenska. (Se närmare Martola 1995a:287 f.)

En grupp av ord där den tvåspråkiga korpusen också kunde vara till hjälp är sammansättningar av bahuvrihityp. Det finns betydligt fler lexikaliserade bahuvrihisammansättningar på finska än på svenska. De

finska orden kan dessutom i regel vara både adjektiv och substantiv och som substantiv kan de ofta avse både egenskapen och bäraren av egenskapen.

Ytterligare en grupp som det vore intressant att undersöka med hjälp av en parallellkorpus är komparativ- och superlativformer av sådana ord vars svenska motsvarigheter inte kan kompareras. Den här typen är synnerligen besvärlig att behandla i en vanlig ordboksartikel (t.ex. *laidempi* 'som befinner sig närmare kanten' till substantivet *laita* 'kant, rand').

Problemet med en parallellkorpus är naturligtvis först och främst att uppbringa samma texter på bägge språken. Om man skall hitta samma ställe i båda texterna kan man inte använda sig av texter där den ena versionen är förkortad eller bearbetad, utan de borde i väldigt hög grad motsvara varandra.

Tanken är att i så hög grad som möjligt dra nytta av det finska korpusmaterialet, så att t.ex. de romaner som finns översatta tas med också på svenska. Många av de fenomen vi vill komma åt är relativt talspråksnära, så romaner utgör därför ett bra material, och också t.ex. barn- och ungdomsböcker. För att balans mellan språken skall nås bör det finnas med översättningar i bägge riktningarna och gärna också översättningar från tredje språk.

Sammanfattning

Svenska avdelningen vid Forskningscentralen för de inhemska språken har planer på dels en utvidgad finlandssvensk korpus, dels en tvåspråkig parallellkorpus. Den existerande finlandssvenska korpusen kommer att byggas ut i samarbete med Institutionen för nordiska språk vid Helsingfors universitet. Tills vidare har det konkreta arbetet inte inletts. I artikeln beskrivs främst de behov av korpusar som gjort sig gällande i arbetet på en finsk-svensk storordbok.

Litteratur

Tryckta källor

Finsk-svensk storordbok. Forskningscentralen för de inhemska språken. WSOY. Under tryckning.

- Gellerstam, Martin 1986: Translationese in Swedish novels translated from English. I *Translation Studies i Scandinavia*. Ed. Lars Wollin and Hans Linquist. Lund: CWK Gleerup, 88–95
- Lehtinen, Marja – Karvonen, Pirjo – Rahikainen, Tarmo 1995: *Tekstikorpukset*. Kotimaisten kielten tutkimuskeskuksen julkaisuja 81. Helsinki.
- Martola Nina 1995a: Från enspråkig till tvåspråkig ordbok. I *Nordiske studier i leksikografi* 3. Skrifter utgitt av Nordisk forening for leksikografi, 283–293
- Martola Nina 1995b: Substantiv avledda av adjektiv och verb – en jämförelse mellan några ordböcker med finska som källspråk. I *LexicoNordica* 2, 89–108

Otryckta källor

- Sinclair, John 1992: Lexicographers' Needs. Pisa Workshop on Text Corpora. (stencil)
- Le Parole. Technical Annex. Nov 5th 1995. Rev. Nov 19th 1995 - V1.5 (stencil)

Övrigt

Språkbanken, Göteborgs universitet
FISC-korpusen, Helsingfors universitet

Ole Norling-Christensen

Klassifikation af korpustekster, og kvantitative mål for sammensætningen af et almensprogligt korpus

The paper gives an introduction to the design and composition of general language corpora, and the problem of statistical representativeness is considered. In order to determine, check and document their composition, a classificatory scheme for text types is needed. The building of a corpus is seen as an iterative process: The original design may set up overall quantitative measures for a few easily distinguishable and rather general text types; after the collection of some texts or text samples, which should be annotated according to a more fine-grained classification scheme, the distribution by more specific classes (e.g. topic under a general class of non-fiction) is measured, and the design criteria are adjusted accordingly. The design of the corpus of The Danish Dictionary and of the corpus of Danish which constitutes part of the European PAROLE project, are used as examples.

Indledning

Det største og bredest sammensatte danske korpus er opbygget til brug for Den Danske Ordbog (DDO). Det omfatter i alt 40 millioner tekstord fordelt på henved 44.000 forskellige tekstprøver fra perioden 1983–92, herunder samtlige tekster (fire millioner tekstord) fra Henning Bergenholtz' DK87–90. Erfaringerne fra opbygningen af DDO's korpus, og også en del af dets tekster, indgår i det europæiske korpus- og ordbogsprojekt PAROLE, som er en fortsættelse af NERC og desuden bygger på bl.a. standardiseringsforslag fra EAGLES.

De tre nævnte akronymer er navne på EU-støttede projekter inden for sprogforskning og sprogteknologi. **NERC** (Network of European Reference Corpora) søgte ca. 1991–93 at sammenfatte den daværende viden om korpora og korpusarbejde med henblik på at rådgive EU (Generaldirektorat XIII i Luxembourg). Hovedresultaterne er samlet i NERC (1995), som desuden rummer en detaljeret plan for et europæisk korpussamarbejde. Inden for **EAGLES** (European Advisory Group on Language Engineering Standards) har en række ekspertgrupper 1993–96 udarbejdet forslag til fælleseuropæiske standarder for opbygning og dokumentation af korpora og for den leksikografiske beskrivelse af EU-sprogene; de fleste af dem kræver dog yderligere afpudsning, og en

række delområder mangler endnu at blive behandlet. Første fase af **PAROLE** (**P**reparatory **A**ction for Linguistic **R**esources **O**rganization for Language **E**ngineering) måtte derfor bl.a. omfatte udarbejdelsen af en fælles norm for sammensætning og opbygning af de europæiske korpora (Norling-Christensen 1996), samt en norm for deres tekniske udformning (Ridings 1996). Under anden fase (1996–98) opbygger de enkelte sproggrupper (14 vesteuropæiske sprog) nu offentligt tilgængelige korpora, som følger disse normer. Endvidere konstrueres for hvert sprog et **leksikon**, dvs. en ordbog der kan anvendes af datamaskiner.

I det følgende gives en introduktion til planlægningen og opbygningen af almensproglige korpora med eksempler fra DDO's korpus, det danske PAROLE-korpus og et subkorpus heraf.

Definitioner

Med udgangspunkt i en klassifikation opstillet af Bernard Quémada skelner Atkins/Clear/Ostler (1992:1) mellem fire slags (elektroniske) tekstsamlinger: Et **arkiv** er et lager af elektroniske tekster uden særlig sammenhæng og ofte i forskellige tekniske formater, fx hidrørende fra forskellige tekstbehandlingssystemer. Et **elektronisk tekstbibliotek** (engelsk: electronic text library, ETL; fransk: textothèque) er en samling af elektroniske tekster i standardiseret format; sammensætningen af biblioteket kan følge visse retningslinier med hensyn til fx indhold, men uden at der er lagt strenge begrænsninger på udvælgelsen. Et **korpus** er da en delmængde af et bibliotek, udvalgt efter eksplicite kriterier og med et bestemt formål for øje. Endelig forstås ved **sub-korpus** en delmængde af korpuset, fx fremkommet som mellemresultat under elektronisk korpusanalyse.

Denne præcisering af korpusbegrebet, som jeg vil søge at fastholde i det følgende, synes i dag at være almindeligt accepteret i fagkredse; men endnu i NERC (1995:63) tales der om

a multifunctional corpus.. [which] .. should contain full texts rather than samples .. [and] .. should be extensively documented, so as to facilitate the selection of specific subcorpora, which can be expanded in order to construct large specialized or task-oriented corpora,

altså snarere et tekstbibliotek. Også Atkins/Clear/Ostler (l.c.) vælger i øvrigt, efter at have givet de nævnte definitioner, for kortheds skyld at

bruge *corpus* som et fælles overbegreb for bibliotek, korpus og subkorpus.

Som eksempler på de "bestemte formål", der har været retningsgivende for sammensætningen af nogle danske korpora, kan nævnes, at Maegaard og Ruus opbyggede (jf. fx Ruus 1995,1:19) DANwORD med henblik på undersøgelse af ordhyppigheder, DK87-90 skulle især tjene som en almensproglig reference for fagsproglige korpusstudier, mens DDO's korpus er beregnet til ordbogsredigering.

Repræsentativitet

Statistiske undersøgelser af et repræsentativt udsnit af befolkningen, eller af en population som statistikerne siger, kan give meget præcise oplysninger om populationen som helhed. Det gælder, hvad enten populationen er Danmarks befolkning på et givet tidspunkt, ti årgange af et bestemt dagblad, jf. Manual (1994:95), eller fx den del af et uddødt sprog som er overleveret. Stikprøvetori er den del af den statistiske teori, som omhandler metoder til udvælgelse og analyse af stikprøver fra en population, der består af et endeligt og kendt antal enheder. Enhederne skal i princippet kunne opregnes på en liste (som for den danske befolkning fx kan være folkeregisteret), fra hvilken man efter visse regler kan udvælge en mindre del til nærmere undersøgelse. Hvis enheden ikke kan defineres eksakt, eller hvis populationens størrelse ikke kendes, kan stikprøvetori ikke give eksakt vejledning, men højst tjene som en kvalitativ rettesnor. Korpusbyggeren må se i øjnene at korpus ikke bliver repræsentativt i statistisk forstand; i stedet må han stræbe efter at korpus, med mindre præcise termer, bliver *eksemplarisk* (Manual l.c.) eller *(af)balanceret* (Atkins/Clear/Ostler 1992:6) med hensyn til det sproglige materiale det skal være en model af. Dog søger Biber (1993) at fastholde ideen om statistisk baseret repræsentativitet, jf. kapitlet "Udvælgelse af tekstprøver" nedenfor.

Det almensproglige korpus

Formålet med DDO's korpus er at det skal udgøre en væsentlig del af det empiriske grundlag for en *deskriptiv ordbog over almensproget*, og det egentlige problem er i første omgang ikke af statistisk natur, men derimod at fastslå, hvad der skal forstås ved almensprog. Først når der foreligger en operational definition af det sprog der skal beskrives, vil statistisk teori eventuelt kunne vejlede med hensyn til hensigtsmæssig udvælgelse af stikprøver.

Skulle ordbogen have dækket et nøjere afgrænset område af sproget, fx ét forfatterskab eller ét bestemt fagområdes sprog, ville problemet om korpus' sammensætning være noget enklere. Ved enkelt-forfatter-skaber kan man overveje at medtage alle tekster fremfor stikprøver. Ved fagsprog, hvor en vis normering ofte vil være ønskelig, kan fagfolk hjælpe med til at udpege fagets centrale og typiske tekster, og man kan supplere med lærebøger el.l. med gode registre. Havde målet endelig været en normativ ordbog over (fx) det danske almensprog, ville man vel have nedsat en komité af "gode danske mænd" til at udvælge "gode danske tekster".

Måske afgrænses almensproget bedst ved udelukkelsesmetoden. DDO valgte at udelukke det sprog, som fagfolk bruger til at kommunikere med andre fagfolk om deres fælles fag. Sproget i videnskabelige afhandlinger, eller i mekanikerens reparations-manualer, hører således efter vor opfattelse ikke til i en almensproglig ordbog. De danske fagsproglige korpora karakteriserede deres tekster ved en simpel sender modtager-model:

lægmand → lægmand
 lægmand → fagmand
 fagmand → lægmand
 fagmand → fagmand

Det er den sidstnævnte kategori, fagsprog i snæver forstand, som vi har udelukket.

Modellen er nok for simpel. Manual (1994:16) indfører endnu en rolle: "halvfagmanden" (the semi-expert), som fx kan være under uddannelse i faget, være fagmand inden for en beslægtet disciplin, eller være popularisator ([fag]journalist, marketingmedarbejder). Anlægger vi et receptionssynspunkt, ser på, hvilket sprog folk udsættes for (når de ikke er beskæftiget med deres fag), indtager halvfagmandens sprog en helt central rolle. Det skrives og tales af de journalister og andre formidlere, der i presse, radio og tv formidler politik, økonomi, økologi, sport, musik, bilteknik, osv. Sådanne tekster er i DDO's terminologi "(alment) fagsprog", og det hører klart med til grundlaget for en almensproglig ordbog.

Hermed har vi opridset konturerne af et *almensprogligt korpus*; det kan være en samling af prøver på så mange og så forskellige tekster som muligt; de må ikke være fag-interne og de skal være rettet mod og/eller produceret af brede dele af befolkningen. Men vi har ikke hermed sagt, hvad *almensprog* er. Man kunne definere det snævert, som den del af sproget alle danskere (/svenskere etc.) er fælles om – den

kerne af ord, som vi alle forstår. Ruus (1995) anviser i sin disputats en korpusbaseret metode til at bestemme det centrale ordforråd, anvender den på sproget dansk, og konstruerer en ordbog over 1117 kerneord i form af et semantisk netværk. Dette er spændende lingvistisk grundforskning; men selv hvis denne ordbog blev omsat til en form, som gjorde det muligt for den almindelige ordbogsbruger at slå op i den, ville brugeren nok blive skuffet. For man slår jo ikke op for at finde ud af det, alle ved i forvejen, men det, man kan være i tvivl om. Sat på spidsen, kunne man sige, at når vi i vores ordbøger definerer det velkendte, det indlysende, så gør vi det mere for vor egen skyld end for brugerens. Dog bør det tilføjes, at ikke al sproglig kompetence er bevidst; én af ordbogens opgaver er at minde om det, man egentlig godt vidste.

Kerneordene er, groft sagt, ord der forekommer meget hyppigt i mange forskellige slags tekster; de vil derfor også optræde hyppigt i det almensproglige korpus. Men derudover vil de mange forskellige tekstarter, omhandlende mange forskellige emner, bidrage med *deres* specifikt hyppige ord. Den bredt anvendelige almensproglige ordbog får dermed ikke kun 1117 opslagsord, men snarere fem eller ti gange så mange, og dermed måske 50.000 eller flere, som ikke hører til kerneordene. De er tværtimod i ordbogen, fordi mange møder dem, men ikke alle er sikre på hvad de betyder eller hvordan de bruges.

Det sidste, hvordan de bruges, er vigtige oplysninger i en ordbog, men er ikke med i "Kerneordene". Noget af det, statistik på et godt korpus kan belyse, hvis korpus er stort nok, er syntaktiske mønstre (valens), samt ordkombinatorikken, altså hvilke ord der typisk forekommer sammen. *Blussende rød* og *strålende hvid* er godt og typisk dansk; *blussende gul* og *strålende sort* er syntaktisk korrekt, men ganske umuligt. Kontaminerede idiommer kan vise sig at være at være hyppigere end de oprindelige: *Gammel rotte* og *garvet politimand* er godt gammelt dansk. Men det meningsløse *garvet rotte* udviser overraskende højt forekomsttal, og må registreres i ordbogen. Sådanne forhold afspejles i korpus og kan eftervises med statistiske metoder, hvis korpus er tilstrækkelig omfattende.

Udvælgelse af tekstprøver

Stikprøveteorien skelner mellem simpel tilfældig udvælgelse og stratificeret udvælgelse.

Ved **simpel tilfældig udvælgelse** har alle individer samme sandsynlighed for at blive udvalgt; sandsynligheden for, at en bestemt slags

individ bliver udvalgt, er simpelthen antallet af denne slags individer divideret med det totale antal individer i populationen. For udvælgelsen af tekstprøver til et korpus rejser denne definition straks tre problemer: Hvilke individer er der tale om (ord, sætninger, afsnit, hele tekster)? – Hvad er det totale antal individer i populationen, som jo er hele det sprog der ønskes en model af? – Og er samme sandsynlighed for ethvert individ overhovedet ønskelig?

Biber (1993:247) anslår, at et korpus, der skulle være repræsentativt i demografisk forstand, altså rumme de sproggenrer som folk typisk bruger i samme forhold som de bruger dem, ville bestå af ca. 90% samtale og 3% breve og notater, mens de resterende 7% måtte deles af alle øvrige genrer,

such as press reportage, popular magazines, academic prose, fiction, lectures, news broadcasts, and unpublished writing. (Very few people *ever* produce published written texts, or unpublished written and spoken text for a large audience). Such a corpus would permit summary descriptive statistics for the entire language represented by the corpus. These kinds of generalizations, however, are typically not of interest for linguistic research.

I stedet for den demografiske repræsentativitet foreslår han derfor, at korpus sammensættes ved stratificeret udvælgelse på en sådan måde, at det bedst muligt belyser den sproglige variation, og han angiver statistiske metoder til at måle om givne sproglige træk er tilstrækkeligt repræsenteret; de vil fx vise, at 90% samtale i et stort korpus er langt mere end nødvendigt for at give et pålideligt billede af denne tekststart.

Ved **stratificeret udvælgelse** inddeles populationen i forskellige lag, **strata**, iht. kriterier, som på forhånd er kendt. Den tilfældige udvælgelse foretages separat for hvert stratum med henblik på, at man efterfølgende skal kunne analysere stikprøverne separat for hvert stratum. Ved befolkningsundersøgelser kan kriterierne fx være alder og køn, for korpusopbygning er fx genre eller teksttype relevante kriterier, og man vil stræbe efter tilstrækkelig repræsentation af hver type og altså tildele sjældent forekommende teksttyper forholdsvis større vægt end de hyppige. Yderligere foreslår Biber, at korpus opbygges ad flere omgange, så udvælgeskriterierne kan justeres undervejs. Jeg kan ikke her gå yderligere ind på Bibers argumentation eller på hans skelnen mellem klassifikation/stratificering efter henholdsvis tekst-eksterne kriterier ("genre" eller "register") og tekst-interne kriterier ("teksttype"), men må henvise interesserede læsere til den pågældende artikel.

Den Danske Ordbogs korpus

Uden at kende Bibers arbejder fulgte vi faktisk i stor udstrækning hans anbefalinger, da vi i 1991 planlagde DDO's korpus. Dog blev størrelsen af den enkelte tekstprøve, det ovenfor omtalte individ, ikke nøjere præciseret; kortere tekster er medtaget i deres helhed, mens der fra længere skrevne tekster, typisk hele bøger, er udvalgt et eller nogle få kapitler, i alt højst 10.000 tekstord. På basis af dikotomierne skriftsprog / talesprog, reception ("offentligt sprog") / produktion ("privat sprog"), og almensprog / (alment) fagsprog, opdeltes korpus i otte strata, og for hvert fastsattes skønsmæssige mål for deres andel af de 40 millioner tekstord. Hver tekstprøve blev annoteret med oplysning om stratum, men desuden en række andre oplysninger, bl.a. om medie, genre, emne og produktions- eller trykår, samt om sprogbrugerens alder, køn m.v. Hermed var indledt en iterativ proces, hvor den første udvælgelse skete efter ret grove kriterier, de otte strata, hvorefter teksternes fordeling med hensyn til de øvrige kriterier løbende kunne kontrolleres og den videre tekstakkvisition justeres derefter. Denne metode bidrog fx væsentligt til, at mange forskellige emner blev repræsenteret blandt fagteksterne. En fyldig beskrivelse af korpus, arbejdet med at opbygge det, og hvad det bruges til, gives i Norling-Christensen & Asmussen (i trykken); en fuldstændig oversigt over klassifikationen efter medie, genre og emne findes i Norling-Christensen (1996).

Parole-korporaene

De korpora, der produceres inden for PAROLE-projektet skal kun dække skriftsprog, er primært til brug inden for sprogteknologi, og skal være offentligt tilgængelige. Allerede disse krav udelukker en hel del af DDO-materialet fra at blive genbrugt, først og fremmest talesproget; men også visse andre teksttyper må anses for mindre relevante i denne sammenhæng. Desuden udelukkes tekster, som af forlag m.fl. er stillet til rådighed med den klausul, at de alene må anvendes til ordbogsarbejdet, samt andre tekster som af ophavsretsmæssige grunde ikke kan mangfoldiggøres og distribueres.

Budgettet for det enkelte sprog er begrænset, og bl.a. af denne grund er de fælles krav til korpussammensætning ret løse. Kun for fordelingen på medier (hvor der bl.a. sættes en overgrænse for andelen af avistekster) og for perioden (højst 20% fra 1970'erne, og intet ældre) er der fastsat bindende retningslinier.

Mindst 250.000 ord skal være grammatisk annoteret (såkaldt **tagging**), minimalt med ordklasse; taggingen skal ikke blot være maskinel, men kontrolleret af mennesker, så den sidenhen kan bruges til træning og/eller afprøvning af taggere, dvs. programmel der automatisk tilordner ordklasse og eventuelt andre morfosyntaktiske oplysninger til tekstord. Til brug for opbygningen af det regelsæt, som skal anvendes ved automatisk tagging af det danske PAROLE-korpus, er der ved stratificeret udvælgelse udtrukket et subkorpus på 100.000 tekstord, som er ved at blive manuelt tagget. De er fordelt på ca. 700 tekstprøver, taget fra hver sin tekst; en prøve består af hele afsnit op til et maksimum af 160 tekstord. Dette design forventes at sikre maksimal spredning mht. sådanne sproglige træk som er af betydning for automatisk tagging; og som en sidegevinst opnås formentlig, at de 100.000 ord kan betragtes som en citatsamling og distribueres uden ophavsretlige problemer.

Afslutning

Som det især er fremgået af omtalen af PAROLE-korporaene, kan korpusbyggeren ikke nøjes med teoretiske overvejelser af lingvistisk, statistisk og sociologisk art. Også praktiske og økonomiske overvejelser hører med: Hvilke tekster kan overhovedet fremskaffes? Hvad koster det at scanne eller indtaste dem, hvis de ikke allerede foreligger i et egnet elektronisk format? Hvor megen annotation er der resurser til at udføre? Hvad med ophavsretten? – Til en lang række formål vil mange korte tekstprøver være at foretrække frem for fulde tekster. Men med fulde tekster (hele bøger) når man langt hurtigere, nemmere og billigere op på den ønskede korpusstørrelse. Det gælder her, som i næsten alle forhold, at det praktisk mulige må afvejes mod det ideelt ønskelige, og at kompromiser ikke kan undgås. Lad det være en trøst, at

In our ten years' experience of analysing corpus material for lexicographical purposes, we have found any corpus – however 'unbalanced' – to be a source of information and indeed inspiration. Knowing that your corpus is unbalanced is what counts. It would be shortsighted indeed to wait until one can scientifically balance a corpus before starting to use one, and hasty to dismiss the results of corpus analysis as 'unreliable' or 'irrelevant' simply because the corpus used cannot be proved to be 'balanced'. (Atkins/Clear/Ostler 1992:6).

Litteratur

- Atkins, Sue / Jeremy Clear / Nicholas Ostler 1992: Corpus Design Criteria. In *Literary and Linguistic Computing* 7.1, 1–16.
- Biber, Douglas 1993: Representativeness in Corpus Design. In *Literary and Linguistic Computing* 8.4, 243–257.
- Manual 1994 = *Manual i Fagleksikografi*. Red. Henning Bergenholtz og Sven Tarp. Herning: Systime.
- NERC 1995 = Calzolari, N. / M. Baker / P.G. Kruyt (eds) 1995: *Towards a Network of European Reference Corpora*. Report of the NERC Consortium, feasibility study, coordinated by A. Zampolli. Pisa: Giardini.
- Norling-Christensen, Ole (ed.) 1996: *Design and Composition of Reusable Harmonized Written Language Reference Corpora for European Languages*. PAROLE (MLAP: 63-386) WP 4 – Task 1.1. København: Det Danske Sprog- og Litteraturselskab.
- Norling-Christensen, Ole / Jørg Asmussen, i trykken: The DDO Corpus of Contemporary Danish. In *Computers and the Humanities* (forthcoming).
- Ridings, Daniel 1996: *Text Representation in PAROLE*. Göteborg.
- Ruus, Hanne 1995: *Danske Kerneord. Centrale dele af den danske leksikalske norm 1–2*. København: Museum Tusulanums Forlag.

Liisa Nuutinen

Korpus över äldre finska

The articles in the Dictionary of Old Finnish are based on the collection of about half a million entries of words and a corpus in machine-readable form, the compilation of which was started in 1992. The corpus contains the essential documents of old Finnish and additionally other texts from different fields. In choosing the material, the goal has been a coverage as complete as possible from the lexicographic point of view. At present, the corpus contains over 2 million word forms in running texts, and the editors of the dictionary can pick the example sentences directly from it. The search programs always go sentence by sentence in looking for words, word forms or sequences or signs. It is also possible to look for alternative words or sign sequences, either exact or approximate ones, or limit the search to morphological elements. When necessary, word lists can be made using KWIC program.

Om Vanhan kirjasuomen sanakirja

Med *äldre finska* avses finskan i de första skriftliga dokumenten från mitten av 1500-talet fram till år 1810. Denna finska beskrivs i *Vanhan kirjasuomen sanakirja* ('Ordbok över den äldre finskan'), som håller på att utarbetas vid Forskningscentralen för de inhemska språken. Tryckta verk och enhetliga manuskript på finska från denna tid utgör basen, så att alla finska ord som förekommer i den här tidens litteratur tas med i ordboken som uppslagsord. Uppslagsordsformen är för det mesta normaliserad till modern finska, och efter det ges de olika varianterna. Sedan följer ordklassangivelse och efter det kommer uppgifter som hämtats ur gamla ordböcker, sedan ordets definition och sist exempen. Ordboken över äldre finska är en citatordbok. Det betyder, att exempen får tala för sig själva genom att de presenteras i betydelsegrupper. Citaten har valts så att de förutom betydelse också visar ordets rektion och typiska kontexter. Dessutom skall de ange uppslagsordets utbredning och skriftliga ålder.

I ordbokens källförteckning ingår litet över tusen verk, av vilka ungefär hälften är ganska korta författningstexter. Utgångspunkten för vår ordbok är ordförrådet hos Mikael Agricola, som skapade det finska skriftspråket. En stor del av vårt material består av översättningar, och därför har originaltexterna på svenska, latin och tyska stor betydelse för ordboksarbetet. En excerptsamling på en halv miljon ord utgör tillsammans med korpusen i maskinläsbar form basen vid redigeringen. Korpusen innehåller i dag drygt 2 miljoner ord.

Ordboken över äldre finska är ett av forskningscentralens stora ordboksprojekt. När den är färdig kommer den att bestå av 6 delar och cirka 5.500 sidor. Hittills har två delar utkommit. De omfattar bokstavsintervallet A till K. De följande delarna planeras utkomma med fyra till fem års mellanrum.

Om korpusens innehåll

Grunden till korpusen över äldre finska lades 1992 då forskningscentralen anslöt sig till Internet och direktionen fattade ett principbeslut om att det skall skapas texter i maskinläsbar form. Då uppstod tanken att vi skulle utnyttja de perforeringsremсор som vi hade fått från förlagen i samband med att några viktiga verk på äldre finska hade satts med moderna bokstäver och getts ut under sina respektive jubileumsår. Verken var Mikael Agricolas samlade verk från åren 1543–1552, den första finska psalmboken av Jaakko Finno från slutet av 1500-talet, översättningen av kyrkolagen av år 1686 samt översättningen av Sveriges rikets lag från år 1734. Dessutom kunde forskningscentralen tack vare ett ekonomiskt bidrag få den första finska Bibeln från år 1642 i elektronisk form, så att den fanns tillgänglig 1992 då den finska Bibeln fyllde 450 år. Efter det här har korpusen utökats så att den i dag består av drygt 2 miljoner ord. Största delen av det nytillkomna materialet har renskrivits på dator, eftersom den ursprungliga texten är skriven i frakturstil. De delar av materialet som inte har varit skrivna i fraktur har vi kunnat skanna. En förteckning över det material som ingår i korpusen ges i ett appendix.

Korpusen kan delas upp i tre ämnesområden: religiös text, varav en del utgörs av psalmer, lagtext och annan profan text. Religiöst material utgör den största andelen av den äldre finskan och därmed också av korpusen. Vi har med de båda viktigaste andliga texterna, alltså Agricolas samlade verk och 1642 års Bibel. Vi har också försökt få med de viktigaste lagtexterna. Nu innehåller korpusen Herr Mårtens översättning av kung Kristofers landslag, som var den första lagtexten som översattes till finska i slutet av 1500-talet, kyrkolagen från år 1686 och Sveriges rikets lag från år 1734. Från verket *Acta Historica Fennica* har vi skannat en samling olika korta lagtexter. Vi hoppas att vi i år skall få in den första finska stadslagen från år 1601 och den andra översättningen av landslagen från år 1609. En samling författningstexter från 1600- och 1700-talet representerar olika livsområden, allt från brännvinsframställning till forsrensning, och eftersom det i allmänhet finns ganska litet världslig text från den här tiden, tänker vi utöka antalet författningstexter i korpusen. Den profana litteraturen

representeras också av 33 stycken almanackstexter från 1700-talet ända från det att almanackor på finska började utkomma i början av 1700-talet. Facklitteraturen representeras av Lizelius *Tietosanomat* som var den första dagstidningen på finska, samt två små böcker av Christfrid Ganander, som representerar dåtidens medicinska uppfattningar om människans och djurens sjukdomar och deras behandling. De två andra böckerna av Ganander består av finska gåtor och en samling fabler. Frosterus *Hyödyllinen Huwitus Luomisen töistä* är en naturvetenskaplig skrift om Guds skapelseverk.

Materialet är sparad i operativsystemet VMS med DEC:s 8-bitars multinationella teckensystem, som är nästan identiskt med ISO:s Latin-1-teckensystem. Forskningscentralen ingår i telekommunikationsnätet FUNET tillsammans med högskolorna och andra forskningscentraler. Genom FUNET är forskningscentralen kopplad till Internet, och via Internet har också utomstående personer tillgång till korpussen. Det kräver dock att man ansöker om en användarlicens av forskningscentralen. Sådana beviljas för forskningsändamål.

Arbetet med korpussen har till stor del utförts vid sidan av det egentliga ordboksarbetet och redaktionen har försökt välja ut de texter som varit mest centrala för redigeringen av ordboken. Vi har också försökt täcka en mängd olika ämnesområden för en så allsidig terminologi som möjligt. Vår strävan är att få med alla texter från 1500-talet, eftersom vi i ordboken alltid försöker ge det första belägget på ett ord.

Kodningen

För att korpussen skall vara så användarvänlig som möjligt har vi i de olika korpusmaterialen strävat efter att förenhetliga såväl formen som kodningen av materialet. Nästan alla verk som finns medtagna är löpande text. Det här innebär att en mening som avslutas med punkt eller liknande bildar en rad i korpussen. I början av varje rad finns en källangivelse där vi kodat in vilket verk det är fråga om. Koderna är desamma som i ordboken.

RWL1759-80-Rak-5:1 1 . § . Weräjä , josta kylän tie käypi , pitää caikilda nabureilda rakettaman ja ylöspidettämän .

RWL1759-80-Rak-5:1 Weräjän omalle pellolle ja niytulle mahta jocainen itze raketa ja sijtä waarin otta .

1Ms-1:0 I. Lucu . MAilman luomises teke Jumala Sanalla ja Hengellä I. Päiwänä / walkeuden .

1Ms-1:1 ALgusa loi Jumala Taiwan ja Maan .

1Ms-1:2 Ja Maa oli autia ja tyhjä / ja pimeys oli sywyden päällä / ja Jumalan Hengi lijckui weden päällä .

1Ms-1:00 Vers . 2. Jumalan Hengi) Silloin ei ollut wielä tuulda / sentähden tällä sanalla (Jumalan Hengi) ymmärretän pyhä Hengi .

JAFro1791-75 Korwalehdet ulkopuolella päätä kokoawat äänen , ja mutkaisen matkan kautta laskewat sen korwan sisälle .

Källangivelsen innehåller för det mesta namnet på författaren eller översättaren och verkets utgivningsår samt uppgift om sidnumret i originalkällan. För lagtexter anges också namnet på lagbalken, samt kapitel och paragraf. I källangivelserna för Bibeln anges bok, kapitel och vers, och dessutom anger vi också om det är fråga om text som inte ingår i verstexten utan t.ex. i förordet, i summarierna före verstexten (i exemplet 1Ms-1:0) eller i efterskriften (1Ms-1:00).

Det finns alltid ett mellanslag före och efter textordet, så att man så exakt som möjligt skall kunna söka ett visst textord. Det här gör att materialet också lämpar sig för morfologiska sökningar och dessutom kan tjäna annan språkforskning. Av samma orsak har alla skiljetecken avskilts från textorden med ett mellanslag. Ord som har delats på två rader i det tryckta materialet har i korpusen sammanfogats till ett odelat ord. Vi har kodat sidbyten mitt i en mening och vid behov har vi också kodat främmande ord, rubriker och vissa diakritiska tecken. Däremot har vi inte kodat typografin eller strukturen i den ursprungliga texten.

Mikael Agricolas samlade verk var vårt första korpusmaterial. Som grund för korpusen ligger en datafil som kom till då Agricolas verk utgavs i hålkortsform i slutet av 1970-talet. Materialet innehåller cirka 490.000 textord. Av tekniska orsaker var hela hålkortsmaterialet skrivet med versaler och följde det tryckta materialets sid- och radindelning. Här är ett exempel, början av psalm 23 ur Agricolas översättning, den som börjar "Herren är min herde, mig skall intet fattas".

DATA: [KORPUS.WANHASUOMI.AGRICOLA]AGRI3PS.TXT

324121 DAUIDIN PSALMI .

324124 HERRA ON MINUN PAIMENEN / EI MITEN MINULTA

324125 PWTU .
 324126 HEN CAITZEPI MINUA WIHERIEISES NITYS / IA
 324127 WIEPI MINUN WIRGOTTOUAN WEDHEN TYGE .
 324128 MINUN SIELUNI HEN WIRGOTTA / HEN WIEPI MINUN
 324129 OIKIALLE TIELLE / HENEN NIMENS TEDHEN .
 324130 IA WAICKA MINE WAELLAISIN PIMIES LAXOS / EM MINE
 324131 PELKEISI MITEKEN PAHUTTA / ETTES OLET MINUN CANSANI / SINUN
 324201 CAPULAS IA SAUWAS MINUN TUKEEUAT .

Det var väldigt besvärligt att läsa och hantera ett material som var skrivet i versaler, och som hade mycket korta rader och som dessutom hade en sexsiffrig källangivelse i början av varje rad. Det här gjorde att vi lät ett dataprogram ändra alla versaler till gemener, utom den första bokstaven i sådana ord som Gud, Jesus, Kristus, Amen o.s.v. Därefter måste nödvändiga ändringar från gemener till versaler göras manuellt genom att korpustexten jämförs med originalet, och det är både arbetsamt och tidskrävande. Dessutom har vi nu ändrat på radindelningen i texten. Vi ordboksredaktörer plockar ju ut de textavsnitt vi vill ha direkt ur korpusen, och vi har ingen nytta av att textens radindelning följer originalet, tvärtom. En kort rad ger inte tillräcklig kontext. Radindelningen har nu ändrats så att en mening i texten motsvaras av en rad i korpusen. I Agricolas bibelöversättningar motsvaras en vers av en rad i korpusen. På så vis omges textordet alltid av en hel mening, och inte av en textrad, som i allmänhet är kortare än en mening i det tryckta verket. En sådan här rad i korpusen kan gå över flera rader på skärmen eller i utskrift, men för datorn är den alltså bara en rad.

DATA:[KORPUS.WANHASUOMI.AGRICOLA]AGRI3PS.SEN

324121 Daidin Psalmi .
 324124 HERRA on minun Paimenen / ei miten minulta pwtu .
 324126 Hen caitzeppi minua wiherieises Nitys / ia wiepi minun wirgottouan Wedhen tyge .
 324128 Minun Sieluni hen wirgotta / Hen wiepi minun oikialle Tielle / henen Nimens tedhen .
 324130 Ja waicka mine waellaisin pimies Laxos / em mine pelkeisi miteken pahutta / Etttes olet minun cansani / Sinun <> Capulas ia Sauwas minun tukeeuat .

Bibeln från år 1642 hade också ursprungligen originalets radindelning. Även den har ändrats så att varje vers bildar en rad i datorn. Verket finns som helhet i korpusen, också förord och register. Och all text som

inte ingår i själva verstexten har kodats (summerier före verstext, efterskrift, rubriker och marginaltexter).

Arbetet med korpusen har lärt oss att lägga in tillräckligt med information i källangivelsen. Men det måste vara rätt sorts information. Vi har exempelvis märkt att radnumreringen för det mesta är överflödig. Däremot har källangivelsen och informationen i den en viktig funktion för jämförelser mellan texterna, t.ex. mellan olika översättningar av en och samma text. Vid redigeringen av ordboken jämför vi alltid Agricolas bibelöversättningar med Bibeln från 1642 och tvärtom. För närvarande är de här jämförelserna tidskrävande. Eftersom det inte finns versindelningar i Bibeln i Agricolas tryckta böcker (och därför inte i den nuvarande korpusen) kan samma ställe i de båda bibeltexterna inte lokaliseras automatiskt. Därför har vi nu börjat koda in verserna i korpusen i enlighet med 1642 års Bibel för att underlätta jämförelsearbetet. Den del av Agricola-korpusen som är färdigkodad ser ut så här:

DATA:[KORPUS.WANHASUOMI.AGRICOLA]AGRI3PS.SEN

A-III-241-[Ps-23:1] Daudin Psalmi . HERRA on minun Paimenen / ei miten minulta pwtu .

A-III-241-[Ps-23:2] Hen caitzepe minua wiherieises Nitys / ia wiepi minun wirgottouan Wedhen tyge .

A-III-241-[Ps-23:3] Minun Sieluni hen wirgotta / Hen wiepi minun oikialle Tielle / henen Nimens tedhen .

A-III-241-[Ps-23:4] Ja waicka mine waellaisin pimies Laxos / em mine pelkeisi miteken pahutta / Ettes olet minun cansani / Sinun <> Capulas ia Sauwas minun tukeeuat .

Det här innebär att vi med ett enda kommando får fram samma vers i 1642 års Bibel och i Agricolas verk. Uppgifterna om bibelbok, kapitel och versnummer står inom parentes, eftersom de inte anges i ordbooksartiklarna. Vi har också löst upp nummerkoden i början av varje rad, så att källangivelsen nu har samma form i korpusen som i ordboken. Vi kommer att föra in likadana koder också i Finno-korpusen för att underlätta jämförelser mellan Finnos psalmbokstexter och de två senare utkomna psalmböckerna från 1605 och 1701. I lagtexternas källangivelser har vi fört in uppgifter om både lagbalk, kapitel och

paragraf så att det är lättare att lokalisera ett textord. Den praktiska erfarenheten har alltså lärt oss att utforma korpusen enligt användarbehovet och att korpusinnehållet kan vara detsamma som i originalverken trots att formen är något helt annat. Det är syftet med en korpus som avgör hur den skall se ut.

För att underlätta de tidskrävande textjämförelserna har vi som målsättning att få in alla sådana texter i korpusen vars ordformer vi måste jämföra sinsemellan. För att reda ut ords etymologi vore det många gånger också idealiskt att ha med de svenska originaltexterna. Det gäller först och främst lagtexterna och den första svenska Bibeln.

Utnyttjandet av korpusen vid ordboksredigeringen

Sättet att utnyttja korpusen vid artikelredigeringen varierar antagligen något från redaktör till redaktör, men som regel tror jag att stommen till en artikel kommer till på basis av materialet i excerptsamlingen, eftersom den har en högre täckningsgrad både tids- och materialmässigt, medan korpusen främst utnyttjas för kompletteringar. Det händer dock att det kommer fram så mycket nytt material vid sökningarna i korpusen att artiklarna måste omstruktureras, t.ex. vad gäller indelningen i betydelsegrupper. Sökningarna i de större och centralare korpusdelarna görs alltid som egna körningar, eftersom de ger ett så rikt exempelmaterial. Om en sökning görs på flera olika ord grupperas sökresultatet ordvis, varefter en jämförelse görs med motsvarande ordboksartiklar, som vid behov kompletteras genom att exempel kopieras och klistras in. För det mesta görs sökningarna över ett visst alfabetavschnitt åt gången och inte på enskilda uppslagsord. I samband med körningarna kommer det ofta också fram nya ord, som inte återfinns i excerptsamlingarna. Dessa får självfallet egna artiklar i ordboken.

Sökprogrammen *search*, *agrep* och *kwic*

I de sökprogram vi använder riktar sig sökningen alltid till en rad i datorn. Man kan söka en eller flera sinsemellan olika teckensträngar. Det går också att söka ordpar, och just för de sökningarna är det väsentligt att korpusraderna är tillräckligt långa. Man kan söka samtidigt i flera olika filer.

Det sökkommando som lämpar sig bäst för vårt arbete heter *search*. Det är ett kommando som ingår i operativsystemet. I sin enklaste form kan kommandot se ut såhär:

search fil(er) teckensträng

det vill säga "sök en viss teckensträng i en eller flera filer". Om datorn hittar teckensträngen visar den träffarna på skärmen tillsammans med kontexten i den ordning de förekommer i korpusen. Man kan lägga till preciseringar för att exempelvis förlänga kontexten kring textordet eller samla statistiska uppgifter.

Ett annat sökkommando som kan användas är *agrep* som hör till UNIX-systemets program av *grep*-typ. *Agrep* används i synnerhet för approximativa sökningar. Det godkänner regelbundna uttryck som sökargument. Med hjälp av så kallade metatecken kan man bilda mer komplicerade söksträngar enligt vissa regler. Man kan meddela hur stor avvikelser får vara från den teckensträng man söker. Man kan också använda optionen *best-match* och då letar algoritmen efter den mest lämpliga träffen och meddelar hur stor avvikelser är. För en användare som inte känner gammal finsk ortografi är *agrep*-kommandot mer användbart än *search*-kommandot.

När man bläddrar i ett stort material är *kwic* ett bra hjälpprogram. Vårt *kwic*-program godkänner också regelbundna uttryck som sökargument. Det listar beläggen i alfabetisk ordning under varandra på skärmen, vilket gör det möjligt att komma åt den äldre finskans mycket varierande skrivformer av ett och samma ord. En av *kwic*-programmets förtjänster är att man kan lägga till en sekundär söknyckel så att beläggraderna i andra hand listas enligt ordet före eller efter själva belägget. På så vis får man lätt fram sammansatta ord som mycket ofta är särskrivna i äldre finska. Exemplet nedan visar en sökning på ordet *tauti* 'sjukdom' med dess böjningsformer där ordet före sökordet är sekundär söknyckel.

```
kwic -r -s-1 "tau[t|d][i|e]" kotiapteekki.sen
```

```
[Gan1788a-40] tu ottaa pois hammas taudin , hauteexi ajoxijn äkämiijn
```

```
[Gan1788a-40] aan , parantaa korwa taudin .
```

```
[Gan1788a-26] 38. Recepti Rewäsin Taudin öljyyn. ( Moder droppar )
```

```
[Gan1788a-8] 16. Jalka taudisa ( podagra ) uloswetää se
```

```
[Gan1788a-23] n alusa , lankeewasa taudisa ; Kelta- jäsenten- ja [Gan1788a-19]
```

```
hewoisille loukkaus taudisa , lusikallinen sisälle
```

```
[Gan1788a-26] 100 tippua , Rewäsin taudisa , raskaisa lapsen
```

```
[Gan1788a-8] n päänskipuin , wattan tauteihin , wiemistymisiin , ja
```

```
[Gan1788a-8] jn , selkä- ja lonkka tautein , hammas tautijn ja pahan
```

```
[Gan1788a-24] hyvä keuhko ja rinta tautein ja raskampaan ja helpompään
```

```
[Gan1788a-18] ain ja saastuttawain Tautein lijkkeisä , warjeelee , ja
```


En annan fördel med programmet är att man får fram också sådana sökord där den sökta teckensträngen finns mitt inne i ordet. Tack vare det kan man söka ihopskrivna sammansättningar på efterledsstammen.

```
kwic -r "tau[t|d]" kotiapteekki.sen
```

```
[Gan1788a-40] ain pureman päälle , jalkatautiijn ( podagraan ) ja
[Gan1788a-7] 12. Punataudisa , yxi plaastari lappu
[Gan1788a-46] 0 pisarta pureisa ja punataudisa , Hammas-taudisa
[Gan1788a-18] ta tulleisa Räkä- ja Päätaudeisa woidellaan päälaki .
[Gan1788a-7] 13. Ruusutaudisa päälle pannaan kahden
[Gan1788a-21] , weren sylkemys ja täwytauti .
[Gan1788a-34] uin woiman kuluttawa ulkotaluti on seurasa .
[Gan1788a-23] sten kansa tullesa ; Wesitautien alusa , lankeewasa
[Gan1788a-18] 3. Wilutaudeisa sisälle otetaan 1 eli
[Gan1788a-40] auttaa 3:nen päiwän wilutaudisa , Lukinwerkko haawan
[Gan1788a-22] ia , perkaa wattan , wilutaudisa hywä , ja ummella
```

Denna funktion möjliggör också morfologiska sökningar.

```
kwic -r "ium" kirkkolaki.sen
```

```
[Flor1688-126-XXV:1] ispanis canssa Auditoriumin / eli sijhen paikkaan
[Flor1688-146-IV:0] / nautita sitä beneficiumita / cuin täsä processis
[Flor1688-104-XXII:2] cuin Upsalan Conciliumis , Anno 1593 . caikilda
[Flor1688-1-I:1] Upsalan Conciliumis wastanotettu / ja
[Flor1688-Aa4b] / ilman Pispan ja Consistorialium saldimat .
[Flor1688-144-I:0] Nijsä Consistoriumeis joisa
[Flor1688-123-XXIV:29] Pispalda ja Consistoriumilda anoman Pappia
[Flor1688-91-XIX:22] n Pispalda ja Consistoriumilda täydel woimalla
[Flor1688-11-II:5] itä Epistola ja Evangeliumi weisattaman Altarin edes
[Flor1688-150-XI:0] ja hänen pyhän Evangeliumins päälle / että minä
[Flor1688-81-XIX:6] / Schouluis / Gymnasiumeis eli Academiois / jolla
[Flor1688-144-I:0] Meidän Hof-Consistoriumis on Meidän
[Flor1688-88-XIX:17] ten ja Caluin Inventarium ynnä Kircon Kirjain
[Flor1688-Bb4b] lla on erinomainen Privilegium Papin walitzemises /
[Flor1688-85-XIX:11] n erinomaisen Privilegiumin Papin walitzemises /
```

Avslutning

Jag har gjort vissa uträkningar över hur stor del av våra ordboks-exempel som i dag kan tas direkt ur korpusen. I de ordboksartiklar jag baserade uträkningen på fanns det 25 till 115 exempel per artikel, och jag kom fram till att 32 till 50 procent av ordboksexemplen fanns i korpusen. Extremvärdena var 24% och 64%. Jag tror att de här siffrorna är representativa för exemplen som helhet, eftersom vår korpus täcker de centralaste och samtidigt största verken över äldre finska. Det här innebär att ungefär hälften av exemplen som kommer att ingå i ordboken inte behöver skrivas av ordboksredaktörerna. Dessutom slipper vi granskningen. Vi kan också vara säkra på att få med alla ord i de centralaste verken. Det här hör till fördelarna med korpusarbetet. Det negativa med det är de stora textmängder som är resultatet av vissa sökningar och som redaktörerna blir tvungna att gå igenom manuellt. Men tack vare korpusen kan vi i alla fall vara förvissade om att vi i vår ordbok får med ett heltäckande och representativt exempelmaterial över det centralaste ordförrådet i äldre finska.

Appendix: korpus över äldre finska

Agricolas verk I–III (1543–1555) (492 000 ord)

Jaakko Finnos Psalmbok c. 1583 (21 000)

Piae Cantiones 1616 (14 350)

Bibeln 1642 (c. 1 milj. ord)

Likpredikningar (1600- och 1700-talet) (176 400)

Herr Mårtens landslag (1500-talet) (52 400)

Kircko-Laki ja Ordningi 1688 'Kyrckio-Lag och Ordning' (42 800)

Ruotzin Waldacunnan Laki 1759 'Sveriges Rikes Lag' (102 400)

Acta Historica Fennica (1500- och 1600-talet) (22 900)

Författningstexter 1600- och 1700-talet (44 500)

Almanackstexter från 1700-talet (53 000)

Lizelius:

Suomenkieliset Tieto-Sanommat 1775–1776 'Finskt Tidningsblad' (24 800)

Ganander:

Suomalaiset Arwotuxet, Wastausten kanssa 1783 'Finska gåtor med svar' (4 450)

Uudempia Uloswalituita Satuja 1784 'Nya utvalda fabler' (28 200)
Maan-Miehen Huone- ja Koti-Aptheeki 1788 'Landtmannens hus-
apotek' (8 300)
Eläinden Tauti-Kirja 1788 'Om husdjurens sjukdomar' (13 300)
Frosterus:
Hyödyllinen Huwitus Luomisen Töistä 1791 'Nyttig förströelse om
Guds skapelseverk' (23 400)
Varia: tryckta verk från 1700-talet (5 500)

(Övers. från finska av Pia Westerberg, bearbetning av Nina Martola)

Jette Pedersen

Sammensætning af et tekstkorpus til en bilingval teknisk fagordbog og forslag til metodiske lemmaselektionskriterier

On the basis of the inherent systematic structure of the subject field(s) in question, the criteria are laid down for the selection of texts for a monofunctional text corpus with a view to a bilingual technical dictionary. Besides, a number of recommendations are made concerning the identification of and selection criteria for associated subject fields and relevant aspects. The need for unambiguous, consistent selection criteria is emphasised, not only in connection with text selection, but also as regards the individual dictionary information categories, in the article exemplified by lemmata. Whereas the selection of certain types of lemmata which must be considered necessary for a technical dictionary to fulfil a number of encoding functions, including text production and L1–L2 translation, can only be made on the basis of a text corpus, the identification of actual LSP terms requires the participation of an expert. In either case, selection presupposes that the necessary selection criteria be established in advance.

1. Empirisk basis

At producere en fagordbog kan karakteriseres som én lang selektionsproces. For at færrest mulige valg baseres på ad hoc afgørelser og tilfældigheder, må der som et led i de prækonceptionelle overvejelser fastlægges et sæt entydige og konsekvente selektionskriterier for samtlige ordbogens informationskategorier.

Blandt fagleksikografens første valg er at afgøre, hvorfra den omtalte selektion skal foretages, eller sagt med andre ord: Hvad skal være ordbogens empiriske basis? Ser man på den fagleksikografiske praksis, finder man, at omtæksten til tekniske fagordbøger sædvanligvis kun i meget vage vendinger oplyser om det materiale eller de forudsætninger, der har ligget til grund for ordbogen, jvf. fx følgende uddrag fra forteksten til Clausen da-eng hhv Chambers: "Det indsamlede materiale findes i en database" og "The software was a library database program called Inmagic which allowed all the material including additions to be indexed and" I begge citater henvises til et tilgrundliggende materiale, men uden at der redegøres nærmere for arten heraf. I andre tilfælde henvises til anonyme personer, der har ydet bistand ved ordbogens tilblivelse, jvf. fx forordet til Dictionary of Computing: "We

would like to thank many people who helped with this dictionary, by reading the text and making useful comments. The cover design is by Peter Cartwright." (sic!). I atter andre tilfælde, således fx Warrern og Skjerk Bilteknik, gives der ingen oplysninger om den empiriske basis overhovedet. Samtidig med at denne uklarhed mht eksisterende tekniske fagordbøgers empiriske basis forekommer at være reglen snarere end undtagelsen, findes der dog også eksempler på en mere reflekteret og informativ praksis. Således indeholder forteksten til Genteknologisk Ordbog en liste over den litteratur, der er blevet benyttet ved udarbejdelsen af ordbogen, mens der i forteksten til Kuçera eng-ty findes en fortegnelse over en række navngivne fageksperter, der har medvirket ved ordbogens tilblivelse, med angivelse af de nævnte personers arbejdsplads samt hvilke fagområder, de hver især repræsenterer.

Der kan på basis af heraf umiddelbart identificeres to muligheder for empirisk basis for tekniske fagordbøger: For det første det i Clausen refererede og i Genteknologisk Ordbog nærmere specificerede materiale i form af **tekstsamlinger**. For det andet kan der inddrages en eller flere **fageksperter**, således som det har været tilfældet i Kuçera, og hvortil vel også "many people" i Dictionary of Computing refererer. Herudover kan man forestille sig, at fagleksikografen har forladt sig på sin egen faglige og sproglige kompetence, dvs. **introspektion**. Endelig kan leksikografen have støttet sig til **belægsamlinger** eller **eksisterende beskrivelser**, herunder allerede foreliggende ordbøger og leksika eller hånd- og fagbøger. Disse muligheder er der dog ingen blandt de 25 undersøgte tekniske fagordbøger, der vedkender sig.

Det forekommer næppe realistisk at antage, at en teknisk fagordbog, hvis intenderede brugere uundgåeligt må omfatte tekstproducenter og oversættere i erhvervslivet, skulle blive til uden hensyntagen til de tekster, der faktisk produceres her. I modsat fald ville det fx næppe være muligt at identificere de forskellige problemtyper, som ordbogen skal medvirke til at løse. Specielt når det drejer sig om tekniske del- og enkeltfagsordbøger kommer hertil, at disse erfaringsmæssigt i stigende omfang allerede i deres udgangspunkt har en mere eller mindre eksplicit tilknytning til en konkret branche, virksomhed, organisation el. lign. – om ikke andet så af økonomiske årsager: De er dyre at producere, og fagordbøger nyder som bekendt ikke offentlig støtte. Det følger derfor, at tekster herfra må indgå som en nødvendig del af den pågældende ordbogs empiriske basis.

I det omfang de omhandler det aktuelle fagområde må alle tekster fra den pågældende branche, virksomhed m.v. betegnes som tematisk relevante. Af hensyn til oplysningernes pålidelighed bør der dog an-

lægges en teksttypologisk indfaldsvinkel, idet selektion bør begrænses til terminologisk og indholdsmæssigt præcist tekstmateriale. Dette indebærer, at reklamer i form af prospekter o.l. udelades, idet denne teksttypes primære formål netop ikke er at oplyse, men derimod at overtale. Den bevidste nedprioritering af det faglige indhold, som ligger i naturlig forlængelse af reklamers direkte kommunikationsfunktion, bevirker, at pålideligheden af denne teksttype må anses for usikker. Af samme grund udelades ligeledes tekstmateriale, der må betegnes som terminologisk tentativt. Der tænkes her specielt på de grundlæggende tekniske specifikationer, som udarbejdes på et meget tidligt stadium af produktudviklingsfasen, nemlig i forbindelse med forskning og udvikling. Som eksempel på teksttyper, der er typiske kandidater til selektion, kan nævnes de forskellige former for produktinformation, monterings-, betjenings-, og driftsvejledninger, standarder og servicemanualer.

Tekniske fagordbøger med en eksplicit tilknytning til en konkret branche eller virksomhed adskiller sig især fra rene del- og enkeltfagsordbøger ved at have en mere kompleks indgang til fagområdet. Ud over et prioriteret del- eller enkeltfag vil der typisk være behov for at medtage informationer fra andre fagområder, der har tilknytning hertil, fx ved at være omfattet af produktprogrammet. I forbindelse med et konkret virksomhedsspecifikt pumpeteknologisk ordbogsprojekt er der eksempelvis identificeret to sådanne associerede områder, idet den pågældende pumpeproducent fremstiller sine egne elmotorer og elektroniske styringer. Til forskel fra pumper produceres de pågældende elmotorer og elektroniske styringer imidlertid ikke som et mål i sig selv, men derimod alene med henblik på at indgå i det prioriterede fags produkter. Dette afspejler sig i virksomhedens tekstproduktion, hvor langt den største del af de informationer, der gives om elmotorer og elektronik, skal findes i litteratur om det prioriterede fagområde. Samtidig kan associerede fagområder give anledning til et vist omfang af selvstændig tekstproduktion. For at sikre at ordbogen vil kunne medvirke til at løse produktions- og receptionsproblemer vedr. hele produktprogrammet, brancheområdet m.v. må også disse tekster medtages. Men i modsætning til hvad der er tilfældet med ordbogens prioriterede fagområde, intenderes der for associerede områder ikke systematisk tematisk dækning, jvf. nedenfor.

Langt størstedelen af tilgængelig virksomheds- hhv. brancheintern litteratur på L2 vil typisk udgøres af relætekster, dvs. "tekster, der enten i sig selv er oversættelser, som kan være lavet af folk, der har målsproget eller kildesproget som modersmål, eller være skrevet som originaltekster af folk, der ikke har målsproget som modersmål."

(Krogh-Hansen 1994:123). De kan dermed ikke frasorteres, men må i korpus gøres identificerbare som sådanne, fx vha en særlig kodering eller ved at samles i et særligt delkorpus. For at sikre at korpus kan levere pålidelige sproglige, herunder ikke mindst grammatiske og kollokationelle, oplysninger på L2, er der altså behov for et betydeligt bredere tekstfundament end det materiale, branchen hhv virksomheden kan levere. Dette behov understøttes yderligere af, at der ved konciperingen af tekniske fagordbøger ofte forudses en bredere brugergruppe end den konkrete branche eller virksomhed.

Et muligt udgangspunkt for opstilling af bibliografi over offentligt tilgængelig litteratur omhandlende det aktuelle fagområde kan eksempelvis være bibliotekernes universelle decimalklassifikation, UDK-systemet, eller eventuelt allerede eksisterende specialbibliografier. I forbindelse med udarbejdelse af bibliografi til den danske del af et dansk-engelsk-fransk tekstkorpus inden for aftaleretten har man således taget udgangspunkt både i UDK-systemet og i en foreliggende emneinddelt dansk juridisk bibliografi (Dyrberg et al. 1991). Hvor der ikke findes sådanne bibliografier, eller hvor UDK-systemet viser sig at være uegnet, fx ved ikke at være tilstrækkeligt finmasket eller ved at være for upræcis i sin indfaldsvinkel, kan det fornødne overblik over grundmængden af tekster etableres vha on-line søgning på overordnede søgetermer i relevante fagbibliotekers databaser. I forbindelse med det her eksemplificerede fagområde er der således foretaget trunkeret on-line søgning i de tekniske bibliotekers databaser på alle sammensætninger med */pump/* som søgeterm.

Før selektion af tekster fra den opstillede bibliografi over de fremkomne titler kan påbegyndes, må der søges etableret et systematisk overblik over de(t) pågældende fagområde(r). Formålet hermed er dels at sikre indfangning af tematisk relevant tekstmateriale, dels at afgrænse det pågældende fagområde i forhold til tilgrænsende områder og dermed muliggøre frasortering af fagområde-eksternt materiale. Faget, således som det afspejler sig i opbygningen af og dermed i indholdsfortegnelsen til de fremkomne titler, kan her tjene som udgangspunkt for udarbejdelse af en egentlig fagsystematik, men det understreges, at den endelige udformning må ske i samråd med en fag-ekspert. Ved den foreløbige udformning koncentrerer man sig i første omgang om de mere generelle værker. Er ordbogens emneområde hhv ét af ordbogens emneområder som i det aktuelle tilfælde pumpe-teknologi, vil det ikke levere de fornødne oplysninger at søge hjælp i værker som *Centrifugal and Axial Flow Pumps*, der, som det fremgår af titlen, alene behandler et eller flere delfagsområder inden for enkeltfaget og derfor ikke, eller i hvert fald kun utilstrækkeligt, kan

antages at foretage den nødvendige afgrænsning i forhold til tilgrænsende områder. Dette gælder ikke blot, hvor der er behov for at afdække et fagområde i sin helhed, fx hvor der er tale om en flerfagsordbog, der blandt sine emneområder medtager et bestemt enkeltfag, eller hvor ordbogen udelukkende intenderer at dække det pågældende enkeltfag. Det gælder tillige for delfagsordbøger, også selv om det pågældende delfag netop er det, der behandles i den pågældende titel. Med henblik på selektion af tekstmateriale til en delfagsordbog er der altså i lige så høj grad som ved enkelt- og flerfagsordbøger behov for at etablere et overblik over et helt enkeltfag, idet det ellers ikke vil være muligt med tilstrækkelig stor sikkerhed at sortere delfags-eksterne værker eller dele af værker fra, samtidig med at det sikres, at det pågældende delfag dækkes og afdækkes i hele dets udstrækning. På dette stadium af arbejdsprocessen må man altså som sagt søge hjælp i bredere anlagte værker. På den anden side må der ikke være tale om for brede værker, idet bredden oftest vil være på bekostning af dybden. Værker, der indeholder en samlet behandling af et enkeltfag, må erfaringsmæssigt anses for de mest velegnede.

For at sikre at den empiriske basis også er til stede for den leksikografiske behandling af de forskellige aspekter, der falder ind under det pågældende fagområde, kan det anbefales at udarbejde en aspektfortegnelse. Potentielle aspekter kan fx identificeres med udgangspunkt i de relevante titlers indeks. I det konkrete eksempel med pumper fremkom herved så forskelligartede aspekter som funktion /arbejdsprincip; anvendelsesmuligheder, fx afvanding, vanding og varmeanlæg; komponenter; materialer, herunder bearbejdning og sammenføjning af materialer; korrosion; grundlæggende elteknik; elektrisk montage; sensorer; pumpemedier; væskestrømning; akseltætninger; koblinger; ventiler og armaturer; rørledninger; måleteknik; installation; vedligeholdelse; normer; diverse pumpebegreber, som fx trykstød i rørledninger, kavitation, kapacitet, løftehøjde og pumpetryk. I bi- og polylingval sammenhæng må de fagområderelevante dele af fagsystematikken samt de forskellige elementer i aspektfortegnelsen være repræsenteret i begge hhv alle delkorpora.

Specielt i forbindelse med selektion af offentligt tilgængeligt tekstmateriale kan det desuden blive nødvendigt at anlægge et aktualitetskriterium. Således at den empiriske basis også kommer til at omfatte den seneste udvikling på området, medtages ud over fag-, hånd- og lærebøger også tematisk relevante tidsskriftartikler, idet udgivelsestidspunktet for de øvrige værker vil være afgørende for, hvor langt det må anses for påkrævet at gå tilbage.

I forbindelse med oprettelsen af et dansk og et tysk genteknologisk tekstkorpus er der tilstræbt kvantitativt ligelig repræsentation af tekstord inden for hver af de fire kommunikationsretninger fagekspert -> fagekspert / fagekspert -> lægmand / lægmand -> fagekspert / lægmand -> lægmand (Lauridsen et al. 1989:10). Samtidig har man intenderet et polyfunktionalt tekstkorpus, idet korpus er sammensat på en sådan måde, at det skal kunne tjene som udgangspunkt for "diskursanalyser, såvel som grammatisk-syntaktiske, leksikografiske samt terminologiske undersøgelser" (Lauridsen et al. 1989:4). Repræsentation af tekster med lægfolk på afsendersiden har efterfølgende vist sig problematisk i forbindelse med den leksikografiske anvendelse af det danske genteknologiske korpus, nemlig som empirisk basis for udarbejdelsen af Genteknologisk Ordbog, idet den fagkompetente del af det leksikografiske team har konstateret, at denne teksttype indeholdt en stor risiko for fejlinformation. Det kan ud fra denne erfaring konkluderes, at hvor et korpus skal tjene som værktøj i en faglig sammenhæng, må det også opbygges på et korrekt fagligt grundlag (Kaufmann 1992:67). I modsat fald vil tekster af typen lægmand -> fagekspert og lægmand -> lægmand, kræve en omfattende selektiv indsats, hvis korpus også, således som det var intenderet, skal kunne anvendes til leksikografiske formål. "Med mindre en ekspert (molekylærbiolog) tages med på råd ved en sådan nødvendig frasortering af materiale, der i faglig henseende er mindre lødigt, vil der kunne drages mange forkerte og modstridende konklusioner ud fra det foreliggende materiale." (Stummann 1992:70-71).

Ovenstående tematiske og teksttypologiske overvejelser resulterede for det her eksemplificerede fagområde, inklusive aspekter og associerede områder, i et korpus i størrelsesordenen 1.390.000 løbende tekstord, fordelt med 900.000 ord på L2 og 490.000 på L1. Den skæve fordeling mellem L1 og L2 skyldes dels, at anvendeligheden af langt størstedelen af de virksomhedsinterne L2 tekster er begrænset, ikke mindst i forbindelse med ækvivalentkontrol og selektion af sproglige oplysninger, idet der er tale om relætekster, dels at tekstproduktion på og oversættelse til modersmålet ikke er intenderede brugsfunktioner.

2. Lemmasektion

For at afdække hvilke lemmaselektionskriterier, der har været anvendt i eksisterende tekniske ordbøger, har jeg undersøgt omteksterne til 25 mono- og bilingvale tekniske fagordbøger. Heraf giver tolv ingen oplysninger overhovedet, mens lemmasektion i de resterende tretten

kun omtales i meget løse vendinger. I en tilsvarende undersøgelse af ni juridiske fagordbøger når Nielsen 1989:59 frem til et lignende resultat: "Det er karakteristisk for alle de undersøgte ordbøger, at de enten ingen oplysninger indeholder om selektionen af lemmata eller, at de kun indeholder korte, vage oplysninger herom." De sparsomme oplysninger står i stærk kontrast til den interesse, emnet vises i ordbogsanmeldelser, jvf. følgende uddrag af anmeldelser af fagordbøger i tidsskriftet *Hermes*, hvor de to første omhandler flerfagsordbøger, mens de sidste fire er anmeldelser af enkeltfagsordbøger:

Møller (1990:148): "Selv om en række centrale termer er medtaget, mangler mindst lige så mange, ikke mindst af moderne karakter."

Riiber/Thomsen (1991:141): "Den konklusion, man må drage på grundlag af disse og andre tilsvarende stikprøver, må være, at DTTO nok har taget munden lidt for fuld ved at påstå, at 'det centrale ordforråd' er medtaget. Dertil er der for store mangler i systematikken. Når hullerne i systematikken snarere synes at være reglen end undtagelsen, kunne man let få den tanke, at udvælgelsen af termer er noget tilfældig, eller måske styret af et hyppighedskriterium, hvilket ikke bør være tilfældet for en fagordbog."

Duvå (1991:148 og 152): "Hvad laver ganske almindelige ord som fx ... *fedte for, mulighed, på grund af* i en ordbog for speciel fagterminologi? ... Er de nyeste fagudtryk medtaget? Forgæves leder jeg efter: ... *børsetik, deregulering* Den manglende lemmatisering gælder dog ikke kun nye udtryk, men desværre gode, gamle, velkendte fagudtryk: *bygge- og anlægsarbejde* ...".

Laursen/Tarp (1991:156): "Men hvorfor kan man f.eks. finde ord som *afladedokumenter, eksport* ..., når helt relevante og hyppigt brugte ord som *anløbshavn, bortfragter, fortolde* ... ikke er medtaget?"

Duvå/Nielsen (1992:117): "Vi kan finde termer, der er relateret til organisation, fx *stabschef* ..., men hvorfor ikke termer som fx *åben-dørpolitik, quality management* ..., som er nogle af nøgleterminerne indenfor moderne management og ledelsesstrategi? Til gengæld kan vi finde adskillige termer inden for regnskabsområdet ... Det kan være svært at se, hvorfor de hører hjemme i en marketingordbog."

Barlach et al. (1992:196): "I denne sammenhæng skal opremses følgende manglende udtryk i de to stikordslister: *økonomisk kriminalitet, criminal code, criminal law* ...".

Ser man på de tretten tekniske ordbøger, hvis omtækst kommer ind på lemmaselektion, omend det altså kun er summarisk, hedder det eksempelvis i *Elektroteknisk Ordbog*: "... indeholder i alfabetisk orden samtlige de i følgende danske standarder indeholdte termer: ..." og *Webster's New World Computer Ordbog*: "Formålet med denne bog har

været at samle så mange som muligt af disse [= computer-terminologiske] ord." I modsætning hertil intenderes i Skibsted en mindre dybtgående dækning: "Bogen omfatter den væsentligste terminologi inden for M, E, P og B retningerne." Det samme er tilfældet med rækken af bilingvale tekniske flerfagsordbøger fra Clausen: "Der er ved udvælgelsen af opslagsordene lagt vægt på at medtage det centrale ordforråd inden for de behandlede emneområder." og L&H: "[Ordbogen] repræsenterer et systematiseret ekstrakt af en translatør-udviklet database." Til sammenligning hedder det i McGraw-Hill: "... the vocabulary of science and technology was surveyed by a select group of consulting editors. The terms ... included in the Dictionary are those considered important in each consultant's particular area of specialisation."

Om den tilgrundliggende empiriske basis for selektion af lemmata kan det af ovenstående citater fra tekniske fagordbøger udledes, at denne har bestået dels af tekstsamlinger (Clausen og L&H), dels af introspektion, her i form af en fagkompetent vurdering (McGraw-Hill). Bortset fra belægsamlinger, som i denne sammenhæng må anses for uegnede, eksisterer for selektion af lemmata principielt de samme muligheder for empirisk basis som for selektion af ordbogens øvrige informationskategorier. Specielt for ordbøger med tilknytning til en konkret branche eller virksomhed gælder det, at selektion må tage særlige hensyn til de tekster, der produceres i den pågældende branche eller virksomhed. Når det drejer sig om del- og enkeltfagsordbøger, vil eksisterende fagordbøger alene kunne have en supplerende funktion, og kun i det omfang, der allerede måtte foreligge en del- eller enkeltfagsordbog for det pågældende fagområde, idet flerfagsordbøger ikke kan anses for tilstrækkeligt dybtgående til at sikre dækning af et fags termer.

Yderligere kan der på basis af citaterne identificeres to grundholdninger med hensyn til dækningen af et konkret fagområde, således som den afspejler sig i ordbogens lemmabestand. Det drejer sig om maksimerende lemmaselektion, som fx i Elektroteknisk Ordbog og Webster's New World Computer Ordbog, henholdsvis minimerende lemmaselektion, jvf. fx Skibsted og Clausen. En maksimerende ordbog er altså en ordbog, der intenderer at dække den størst mulige del af et bestemt fags ordforråd. Der er her tale om et yderst arbejds- og pladskrævende ordbogsprojekt, og som det var tilfældet med de to ovenfor citerede ordbøger, der i henhold til deres omtækt intenderede maksimerende lemmaselektion, vil maksimerende ordbøger da også typisk begrænse sig til et enkelt fagområde. Når bortses fra enkelte, meget omfattende flerbindsværker, er flerfagsordbøger i langt de fleste til-

fælde minimerende, idet de kun intenderer at medtage det centrale ordforråd inden for hvert fagområde, jvf. ovenstående citater fra omteksterne til Skibsted, Clausen og McGraw-Hill.

Som det fremgik af ovenstående uddrag fra fagordbogsanmeldelser, kan en ordbogs begrænsning til et enkelt fagområde omvendt ikke ses som nogen garanti for, at den pågældende ordbog også er maksimerende, dvs. dækkende i forhold til fagets ordforråd. Samtidig konstateredes i samtlige anmeldelser en række eksempler på lemmata, som måtte anses for irrelevante for det pågældende fagområde. Endelig fandtes der eksempler på, at det ikke var lykkedes minimerende fagordbøger at identificere det centrale ordforråd, således som det fx var intenderet i flerfagsordbogen fra *Clausen*, hvor anmelderne antyder, at lemmaselektion har været styret af et frekvenskriterium. Med et konkordansprogram lagt ind over et maskinlæsbart tekstkorpus er det en simpel sag at generere et alfabetisk indeks over samtlige de i tekstkorpus forekommende ortografiske ord med angivelse af den frekvens, hvormed de forekommer i korpus. Med udgangspunkt i en sådan alfabetisk frekvensliste kan der herefter anlægges et rent objektivt baseret frekvenskriterium for selektion af lemmata. I den polylingvale Grundwortschatz Naturwissenschaften, der intenderer at dække fagområderne kemi, matematik og fysik, har man koblet frekvens med et tekstspredningskriterium, hvilket i ordbogens engelsk-tyske del har resulteret i følgende komplette liste over lemmata under bogstavet *w*: *wall, want, wash, water, wave, wavelength, way, we, weak, weigh, weight, well, well-known, what, whatever, wheel, when, whence, whenever, where, whereas, whereby, whether, which, whichever, while, white, who, whole, whose, wide, widely, width, will, wire, wish, with, withdraw, within, without, word, work, worker, world, worth, would*. Det er her om muligt næsten mere end umuligt at forestille sig, hvilke brugere der i forbindelse med hvilke brugssituationer vil kunne drage nytte af denne type lemmata – de intenderede fagområder taget i betragtning.

En løsning på de ovenfor eksemplificerede lemmatiseringsproblemer forudsætter en teoretisk afklaring af de for lemmaselektion tilgrundliggende kriterier. Dette indebærer konkret, at der som første skridt opstilles et sæt kriterier for identifikation af det prioriterede fagområdes termer. Dette gælder, uanset om der er tale om en maksimerende del- eller enkeltfagsordbog eller en minimerende flerfagsordbog. I førstnævnte tilfælde vil nemlig alle fagtermer skulle identificeres med henblik på selektion, mens det i sidstnævnte tilfælde ikke vil være muligt at selekttere det centrale ordforråd, medmindre man først har etableret et overblik over hele ordforrådet, jvf. fx diskussionen af

Clausen ovenfor. Der foreslås med henblik herpå udarbejdet systematiserede oversigter over det pågældende områdes fagterminologi. Bortset fra at dette arbejde må involvere et tæt samarbejde mellem leksikograf og fagekspert (hvor der ikke er personsammenfald mellem de to), er det næppe muligt at opstille et sæt universelle retningslinier for en terminologisk strukturering af samtlige teknikkens emneområder. Der kan hentes inspiration i lære- og andre fagbøger samt i virksomhedsinternt materiale, hvor tekster af typen eksploderede billeder og snitmodeller med ledsagende tekst må anses for særligt velegnede, når det drejer sig om etablering af et overblik. På baggrund heraf foreslås der udarbejdet såkaldte termsystematikker med udgangspunkt i de forskellige produkttyper, som igen kan tilordnes en systematiseret, funktionelt baseret fagsystematik (Pedersen 1995a:83–90 og Pedersen 1994:203–227).

Mens en fagordbog, i hvilken lemmabestanden er begrænset til rene fagtermer, ofte vil være tilstrækkelig til receptionsformål, anses en sådan rent terminologisk ordbog ikke at kunne yde den fornødne hjælp i forbindelse med L2 tekstproduktion og L1-L2 oversættelse. I modsætning til hvad der gør sig gældende for terminologiske ordbøger, forudses der altså hvor det drejer sig om andre fagordbøger end rene receptionsordbøger at være behov for en betydeligt bredere lemmaforståelse end den, der begrænser sig til egentlige fagtermer. Branche- og virksomhedsspecifikke tekniske fagordbøger medtager som nævnt også emneområder, som ligger uden for det prioriterede fagområde, men som fx er karakteriseret ved at være omfattet af produktprogrammet. Dette indebærer, at der også må selekteres fagudtryk herfra i det omfang, disse indgår i branchens hhv virksomhedens tekstproduktion. I modsætning til hvad der er tilfældet med det prioriterede fagområdes termer er der altså for associerede områder ikke behov for systematiseret terminologisk dækning. Yderligere kan der være behov for at selektere fagudtryk fra en række mere perifere emneområder, som ikke direkte er omfattet af produktprogrammet, men fx relaterer sig til de for fremstillingen heraf tilgrundliggende materialer, processer m.v. I lighed med hvad der var tilfældet for associerede områder, vil der også for perifere fagområder alene være behov for tekstuel betinget selektion. Det forudsættes desuden, at de pågældende udtryk optræder i de tilgrundliggende tekster med en vis fastlagt minimumsfrekvens og tekstspredning. Tallene herfor ansættes i forhold til den samlede størrelse af tekstkorpus. Der er dermed for associerede hhv perifere områder tale om varierende grader af minimerende, tekstbaseret lemmaselektion, idet der for selektion af fagudtryk fra perifere områder

er indlagt yderligere to kriterier. De ovenfor foreslåede selektionskriterier kan herefter opsummeres som følger:

- (1) fagtermer i henhold til termsystematikker for det prioriterede fagområde,
- (2) fagudtryk fra associerede fagområder, der indgår i branchens hhv virksomhedens tekstproduktion,
- (3) fagudtryk fra perifere fagområder, der optræder i tekster om det prioriterede fagområde med en vis minimumsfrekvens og med en vis tekstspreddning.

Ovenstående kriterium (1), der vedrører maksimerende selektion af termer inden for ordbogens prioriterede område, tager alene udgangspunkt i selve faget, nærmere betegnet i viden herom. Mens metoden således sikrer en terminologisk dækkende ordbog, er den ingen garanti for, at de pågældende fagtermer faktisk optræder i de tekster, der produceres inden for faget. Omvendt er der i forbindelse med de minimerende, tekstbaserede fremgangsmåder (2) og (3) ingen garanti for, at der overhovedet vil blive selekteret deciderede fagtermer fra de pågældende områder.

Herudover må der tages højde for de aspekter, der optræder i den i forbindelse med oprettelse af tekstkorpus udarbejdede aspektfortegnelse, idet der selekteres

- (4) fagudtryk, der må anses for særligt vigtige eller relevante for det prioriterede delfagsområde.

For at en fagordbog skal kunne gøre fyldest i forbindelse med brugsfunktionerne tekstproduktion på og oversættelse til fremmedsproget foreslås tillige selekteret

- (5) øvrige leksikalske enheder, der i forbindelse med tekstproduktion på og oversættelse til fremmedsproget frembyder særlige grammatiske eller kollokationelle problemer.

Også i forbindelse med kriterium (5) er der tale om tekstbaseret selektion, idet identifikation af denne problemtype kræver medvirken af en fagsprogskyndig lingvist. Vedrørende selektion og præsentation af grammatiske og kollokationelle oplysninger i tekniske fagordbøger henvises til Pedersen (1995b:121–142 hhv 1995c:60–73).

3. Afsluttende bemærkninger

Rene terminologiske ordbøger er som nævnt bl.a. karakteriseret ved at begrænse sig til egentlige fagtermer, foruden ved en mikrostruktur, som er fattig på sproglige, men rig på faglige oplysninger. Denne ordbogstypes primære brugsværdi er dermed i forbindelse med tekstreception og som et supplement til egentlige fagbøger. Til forskel herfra intenderer specielt bi- og polylingvale fagleksikografiske produkter typisk at yde hjælp i forbindelse med tekstproduktion på og oversættelse til fremmedsproget. For at kunne opfylde disse funktioner, er det bl.a. påkrævet med en lemmapforståelse, der er ganske betydeligt bredere end den, der begrænser sig til vidensbaseret selektion af egentlige fagtermer. Dette gælder ikke mindst, når vi taler om den hastigt voksende del af fagleksikografien, som tager udgangspunkt i eller orienterer sig mod en konkret branche eller virksomhed. Alene relevant tekstmateriale vil kunne levere de fornødne oplysninger til hjælp for selektion af lemmata efter kriterierne (2), (3) og (5), samt i vid udstrækning også (4), ovenfor, kontrol af ækvivalenter til disse lemmata samt de for L2 sprogproduktion og L1-L2 oversættelse nødvendige grammatiske og kollokationelle angivelser. Det er altså her af største vigtighed, at fagleksikografen har et monofunktionalt tekstkorpus til sin rådighed, dvs. et korpus som er sammensat med dette specielle formål for øje og efter nøje fastlagte tematiske, teksttypologiske og øvrige kriterier, således som det er beskrevet ovenfor. For at sikre at korpus opbygges på et korrekt fagligt grundlag, skal igen fremhæves nødvendigheden af en fagkompetent vurdering allerede på tekstselektionsstadiet af den leksikografiske proces.

Lige så uomgængeligt nødvendigt det er med kriterier for selektion af tekster **til** korpus, lige så påkrævet er det at fastlægge kriterier for selektion af informationer **fra** korpus. Sidstnævnte er her eksemplificeret ved selektion af lemmata men må i videst muligt omfang forhåndsfastlægges for samtlige ordbogens informationskategorier, ikke blot under hensyntagen til den pågældende ordbogs fagområde(r) men også med henblik på ordbogens forudsete brugere og brugsfunktioner samt de problemtyper, ordbogen skal medvirke til at løse.

Litteratur

- Barlach, Richard/Henning Bergenholtz/Wilhelm Gubba 1992: [Anmeldelse af] Steen Hjelmblink: Retsplejeordbog. København: Munksgaard 1991. I: *Hermes* 9, 193–196.
- Stepanoff, A. J. 1957: *Centrifugal and Axial Flow Pumps. Theory, Design, and Application*. New York etc.: John Wiley & Sons.

- Chambers 1988 = Walker, Peter M.B. (ed.): *Chambers Science and Technology Dictionary*. Edinburgh/Cambridge: W. & R. Chambers and Cambridge University Press.
- Clausen da-eng 1990 = *Dansk-engelsk teknisk ordbog*. København: Grafisk Forlag.
- Dictionary of Computing 1988: *Dictionary of Computing*. Middlesex: Peter Collin Publishing.
- Duvå, Grete 1991: [Anmeldelse af] Orlando Gorlindo: Erhvervs- og økonomiordbog – Dictionnaire économique et de l'entreprise. Dansk-Fransk. Danois-Français. GTO Forlag 1990. I: *Hermes* 7, 147–152.
- Duvå, Grete/Anne Ellerup Nielsen 1992: [Anmeldelse af] Anne Møllmann: Dansk-fransk marketing ordbog. Foreningen til Unge Handelsmænds Uddannelse 1990. I: Handelshøjskolen i Århus: *Hermes* 8, 111–126.
- Dyrberg, Gunhild/Dorrit Faber/Steffen Leo Hansen/Joan Tournay 1991: *Oprettelse af fagsproglige tekstkorpora. Engelsk-fransk-dansk juridisk sprog aftaleret*. Handelshøjskolen i København: (= ARK 60)
- Elektroteknisk Ordbog 1983 = *Elektroteknisk Ordbog. Engelsk-dansk. Dansk-engelsk*. Udarbejdet og vedtaget af Dansk Elektroteknisk Komité. København: Dansk Standardiseringsråd.
- Genteknologisk Ordbog 1992 = Kaufmann, Uwe og Bergenholtz, Henning: *Genteknologisk Ordbog. Dansk-engelsk/engelsk-dansk*. København: Gad.
- Grundwortschatz Naturwissenschaften 1988 = Hoffmann, Lothar: *Grundwortschatz Naturwissenschaften. Russisch-Deutsch Englisch-Deutsch Französisch-Deutsch*. Leipzig: VEB Verlag Enzyklopädie.
- Kaufmann, Uwe 1992: Anvendelse af det danske genteknologiske tekstkorpus ved udarbejdelsen af Genteknologisk ordbog, med specielt henblik på udvælgelsen af eksempler. I: Gert Engel (ed): *Korpuslingvistik i Fagsprogsforskningen*. Hindsgavl Slot 26. og 27. november. [Kolding: Handelshøjskole Syd], 56–68.
- Krogh-Hansen, Niels 1994: Oversættelse af fagsproglige tekster. I: Hjørnager Pedersen, Viggo/Niels Krogh-Hansen (eds): *Oversættelsehåndbogen*. København: Munksgaard.
- Kucera eng-ty 1989 = Kucera, A.: *The compact dictionary of exact science and technology*. Volume I: English-German. Wiesbaden: Brandstetter Verlag.
- L & H 1990 = Langkilde, Charlotte et al: *L & H Teknisk Ordbog dansk-engelsk /engelsk-dansk*. L & H Ordbøger.

- Lauridsen, Ole/Ingrid Skytte Andersen 1989: *Rapport over oprettelsen af et dansk og et tysk maskinlæsbart tekstkorpus inden for området genteknik*. København: Handelshøjskolen i København (= ARK 66)
- Laursen, Anna-Lise/Sven Tarp 1991: [Anmeldelse af] Orlando Galindo: Dansk-spansk fagordbog (teknisk forslag a/s 1979) og Dansk-spansk erhvervs- og økonomiordbog (GTO Forlag 1989). I: *Hermes* 7, 153–162.
- McGraw-Hill 1974 = *McGraw-Hill Dictionary of Scientific and Technical Terms*. Editor in Chief: Lapedes, Daniel N. New York etc. : Mc-Graw-Hill.
- Møller, Bernt 1990: [Anmeldelse af] Dansk-fransk teknisk ordbog. Fonden for Fagsproglig Leksikografi. Handelshøjskolen i København og Aschehoug Dansk Forlag A/S 1989. I: *Hermes* 5, 143–152.
- Nielsen, Sandro 1989: Kritisk oversigt over engelske og danske juridiske ordbøger. I: Handelshøjskolen i Århus: *Hermes* 2, 55–75.
- Pedersen, Jette 1994: Anskueliggørelse af faglige relationer i brancheordbøger. I: *LexicoNordica* 1, 203–228.
- Pedersen, Jette 1995a: Systematic classification. I: Henning Bergenholtz/Sven Tarp (eds): *Manual of Specialised Lexicography*. Amsterdam: Benjamins, 83–90.
- Pedersen, Jette: Grammatiske oplysninger i tekniske fagordbøger. I: *LexicoNordica* 2, 1995b:121–142.
- Pedersen, Jette 1995c: The Identification and Selection of Collocations in Technical Dictionaries. I: *Lexicographica* 11, 60–73.
- Riiber, Theis/K.T. Thomsen 1991: [Anmeldelse af] Dansk-tysk teknisk ordbog. Fonden for Fagsproglig Leksikografi, Handelshøjskolen i København og Grafisk Forlag A/S, København 1990. I: *Hermes* 7, 139–145.
- Skibsted 1971 = Skibsted, Sigurd: *Teknisk engelsk-dansk ordbog*. København: Polyteknisk Forlag.
- Skjerk Bilteknik 1991 = Skjerk, Ebbe: *Bilteknisk Ordbog, engelsk-dansk*. København: Teknisk Forlag.
- Stummann, Bjarne 1992: Anvendelsesmuligheder og fagligt indhold af det danske genteknologiske tekstkorpus. I: Gert Engel (ed): *Korpuslingvistik i Fagsprogsforskningen*. Hindsgavl Slot 26. og 27. november [Kolding: Handelshøjskole Syd], 69–74.
- Warrern da-eng 1988 = Warrern, Allan: *Dansk-engelsk teknisk ordbog*. København: Clausen.
- Webster 1984 = *Webster's New World Computer Ordbog*. Frederiksund: Thorsgaard.

Hans Kristian Mikkelsen

Leksikografiske kvalitetsindikatorer. Selvforståelse som udgangspunkt for evaluering

The article deals with the contribution of lexicographers to the question of dictionary quality, which is here understood as three user-based requirements: applicability, reliability and accessibility. Explicit reflections on these problems are primarily found in the outside matter of dictionaries. To chart what can be termed the common lexicographic consciousness on the quality of dictionaries, 155 bilingual LSP dictionaries have been analysed. Above and beyond this common base on quality, indication is made of either divergent views or more advanced, individual ideas of dictionary quality. The results can be viewed both as a metalexigraphic state of the art as far as the practising lexicographers are concerned, and as an aid to the dictionary critic, making it easier for him to put the right questions.

Indledning

Som forbrugere er vi alle interesserede i, at de produkter og ydelser, vi bruger penge på, er så gode som muligt. De må med andre ord være kendetegnet ved en høj kvalitet. Så langt kan vi alle være enige, men når det kommer til en nærmere afklaring af, hvad vi skal forstå ved kvalitet, og hvordan vi kan måle den, skilles vandene. Hvem har kompetencen til at afgøre, hvad der er godt og dårligt? Hvilke kriterier sætter vi som afgørende for kvalitetsdommen?

Når det drejer sig om materielle og immaterielle forbrugsgoder, er det normalt at betragte forbrugeren/kunden som den højeste dommer. Dette gælder også, når forbrugsgodet er en ordbog, hvilket de senere års øgede interesse for brugs- og brugerforskning klart illustrerer. Samtidig må man dog erkende, at brugeren stadig står som "der bekendte Ubekannte", både når det gælder adfærden ved benyttelse af ordbog og når det drejer sig om de situationer, som motiverer til opslag i ordbøger¹. Generelt må man derfor konstatere, at forskningen i

¹ Udtrykket "der bekendte Ubekannte" bruges af Herbert Ernst Wiegand (1977:59), som i samme forbindelse opregner ikke færre end 14 punkter, hvor metaleksikografien mangler empirisk funderet viden om brug og bruger. Samme Wiegand kalder forskning i ordbogsbrug en

ordbogsbrug endnu kun lige er begyndt. Der er en lang række både praktiske og teoretiske vanskeligheder forbundet med gennemførelse af brugsundersøgelser. Til de første hører manglende villighed hos fx professionelle oversættere til at lade sig teste, mens de sidste bl.a. vedrører fortolkningen af de empiriske data til leksikografiske hypoteser². Vi er således nødsaget til at erkende, at det endnu ikke vha. brugsundersøgelser er afklaret, hvad der må kræves af ordbøger. Og mht. fremtidsudsigterne i denne henseende må man også slå koldt vand i blodet, for hvis sådanne undersøgelser overhovedet skal kunne bruges, må der tilføjes et ikke ubetydeligt udefrakommende (= subjektivt) element af hypotesedannelse, tilrettelæggelse af spørgsmål og fortolkning af svar. De "rene" brugs- og brugerdata får man aldrig.

Med denne erkendelse in mente kan vi – uden dog samtidig helt at glemme brugerperspektivet – prøve at vende blikket mod andre og mere upåagtede kilder til forståelse af leksikografisk kvalitet. Her tænker jeg på de erfaringer og den viden, som leksikografer har akkumuleret gennem generationer. Det er denne bevidsthed, udtrykt i ordbøgernes selvforståelse, jeg vil beskæftige mig nærmere med i denne artikel. Som jeg ser det, er der to gode grunde til at rette lyset mod selvforståelsen. Dels får man, sådan som jeg netop har lagt op til, indsigt i den kollektive leksikografiske bevidsthed og dermed adgang til en række begreber, som uddyber kvalitetsbegrebet. Dels bør man, når man fx som anmelder står over for at skulle foretage en evaluering af en konkret ordbog, tage udgangspunkt i ordbogens koncept. En ordbog bør vurderes på sine egne præmisser, og disse fremgår netop af selvforståelsen³.

af metaleksikografiens fire hjørneste. De øvrige er ordbøgernes historie, form og kritik (Wiegand 1987).

² En af dem, der har skrevet mest om undersøgelser af ordbogsbrug er Martha Ripfel, fx Ripfel 1990 og Ripfel/Wiegand 1988. I sidstnævnte formuleres "fire bud" som vejen til en forbedret brugsforskning: 1) uafhængighed af forlagsinteresser, 2) udvikling af et teoretisk grundlag (for at undgå blind empiri), 3) erkendelse af brugsforskningens empirisk sociologiske art og 4) knytning af forbindelse til andre metaleksikografiske discipliner.

³ Evaluering af ordbøger kan også foretages uden hensyntagen til ordbogens koncept. Den mest nærliggende måde er at tage udgangspunkt i sine egne leksikografiske behov og vurdere en given ordbog ud fra, om man i en opslagssituation kan finde nogle troværdige oplysninger, der er i stand til at løse det problem, der har foranlediget opslaget. Fordelen ved denne form for testning er, at der er en god chance for, at den konkrete vurdering bliver ægte. Problemet er, at det testede problem muligvis ikke er repræsentativt for ordbogen, og at evaluatoren ikke er repræsentativ for brugerne.

Leksikografiske kvalitetsaspekter

Ordbøger hører ligesom andre opslagsværker såsom leksika, vejvisere, telefonbøger, lovsamlinger, køreplaner osv. til kategorien "brugstekster". Fælles for dem alle er, som Axel Andersen (i DaOp70–71) udtrykker det, at de er "litteratur i arbejdstøjet", bøger som ikke er beregnet til at blive læst, men til at blive brugt. Brugstekster er redskaber, som man anvender for at løse et problem. Derfor er det også oplagt, at de kvalitetskrav, man stiller til ordbøger, baseres på brugerens behov, ønsker og forventninger samt på den forudsætning, at der er tale om et stykke værktøj. Ligesom andet værktøj kan ordbogen betragtes som en genstand, der er **formet** på en bestemt måde af et bestemt **materiale** med et bestemt **formål** for øje. Forudsætningsforholdet er – set fra leksikografens side – det omvendte: ordbogens intention eller tiltænkte funktion sætter rammer for selektion af materialet, og denne afgrænser for sit vedkommende mulighederne for præsentation af materialet. God kvalitet inden for hvert af disse aspekter indebærer, at ordbogen (og redskabet generelt) er anvendelig, pålidelig og tilgængelig. Altså:

OPERATION	KVALITETSKRAV
funktion	anvendelighed
selektion	pålidelighed
præsentation	tilgængelighed

Kvalitetsbegrebet er således foreløbig blevet oversat til tre brugsorienterede krav (henholdsvis behov eller forventninger), der kan relateres til tre forskellige faser/operationer i arbejdsprocessen. Næste opgave er at få oversat hvert af disse krav til nogle mere målelige størrelser: Ved ordbogens tilgængelighed forstår jeg dens egenskaber som tekst. Her tænker jeg på opbygningen af ordbogsartiklerne (mikrostruktur), af ordlisten/indgangen til hoveddelen (makrostruktur), af ordbogen som helhed, dvs. inkl. omtækster (megastruktur) samt af krydshenvisninger fra artikel til artikel eller mellem andre dele (mediostruktur). Ved ordbogens pålidelighed forstår jeg dens gengivelse af en substans uden for teksten: sproget og den virkelighed, vi lever i. Endelig mener jeg med ordbogens anvendelighed dens tilpasning til bestemte brugsituationer og målgrupper.

Min tese er, at enhver ordbog fra leksikografens side tilstræbes at være anvendelig, pålidelig og tilgængelig. Heraf følger også, at man ved evaluering bør søge at afklare, i hvilket omfang den enkelte ordbog lever op til disse tre kvalitetskrav. For at kunne evaluere en ordbog på dens egne præmisser må man allerførst finde ud af, hvad det er for overvejelser, der ligger bag udarbejdelsen af ordbogen. Denne selvforståelse vil kun fremgå implicit af ordbogens hoveddel(e), hvorimod man ved undersøgelse af omteksterne (fx forord og brugsvejledning) har mulighed for at finde eksplicite udsagn.

Den eksplicite selvforståelse er, når det drejer sig om konkret ordbogsevaluering, udgangspunkt for evaluatorens arbejde med at danne sig et overblik over den enkelte ordbogs samlede selvforståelse. Fordelingen af eksplicit og implicit selvforståelse varierer meget i ordbøgerne, bl.a. fordi leksikografernes (og forlagenes) holdning til omtekster er forskellig. Hvis deres udgangspunkt er, at brugeren alligevel ikke læser omteksterne, skal man ikke regne med at finde ret meget eksplicit stof. Dette betyder ikke nødvendigvis, at der ligger færre overvejelser bag en sådan ordbog end bag en ordbog med mere fyldige omtekster, men det medfører, at evaluatorens adgang til selvforståelsen er mindre direkte og mere besværlig.

Nu er målet med denne artikel at få kortlagt omfanget og arten af de eksplicite selvforståelsesoplysninger, idet en sådan oversigt vil kunne tjene som værktøjsskuffe for den, som skal evaluere en ordbog. Jeg har valgt at afgrænse mit materiale til bilingvale fagordbøger udgivet i Danmark i perioden 1900–1992. Desuden er det en betingelse, at dansk indgår som det ene af de to sprog. Ved fagordbøger forstår jeg ordbøger, hvis indhold er fagligt markeret⁴. I alt har jeg registreret 155 titler, der opfylder disse betingelser. Af dem er 9 fra perioden 1900–1924, 10 fra 1925–1949, 23 fra 1950–1969 og 113 fra 1970–1992. Det skal noteres, at flere af titlerne er kommet i flere udgaver og/eller oplag.

Følgende parametre inden for de tre leksikografiske (og meta-leksikografiske) kvaliteter undersøges:

Anvendelighed:	brugssituation brugerprofil afsenderintention
Pålidelighed:	virkeligheden (dvs. faget) sproget
Tilgængelighed:	megastruktur makrostruktur

⁴ For en oversigt over fordelingen af danske fagordbøger på fag samt over deres historiske udvikling henvises til Mikkelsen 1994.

mikrostruktur
mediostruktur

Indledningsvis har jeg gennemgået alle ordbøgernes omtekster for straks at udskille det mest interessante materiale. Udskilningen foretages efter flg. fremgangsmåde: Hver kvalitetsindikator vurderes efter en skala på tre trin: 0, 1 og 2. Værdien 0 tildeles i de tilfælde, hvor der i omteksterne til en given ordbog ikke optræder eksplicite oplysninger om den pågældende kvalitetsindikator. Værdien 1 tillægges omtekster, hvor kvalitetsindikatoren beskrives eksplicit, men uden at fremstillingen går i dybden med problemet. Endelig gives værdien 2 til fremstillinger, som stikker lidt dybere end førnævnte, hvorved de efter mit skøn får en særlig vægt, når formålet som her er at behandle omteksternes metaleksikografiske værdi generelt for at nå frem til et instrumentarium til hjælp ved selvforståelsesanalyse. (Det skal endnu engang understreges, at værdiudmålingen ikke er tænkt som udtryk for den enkelte ordbogs metaleksikografiske kvalitetsniveau, der som nævnt forudsætter inddragelse af den implicite selvforståelse.) Af tabellen nedenfor fremgår, hvorledes i hvilket omfang kvalitetsindikatorerne er beskrevet i omtekstmaterialet:

	0	1	2
Brugssituation	57 (37%)	92 (59%)	6 (4%)
Brugerprofil	68 (44%)	84 (54%)	3 (2%)
Afsenderintention	103 (66%)	45 (29%)	7 (5%)
Fag/virkelighed	50 (32%)	91 (59%)	14 (9%)
Sprog	80 (52%)	56 (36%)	19 (12%)
Megastruktur	105 (68%)	46 (30%)	4 (3%)
Makrostruktur	87 (56%)	56 (36%)	12 (8%)
Mikrostruktur	47 (30%)	83 (54%)	25 (16%)
Mediostruktur	126 (81%)	21 (14%)	8 (5%)

I min undersøgelse anvendes værdierne på følgende måde: Værdi 1 bruges til at trække de grove selvforståelseslinjer op, dvs. til at registrere hvilke perspektiver der overhovedet spiller en rolle i selvforståelsen. Denne viden betragter jeg som udtryk for den kollektivt akkumulerede leksikografiske bevidsthed. Værdi 2 anvendes til at nuancere dette billede de steder, hvor det er muligt, og sådanne tekster kan betragtes som individuelle bidrag til udbygning af den fælles viden.

Mit mål er at konstatere, i hvilket omfang netop de udvalgte 155 bilingvale fagordbøger kan give et grundlag for en fortolkning af de overordnede kvalitetskriterier anvendelighed, pålidelighed og tilgængelighed. (Hvis jeg havde inddraget mere materiale, fx mono- og

polylingvale fagordbøger – for ikke at tale om almene (= ikke fagligt markerede) ordbøger, ville jeg naturligvis også have fået endnu mere omtekstmateriale. Men samtidig ville den selvforståelse, der kun eller hovedsagelig er aktuell for disse andre ordbøger, gribe forstyrrende ind.) I et afsluttende kapitel anføres de 155 undersøgte fagordbøger med deres tildelte points.

Kvalitetsindikatorer i bilingvale fagordbøger

Brugssituation

Flg. kort over brugssituationer tegner sig på grundlag af de i alt 92 "værdi 1-ordbøger":

- 1 reception
 - 11 generel
 - 111 læsning
 - 112 oversættelse
 - 12 speciel
- 2 produktion
 - 21 generel
 - 211 formulering
 - 212 oversættelse
 - 22 speciel

Hovedgrænsen går mellem receptions- og produktionsordbøger, dvs. mellem ordbøger som bruges ved henholdsvis oversættelse fra og oversættelse til fremmedsprog. Inden for hver af disse hovedtyper kan man skelne mellem ordbøger, der er knyttet til en bestemt teksttype, og ordbøger hvis anvendelse i denne henseende er af generel karakter. Sidstnævnte kan videre inddeles efter, i hvilket omfang modersmålet inddrages. Ved oversættelse spiller modersmålet en væsentlig rolle: ved reception når der skal komme et tekstuel produkt ud af den interlingvale kommunikation, og ved produktion når udgangsteksten på en eller anden måde skal sætte sig spor i målteksten. Ved læsning og (fri) formulering er modersmålet mindre vigtigt, fordi det snarere drejer sig om forståelse, henholdsvis at få sin intention omsat til (fremmed)sprog.

Det skal bemærkes, at den enkelte ordbog udmærket kan forstå sig selv som tjenende flere funktioner. Ikke alene kan en ordbog bekende sig til både reception og produktion (fx Fryd 1977) – men også funktioner, som ligger uden for de egentligt bilingvale, kan tilgodeses. Her tænker jeg især på formidling af faglig viden i form af encyklopædisk,

dvs. ikke-sprogligt, stof (fx Eriksen/Helms/Rømer 1979). Denne funktion varetages primært i monolingvale fagordbøger. En anden "fremmed" funktion er nomenklaturfunktionen, som sigter på at tilvejebringe de rette termer for et givet fagområde. Her er det ofte praktisk at tage udgangspunkt i et sprog (typisk engelsk), hvor der allerede foreligger en standard, men formålet med denne slags ordbøger (termordbøger i modsætning til fagordbøger) er primært at etablere en lignende standard på et andet sprog (fx Eskildsen 1931–32).

Det specielle funktioner (12 og 22) kan være mere eller mindre udprægede. Terkelsen (1977) burde måske slet ikke optræde i mit ordbogsmateriale, men så kan den i det mindste bruges til at markere, hvor grænsen mellem 12 og de rene glosarer går. Ved et glosar vil jeg forstå en ordbog, som er bundet til en enkelt tekst (ikke teksttype). Oppenheim (1933) som eksempel på 22 er ikke knyttet til en bestemt tekst, men til en teksttype: brevet.

Flg. seks ordbøger har fået værdien 2: Ludvigsen (1913), Ludvigsen (1919), Maul (1944), ByAn86, Henriksen (1991), Kaufmann/Bergholtz (1992). De udvider ikke skemaet ovenfor, men nuancerer nogle af dets enkelte punkter.

Ludvigsen (1913 og 1919) hæfter sig specielt ved problemerne med at oversætte præpositionalforbindelser og andre udtryk til korrekt tysk. Desuden skelner han mellem to måder at bruge en ordbog: 1) Man kan som urutineret oversætter søge et ord ved hjælp af et andet ord, 2) eller som mere rutineret kontrollere et ord, som man allerede kender. Denne viden giver ham ideen til det særlige koncept, som Ludvigsen – såvidt jeg ved – er ene om.

Maul giver et klart ræsonnement for sin ordbogs berettigelse: Automekanikeren må have et indgående kendskab til bilens konstruktion og pleje. Dette fås gennem læsning af faglitteratur, som kun for en mindre dels vedkommende findes oversat til dansk. En fagordbog er nødvendig, fordi en almenordbog kun vil rumme de færreste af de særlige ord og udtryk.

ByAn86, som eksplicit anfører både funktionerne 111 og 211, siger også nærmere, hvad det er for forskellige dokumenttyper inden for bygge- og anlægsteknisk litteratur, der skal kunne læses og formuleres vha. ordbogen (nemlig licitationsmateriale, kontrakter, specifikationer mm.). Desuden fremgår det af forordet, at ordbogen ikke er tænkt som et leksikon med detaljerede forklaringer. Hermed siges det klart til bruger og evaluator, at det vil være direkte forkert at kræve, at ordbogen skal kunne hjælpe til mere end reception og produktion. Faglig encyklopædisk viden er ikke på dagsordenen. Hvis man vil

klandre bogen for at mangle denne dimension, er det således snarere konceptet end produktet, man skal angribe.

En lignende advarsel udstikkes af Henriksen, når han advarer mod at bruge ordbogen til andet end det, den er tænkt som: et hjælpemiddel til forståelse af spanske tekster (funktion 111) og til oversættelse til dansk (funktion 112). Advarslen gælder specielt den fristelse, det kunne være at anvende det dansk-spanske register i 211/212-funktion.

I Kaufmann/Bergenholtz (1992) behandles funktionsproblematikken særlig indgående. Det erklæres for direkte uønskeligt at udarbejde en fagordbog til en enkelt funktion, fordi hverken marked eller ekspertiseudbud muliggør dækkende behandling af de enkelte funktioner for samtlige fagområder. Ordbogen erklæres derfor at have flg. syv grundfunktioner – som samtidig siges at være samtlige mulige funktioner (Kaufmann/Bergenholtz 1992:8):

1. reception af danske tekster
2. produktion af danske tekster
3. reception af engelske tekster
4. produktion af engelske tekster
5. oversættelse fra dansk til engelsk
6. oversættelse fra engelsk til dansk
7. information om molekylærbiologi

Inddelingen ligner, men er ikke identisk med funktionskortet ovenfor. Forskellen er tildels af terminologisk art, idet funktion nr. 3 synes at svare til min nr. 111, ligesom nr. 4 svarer til 211. 5 og 6 modsvares af mine 212 og 112. De øvrige Kaufmann/Bergenholtz-funktioner findes ikke på mit kort, fordi de i virkeligheden er monolingvale funktioner. Det ligger nemlig implicit i teksten, at der udelukkende opereres med en dansk bruger. Ordbogen bygger således på et temmelig kompliceret koncept, hvilket også berettiger den efterfølgende modificering, tydeligst i funktionen "produktion af tekster" (danske og/eller engelske?), som kaldes en "sekundær brugermulighed".

Brugerprofil

De 84 værdi 1-omtekster rummer følgende brugerspecificering:

- 1 dansksproget
 - 11 lægmand
 - 111 borger
 - 112 sprogmand
 - 1121 praktiserende
 - 1122 studerende
 - 11221 akademisk uddannelse
 - 11222 teknisk uddannelse
 - 12 fagmand
 - 121 fagets fagmand
 - 1211 praktiserende
 - 1212 studerende
 - 12121 akademisk uddannelse
 - 12122 teknisk uddannelse
 - 122 anden fagmand
- 2 fremmedsproget

Den indledende distinktion mellem brugere med dansk som modersmål og brugere med dansk som fremmedsprog ekspliciteres i mindre grad end flere af de følgende inddelinger, især når ordbogen – sådan som det er den uskrevne regel bag en dansk produceret ordbog – er henvendt til danskere. Kun et par enkelte fagordbøger er primært rettet mod indvandringsgrupper (H.Balahang 1989 og N.Balahang 1989).

Nogle fagordbøger henvender sig eksplicit til enhver (dansker) uanset grad af indsigt i det pågældende fagområde (fx Glahn 1990), men det er dog mere almindeligt, at rette opmærksomheden mod lægfolk eller fagfolk. Førstnævnte gruppe underinddeles yderligere i almindelige samfundsborgere (111) og folk med almensproglige færdigheder inden for det givne fremmedsprog (112). Sprogfolk kan underinddeles efter, hvorvidt de er færdiguddannede (1121) eller under uddannelse (1122) – og sidstnævnte gruppe yderligere efter uddannelsens art: akademisk (11221) eller lavere (11222).

En lignende underinddeling finder man under fagfolk. Nogle ord-bøger henvender sig til professionelt interesserede fra andre fag, fx har miljøordbogen MiDa87 politikere og administratorer som målgruppe sammen med fagets egne teknikere (122). Blandt fagfolk skelnes der mellem dem, der er færdiguddannede (1211) og dem, der endnu studerer (1212). Her er det den færdiguddannede gruppe, der undertiden videreinddeles efter arten af deres uddannelse: akademisk (12121) eller teknisk (12122).

Funktionerne kan kombineres på forskellig vis. Fx er det almindeligt at nævne både (interesserede) lægfolk og fagfolk som målgrupper, jf. fx Abrahamsen (1987), som er rettet mod "brugere og edb-folk", og Jenssen (1923), som gerne skulle sælge hos både "planteelskere" og

gartnere samt det akademiske ekspertklientel: botanikerne. Andre kombinerer ved at lave én målgruppe med forskellige træk, fx sprogligt ukyndige fagfolk hos Hannibal (1958). Mange fagordbøger udpeger specielt korrespondenter og oversættere/translatører som målgruppe. Det er en gruppe, som i sig selv kan ses som forenende sproglig (1121) og faglig (12122) indsigt.

Kun tre omtekster skiller sig ud ved at rumme lidt uddybende stof. Det drejer sig om Maul (1944), Jørgensen/Jordal/Andersen/Smith (1986) og Kaufmann/Bergenholtz (1992). Maul specificerer sin målgruppe lidt mere end det er almindeligt, idet han ikke bare taler om fagfolk, men om "Arbejdsledere paa Autoværksteder, Mekanikere og Lagerfolk" samt tekniske skoler.

Jørgensen/Jordal/Andersen/Smith (1986) ligger på grænsen mellem afhandling, fagordbog og termordbog⁵. Udpegningen af målgruppe ses i tæt sammenhæng med ordbogens brugsfunktioner (oversættelse og terminologisk kompetence), idet "såvel folk med en overvejende sproglig som folk med en overvejende teknisk baggrund" indgår. Det, der udmærker specielt denne omtekst, er også, at selektion og præsentation af materialet sættes i relation til målgruppen.

I pagt med den allerede omtalte intention om at tilgodese syv forskellige funktioner eller brugssituationer definerer Kaufmann/Bergenholtz (1992) også målgruppen/målgrupperne bredt:

Det har været vores ønske at udarbejde en ordbog, som kan tjene som værktøj for alle, der på den ene eller anden måde kommer i berøring med genteknologien og ved hjælp af ordbogen ønsker at skaffe sig en vis viden om faget og dets sprog. Mere konkret kunne dette være medarbejdere inden for stat, amt, kommune og i den private sektor, som i forbindelse med ændrede medicinske behandlingsmetoder, patentansøgninger, levnedsmiddelkontrol, miljøspørgsmål mv. har brug for hjælp i deres arbejde. Der er således tale om en bred vifte af personer, hvis baggrund kan være medicin, teknik, farmaci, biologi og/eller journalistik. Hertil hører også, at mange translatører, som beskæftiger sig med oversættelse af tekster med genteknologisk indhold, kan få gavn af ordbogen (Kaufmann/Bergenholtz 1992:5)

Dertil kommer, at der i forbindelse med modificeringen og den mere eller mindre skjulte prioritering af brugsfunktionerne stilles nogle

⁵ Forfatterne siger også dette selv: "Det skal allerede her understreges, at der med dette værk ikke er tale om en "ordbog" i gængs forstand, da der i værket kun behandles 109 begreber med tilhørende 340 termer." (Jørgensen/Jordal/Andersen/Smith 1986:7). Man bemærker også den typiske terminologiske onomasiologiske (fra-indhold-til-udtryk-) indfaldsvinkel.

forudsætningskrav til brugerne. Fx forudsættes der "en vis fortrolighed med videnskabelig sprogbrug, både på engelsk og dansk og et minimum af kendskab til biologi og kemi." (Kaufmann/Bergenholtz 1992:8)

Ingen af omteksterne redegør for, hvordan de har fundet frem til en bestemt målgruppe (i det omfang de altså udpeger én: 44% gør det ikke!). Der er altså ikke meget, der tyder på, at der er gjort nogen fremskridt hen imod en afdækning af "der kendte Ubekendte"!

Afsenderintention

De i alt 45 værdi 1-omtekster indeholder oplysninger om afsenderen, som jeg mener at kunne sammenfatte i flg. kort:

- 1 holdning
 - 11 deskriptiv
 - 111 registrerende
 - 112 forklarende
 - 12 normativ
 - 121 vejledende
 - 122 standardiserende
- 2 autoritet
 - 21 faglig
 - 211 praktisk
 - 212 teoretisk
 - 22 sproglig
 - 221 praktisk
 - 222 teoretisk
 - 23 pædagogisk

Redegørelsen for afsenderintentionen går i to retninger: den ene udtrykker afsenderens forhold til stof og modtager (1), den anden hans kvalifikationer for overhovedet at kunne regne med at blive taget alvorligt (2).

Underinddelingen af 1 i deskriptiv og normativ/præskriptiv er vigtig, fordi den viser, i hvilket omfang afsenderen involverer sig i forhold til sin modtager. De fleste fagordbøger er af 11-typen, hvilket indebærer, at afsenderen stiller sig betragende og beskrivende uden for sit stof. En mindre del, hovedsagelig dem der nærmer sig egentlige termordbøger, kendetegnes ved, at afsenderen selv tager ansvaret for stoffet og forlanger, at modtageren skal efterkomme de regler, han udstikker (12). Disse to grundholdninger finder man i "slapper-" og "strammerversioner": Den objektivt registrerende beskrivelse (111), fx

repræsenteret af Møllmann (1990), som understreger sin bevidste manglende hensyntagen til normativiteten i skikkelse af Det franske Akademis anbefalinger er et udtryk for streng deskriptivisme, ligesom den standardiserende holdning står for den strenge normativisme (122). Påbudene understreges fx som i InOr91⁶ med reference til DS- og ISO-standarder. Mere almindelige er de fagordbøger, der søger et kompromis mellem 11 og 12 i form af enten en forklarende deskriptivisme (112) eller en vejledende normativisme (121).

Afsenderen kan understrege forskellige slags kvalifikationer – hvilke det lønner sig at understrege må naturligvis ses i forbindelse med brugssituation og brugerprofil. Materialet tillader at skelne mellem tre forskellige slags autoritet: en faglig (21), en sproglig/ leksikografisk (22) og en pædagogisk (23). 21 er oplagt at fremhæve, når det gælder termens korrekthed i såvel udtryk som indhold; 22 når det drejer sig om fx de grammatiske og stilistiske oplysningers nøjagtighed; 23 når målgruppen er defineret som elever eller studerende, og indlæring er i højsædet. 21 og 22 kan evt. være splittet op i en praktisk (211) og en teoretisk (212) baseret indsigt. På titelbladet understreges 211 fx i HaKo23: "Skrevet af en praktisk Købmand til Brug for praktiske Købmænd", mens 212 også kan være inddraget som en instans, der kontrollerer fagleksikografernes arbejde – fx TyDa91, som dermed også er et eksempel på kombination af forskellige afsenderprofiler.

Syv ordbøger rummer stof, der kan nuancere oversigten ovenfor. Jacobsen/Toke (1955) uddyber 121 ved at understrege, at der er grænser for, hvad ordbogen kan yde, selv inden for det område, den er specialiseret i:

Vi vil gerne understrege vanskeligheden ved at oversætte varebetegnelser nøjagtigt. Oversættelsen af ordet "jernplade" afhænger for eksempel af formålet med anvendelsen. Det er umuligt at medtage dem alle i en ordbog. Vore oversættelser af varebetegnelser er derfor tænkt til brug ved anmodning om brochurer eller prøver, medens det tilrådes at foretage ordreafgivelse efter modtagelse af en prøve eller på grundlag af sælgerens brochure, varebeskrivelse el. lign. (Jacobsen/Toke 1955:Forord)

Scientologyordbogen er speciel ved at modificere 112 eller 121 i en religiøs retning:

⁶ "Den [ordbogen] er ikke udarbejdet af en professionel ordbogsredaktør, men er en alfabetiseret samling af tekniske ord, der har det officielle stempel "Dansk Standard". Og et ekstra kvalitetsstempel – mange af de tekniske ord er godkendt ikke alene europæisk, men også internationalt." (InOr91:Forord)

Scientology-ord og deres definitioner er vejen til et nyt syn på og en ny forståelse af livet. At forstå dem hjælper dig til at leve bedre. Frem for alt er denne lille ordbog overladt dig for at hjælpe dig ad sandhedens vej som er Scientology. (Hubbard 1970:8)

Eriksen/Helms/Rømer (1979) præciserer 122:

Det skal fremhæves, at de til termerne hørende definitioner kun kan være genstand for standardisering, hvad angår den begrebsmæssige afgrænsning og sammenknytning af termerne; ordlyden i definitionerne og i anden ledsagende tekst er ikke i sig selv at betragte som standardiserbar. (Eriksen/Helms/Rømer 1979:7)

og kombinerer med 121:

Som det vil fremgå, er der i enkelte tilfælde anført danske synonymer. I sådanne tilfælde anbefales det synonym, der står først. (Eriksen/Helms/Rømer 1979:8)

Kaufmann/Bergenholtz (1992) kombinerer også forskellige holdninger. Selektion og præsentation af eksempler styres således af nogle overordnede principper om, at eksemplerne dels skal have en vejledende funktion (121, fagligt eller sprogligt), dels skal fremstå på samme måde som i kildematerialet – altså et ekstremt deskriptivt princip (111):

Vi er ikke ansvarlige for eksemplerne, hverken for indholdet eller for sprogbugen. Der vil derfor i visse tilfælde være en diskrepans mellem artiklens sproglige oplysninger og eksemplets sprogbrug [...] Således kan vi heller ikke i alle tilfælde stå inde for eksemplernes faglige korrekthed, i særdeleshed deler vi ikke uden videre hverken explicit genteknologisk-kritiske eller uindskrænket fremtidsdrømmende citater... (Kaufmann/ Bergenholtz 1992:19f)

Samme ordbog indleder sit forord med at hævde nødvendigheden af at kombinere fagspecialistens (212) kvalifikationer med sprogspecialistens/leksikografens (222). Der refereres i den forbindelse til tesaerne fra en leksikografisk kongres i Bad Homburg i 1975.

TeRe50 fremhæver sin faglige autoritet ved at give en historisk redegørelse for arbejdet i den ekspertgruppe, som har ansvaret for ordbogen. Som udvalg nedsat af Terminologicentralen under Akademiet for de tekniske Videnskaber kan autoriteten næppe blive større. Medlemmerne af udvalget nævnes ved navn, titel og institution.

EnDa83 refererer til såvel faglig som sproglig autoritet, sidstnævnte i form af Dansk Sprognævn, som har gennemset de danske termer. DaDa85, som er en dansk-engelsk og dansk-dansk dataordbog efter norsk forlæg, beskriver indgående forlæggets videnskabelighed, lige-

som også kvaliteten af den danske bearbejdning sikres af en dansk redaktør, der nævnes ved navn, grad, titel, og institution.

Fag

De fleste fagordbøger rummer et minimum af oplysninger om det fagområde, de dækker. Nedenfor har jeg fortolket disse informationer i et forholdsvist groft UDK-net:

0	Information og dokumentation
1	Filosofi og psykologi
2	Religion
3	Samfund
<3>...32	Samfundsvidenskab, statistik, politik
33	Samfundsøkonomi
34	Jura
35...39	Offentlig forvaltning, socialvæsen og forsikring, uddannelse og undervisning, metrologi, etnografi
5	Matematik og naturvidenskab
6	Anvendte videnskaber
61	Medicin
62	Teknik, alment
<62>	Teknik, alment
620	Materialeprøvning, materialekundskab, energiøkonomi
621	Maskinteknik
622...623	Minedrift, våbenindustri
624...628	Bygningsingeniørvæsen, trafikanlæg, offshore-konstruktioner, teknisk hygiejne
629	Trafikmidler
63...64	Landbrug, skovbrug, jagt, fiskeri, husholdning
65	Virksomhedsøkonomi, alment
<65>	Virksomhedsøkonomi, alment
650...651	Virksomhedsledelse, kontororganisation
654...656	Telekommunikationsdrift, transportvæsen
657...659	Regnskabsvæsen, driftsøkonomi, reklame og konsulentvirksomhed
66...69	Kemisk-teknisk industri, andre industrier, byggeri
7	Kunst og kultur
8	Sprog og litteratur
9	Historie og geografi

Dette net er velegnet som illustration af ordbøgernes fordeling på fag. Klasserne 0, 1, 2 samt 8 og 9 er faktisk ikke repræsenteret i mit værdi 1-

materiale, men jeg tager dem med for overblikkets skyld⁷. Underinddelingerne viser, at de fleste fagordbøger befinder sig inden for samfundsfagene og de anvendte videnskaber, specielt teknik og virksomhedsøkonomi.

I materialet har jeg fundet 14 omtekster, der beskriver det faglige indhold mere detaljeret. Ludvigsen (1913) og (1919) på den ene side og Jørgensen/Jordal/Andersen/Smith (1986) på den anden er modsætninger mht. afgrænsning af fagområde. Ludvigsens ordbøger er først og fremmest beregnet til oversættelse af erhvervskorrespondance, hvorfor der mere satses på at dække i bredden end i dybden, endog i et sådant omfang at den almensproglige del af korrespondanceleksikken også omfattes:

Thi det maa man naturligvis ikke se bort fra, at paa 7–800 Sider, fordelt paa to Dele, kan man ikke vente sig alt, og et Vareleksikon er det jo ikke; de tekniske Gloser inden for den enkelte Branche gaar jeg ud fra, at Korrespondenten kender gennem Kataloger, Prislister, Fakturaer etc.; en Ordbog, hvori tillige alle de tekniske Betegnelser paa Varer og lgn. skulde findes, vilde blive flere Bind tyk. At den dansk-tyske Del er blevet noget større end den tysk-danske, ligger simpelt hen deri, at der i denne Del er blevet optaget hel Del almindeligere Gloser der mere eller mindre strejfer Handelssproget, for at det ikke skal være nødvendigt for den første den bedste Glose at have en almindelig Ordbog ved Haanden. (Ludvigsen 1913:Forord)

Jørgensen/Jordal/Andersen/Smith (1986) derimod holder sig inden for et ganske bestemt område: naturgasterminologi. Den faglige bearbejdning er meget omfattende, idet det der på ægte terminologisk vis begyndes med at etablere et universelt begrebssystem, som de russiske og danske termer så defineres efter.

Ordbøgerne fra Fonden for Fagsproglig Leksikografi ved Handelshøjskolen i København (DaFr89, DaEn90, DaTy90, FrDa90, EnDa91a og TyDa91) er skåret over samme læst, også når det gælder den faglige dækning. De skal dække det centrale ordforråd inden for "moderne teknisk sprog", som det hedder i forordet til DaFr89. I en anden omtekst, Emneforkortelser, opregnes ikke færre end 129 fagkoder (hvilket kunne tyde på en vis uoverensstemmelse mellem centrum og periferi: et oplagt emne at undersøge i en kritisk analyse).

I to ordbøger af nyere dato finder man større faglige beskrivelser. Henriksen (1991) giver en kontrastiv redegørelse for forholdet mellem

⁷ Se også Mikkelsen 1994. Klasserne 0, 1, 2, 8 og 9 findes primært beskrevet i monolingvale ordbøger.

det spanske og det danske retssystem med særligt henblik på de forskelle "der erfaringsmæssigt giver anledning til problemer i forbindelse med oversættelse mellem de to landes sprog." (Henriksen 1991:399) Kaufmann/Bergenholtz (1992) giver en introduktion til faget (genteknologi), som dels kan læses som en selvstændig, ordbogs-uafhængig tekst, dels er ment som en støtte ved ordbogsbenyttelse, når der måtte være behov for en redegørelse af den type, som man finder i større encyklopædier.

Sprog

De 56 værdi 1-omtekster giver materiale til flg. skema:

- 1 opslagsord
 - 11 ordklasse
 - 11... (de enkelte ordklasser)
 - 12 ordtype
 - 121 term
 - 122 nomen
 - 123 problemord
- 2 oplysning
 - 21 leksik
 - 211 -nym
 - 212 forklaring
 - 213 kollokation
 - 22 grammatik
 - 221 fleksion
 - 222 derivation
 - 223 komposition
 - 224 syntaks
 - 23 fonetik
 - 24 stilistik

De sproglige oplysninger vedrører dels de ord, der skal optræde som opslagsord (1), dels de oplysninger, som gives til opslagsordene (2). Under 1 kan der skelnes mellem de traditionelle ordklasser (11) og hvad jeg har valgt at kalde "ordtyper" (12). Sidstnævnte dækker lingvistiske begreber som term og nomen. Kategorien problemord møder man fx hos Rosbach/Thornfield (1981), der selekterer danske ord, hvis oversættelse erfaringsmæssigt indebærer store sproglige problemer. Nominer (122) kan nævnes som noget, der ofte bevidst fravælges, jf. citatet ovenfor fra Ludvigsen (1913), hvor varenavne ville få ordbogen til at eksplodere, men der er også ordbøger, der netop vælger at

beskrive dette stof (fx Jonsson 1988 med navne på islandske heste). For ordbøger af terminologisk karakter er det almindeligt at fokusere på substantiver (11...), mens de ordbøger, som mere tilgodeser de kommunikative funktioner reception og produktion, typisk vil inkludere flere ordklasser. Fx specialiserer Unnerup-Madsen (1989) sig i præpositionale udtryk, som er særdeles relevante ved oversættelse af erhvervsproglige tekster. Fagordbøgerne vil normalt udvælge deres stof fra et tekstmateriale (evt. et tekstkorpus), mens termordbøgerne vil basere deres selektion på faglige kriterier⁸.

2 lader sig inddele i traditionelle sprogvidenskabelige kategorier. Jeg bruger betegnelsen "-nym" (211) for en række ord, der står i et semantisk paradigmatiske forhold til opslagsordet. Med "semantisk paradigmatiske" mener jeg synonyme, antonymer, hyponymer osv., der kendetegnes ved et indholds-baseret slægtskab med de ord, som de står i -nym-forhold til. Dette forhold er af paradigmatiske art, fordi der er tale om en enten-eller-relation. Både-og-relationen (= den syntagmatiske) repræsenteres i skemaet af kollokationen (213). I bilingvale ordbøger er ækvivalenten (= det interlingvale synonym) et nærmest obligatorisk -nym-ord. 212 er den anden udbredte – i særdeleshed i monolingvale ordbøger – leksikalske oplysning, nemlig forklaringen (eller definitionen). Dertil kommer de øvrige mulige oplysningskategorier: diverse grammatiske (22), foniske (og grafiske) (23) samt stilistiske (24). Til de sidstnævnte regner jeg også leksikografiske eksempler. Oppenhejms (1933) "vendinger" hører til denne kategori.

I alt 19 ordbøger har jeg givet værdien 2. Dog ikke alle fordi de udmærker sig ved lingvistisk indsigt, jf. Hubbards interessante begrundelse for at selekttere forklaringer/definitioner frem for synonyme til oplysningssiden:

... at sætte et **andet** ord ind i stedet for det oprindelige er at forkludre det hele. Den **rigtige** fremgangsmåde er at gennemse, definere rigtigt og forstå **det** ord, der er anvendt. (Hubbard 1979:7)

Jacobsen/Toke (1955) har mere grund under sig, når de peger på, at der under ganske bestemte forhold kan være tale om manglende ækvivalens (og dermed manglende interlingval synonymi). Desuden giver de et bud på løsning af problemet uden at gribe til parafrasering i form af en forklaring:

⁸ Jeg behandler denne selektionsproblematik i Mikkelsen 1994a:197, hvor jeg skelner mellem et tekstbaseret og et vidensbaseret selektionskriterium.

I tilfælde, hvor egentlig oversættelse ikke er mulig, for eksempel ved stillinger og navne på institutioner har vi alligevel anført en glose med betegnelsen: (*omtrent*), idet korrespondenten ofte vil være tvungen til at skrive noget, der kan lede englænderens tanke i retning af det danske ord. (Jacobsen/Toke 1955:Forord)

Poulsen (1987) er en af de få, som gør særligt meget ud af ikke blot de semantiske, men også syntaktiske oplysninger. De undtagelser, der er i denne henseende, findes i erhvervsproglige ordbøger til brug for korrespondenter og oversættere, hvilket ikke kan undre, eftersom deres opgave er at fremstille sammenhængende og korrekte tekster. Poulsen beklager samtidig, at der pga. manglende forskningsinteresse for emnet simpelthen mangler viden om den danske opslagsords semantik og syntaks. Desuden berører han spørgsmålet om det empiriske grundlag (= kildematerialet), som ligeledes er mangelfuldt⁹. Det er også i erhvervs- (eller handels-) sproglige ordbøger, at man finder grammatiske og stilistiske omtekster. Poulsens ordbog rummer således oplysninger om korrekt titulering, opstilling af brev samt en tematisk ordnet redegørelse for købs- og salgsbetingelser. Claudi (1931) giver en (ganske vist temmelig primitiv) redegørelse for den grafiske side af sprog: alfabet, retskrivning og tegnsætning. Hansen (1988) indeholder et kontrastivt grammatisk afsnit. Jørgensen (1983) beskriver zoologiske nominers særegne såkaldt bi- og trinominale struktur – et kompositionsanliggende (223).

Jenssen (1923) og Abrahamsen (1987) fortæller, hvilken slags ord de ikke har selekteret som opslagsord. For førstnævntes vedkommende drejer det sig om nogle sjældent anvendte synonymmer, lokale artsnavne, slægtsnavne som ikke forekommer i kildematerialet mm. (Jenssen 1923:vi). Flg. passage fra Abrahamsen tyder på en klar selvforståelse mht. den sproglige afgrænsning:

Følgende områder er ikke forsøgt dækket af denne edb-ordbog, selv om flere ord i de forskellige kategorier er medtaget.

- forskellige virksomheders lokale jargon
- navne og forkortelser for specifikke produkter
- forkortelser
- bibliografier [!] over edb-personligheder
- navne på edb-firmaer. (Abrahamsen 1987:8)

⁹ Det skal dog noteres, at Statens Humanistiske Forskningsråd netop i slutningen af 80erne iværksatte initiativområdet Fagsproglig Kommunikation, som bla. resulterede i fremstillingen af nogle større fagsproglige tekstkorpora.

Kaufmann/Bergenholtz (1992) omtaler udførligt selektionen af såvel opslagsord som oplysninger, både til den dansk-engelske og til den engelsk-danske del. For opslagsordenes vedkommende nævnes kilde-materiale og udtagningsmetode, og for oplysningernes vedkommende sættes selektionen i relation til ordbogens funktionelle træk (brug, bruger, afsender), fx:

De enkelte ordbogsartikler indledes med **opslagsordet**, der er forsynet med en **grammatisk angivelse**, når det skønnes, at brugeren ud fra sit kendskab til dansk almensprog ikke eller ikke sikkert vil have den fornødne sprogviden, hvis han/hun vil producere en tekst under brug af det danske opslagsord. (Kaufmann/Bergenholtz 1992:10)

Undertiden kan dog manglende forekomst af tekstuel belæg betyde, at et eksempel (24) ikke anføres, selv om det funktionelle behov er der. Dette kan ses som et ekstremt udtryk for det, jeg i afsnittet om afsenderintention kaldte streng deskriptivisme. Iøvrigt er samme ordbog et eksempel på en meget bred dækning i valget af oplysninger. Den dansk-engelske del har således flg. oplysningstyper: "grammatik", dvs. fleksion (221), engelsk ækvivalent (211) – som oven i købet kan have sine egne grammatiske oplysninger tilknyttet – faglig forklaring (212), eksempel (24), kollokation (213), synonymer og antonymer (211).

Megastruktur

Jeg har fundet 46 værdi 1-omtekster til belysning af den mega-strukturelle selvforståelse. Ved megastruktur forstår jeg disponeringen af de største af ordbogens komponenter: diverse omtekster og hoved-del(e).

- 1 hoveddel
 - 11 én
 - 12 flere
- 2 omtekst
 - 21 forord
 - 22 indholdsfortegnelse
 - 23 forkortelser
 - 24 grammatik
 - 25 fagbeskrivelse
 - 26 illustrationer/billeddel
 - 27 tekstdel
 - 28 uregelmæssige verber
 - 29 register/indeks
 - 30 bibliografi

Kortet viser den basale opdeling mellem den obligatoriske, ordbogs-konstituerende hoveddel og de mere fakultative omtekster. Omteksterne kunne evt. indledende opdeles efter disses placering i forhold til hoveddelen(e), altså i for-, mellem- og bagtekst. En sådan skelnen optræder dog ikke i de selvforståelser, som ligger til grund for kortet, hvorfor der udelukkende nævnes eksempler på de typer af omtekster, der forekommer. Rækkefølgen svarer nogenlunde til den typisk forekommende.

Flere hoveddele (12) finder man i fx bilingvale ordbøger, der tjener såvel reception som produktion. Fryds revisionsordbøger er eksempel herpå – omend brugsfunktionerne kun fremgår implicit. Pedersen (1940) er en hybrid mellem bilingval og polylingval ordbog, idet den har fire hoveddele: tysk-dansk, engelsk-dansk, fransk-dansk og dansk-tysk/engelsk/fransk: et ikke usædvanligt megastrukturelt koncept. Et andet ganske hyppigt koncept er en monolingval del bestående af danske lemmata (= strukturerede opslagsord) med tilhørende forklaringer/definitioner på dansk og en bilingval del indeholdende lemma og ækvivalent. Konceptet er almindeligt i termordbøger, men også i andre slags ordbøger, hvor funktioner og materiale kan være så forskelligartede, at leksikografen foretrækker at præsentere stoffet adskilt. I Oppenheim (1933), A.Moth-Lund (1933), D.Moth-Lund (1933) og Lohmann (1956) er stoffet så forskelligt, at det foretrækkes placeret i hver sin hoveddel: en bilingval sætningsbog og en bilingval ordliste.

Det kan være vanskeligt at skelne skarpt mellem 12 og 11+29, men der tales også i mange tilfælde om en registerdel, som kun har en hjælpefunktion i forhold til en hoveddel. Når hoveddelen er systematisk ordnet, sådan som det er tilfældet i mange termordbøger, er det nødvendigt med en alfabetisk indgang i form af et register. Sådanne

termordbøger er ofte enten monolingvale eller polylingvale, men der findes også bilingvale, fx DsEl69. Samme behov for et alfabetisk register finder man fx i Stürup/Eilertzen (1977), en menuguide med en tematisk ordnet hoveddel. Henriksen (1991) er et eksempel på, at en omtekst, den tidligere omtalte faglige kontrastive oversigt, i sig selv forsynes med et register.

Nogle ordbøger rummer tekstdele (27). I de ovenfor nævnte sætningsbøger (Oppenheim 1933 osv.) finder man fx et afsnit med eksempler på breve. Det er også oplagt for menuguiden at aftrykke et fransk spisekort fra Alsace. – Alternativet ville være at præsentere teksterne opsplittet som stilistiske eksempler i de enkelte ordbogsartikler.

Der er en tendens til bevidst at undgå for mange omtekster. JoVe89, som først og fremmest behandler engelske lemmata rummer i sin hoveddel også danske henvisningslemmata, dvs. sådanne lemmata som ikke rummer nogen mikrostrukturel angivelse ud over en henvisning til et andet lemma. Denne fremgangsmåde er konsekvent, hvilket reelt vil sige, at der er tale om, at et dansk register til den engelske hoveddel er blevet integreret i denne¹⁰.

Fire ordbøger har værdien 2. Ludvigsen (1913) og (1919), fordi der er tale om et unikt koncept, eftersom ikke alene den dansk-tyske (henholdsvis dansk-engelske), men også den tysk-danske (henholdsvis engelsk-danske) hoveddel er tænkt som bistand til den bruger, som oversætter fra dansk til tysk (/engelsk). Altså har begge dele en produktiv funktion, jf. omtalen i afsnittet om brugssituation. Den udførlige omtale af stoffets disponering i Jørgensen/Jordal/Andersen/Smith (1986) er interessant som eksempel på en typisk terminologisk bestemt megastruktur. Kaufmann/Bergenholtz (1992) forklarer i forordet og afsnittet "Informationer om ordbogens opbygning og baggrund" bogens opbygning og meningen med de enkelte afsnit. Afsnittet er specielt, fordi det ikke er rettet til ordbogens kernebruger, men til andre leksikografer.

Makrostruktur

¹⁰ Man kan undre sig over, at denne ganske radikale og utraditionelle løsning ikke begrundes nærmere – i hvilket fald jeg ville have givet omteksten værdien 2. Beskrivelsen lyder lakonisk: "Ordbogen er ordnet alfabetisk, således at uanset om opslagsordet er dansk eller engelsk, slås det op i samme alfabet. [...] Er opslagsordet dansk, henvises der til et eller flere engelske opslagsord. [...] Er opslagsordet engelsk vil artiklen se ud som følger ..." (JoVe89:5)

56 ordbøger har værdien 1 for makrostruktur. Makrostruktur er betegnelsen for den måde de strukturerede opslagsord, lemmata, er ordnet indbyrdes.

- 1 alfabetisk
- 2 systematisk

Til trods for beskrivende omtale i de ovennævnte ordbøger kan der ikke spores nogen synderlig metaleksikografisk bevidsthed hos de praktiserende leksikografer om dette spørgsmål. Det skyldes sikkert, at fx den alfabetiske ordning af lemmata anses som et selvfølgeligt og mindre vigtigt anliggende. Oftest konstateres blot, at opslagsordene i den pågældende ordbog er alfabetisk ordnet. Af samme grund er det ikke særlig svært at finde eksempler på uheldig og inkonsekvent ordning i ordbogsmaterialet¹¹.

Den alfabetiske makrostruktur beskrives i enkelte ordbøger lidt nøjere. Fx siges i ordbøgerne fra Fonden for Fagsproglig Leksikografi:

Opslagsordene står i alfabetisk orden, dog er nogle flerleddede opslagsord anført under et andet led end det første, hvis dette andet led anses for det centrale led eller det led, hvis oversættelse formodes at volde problemer. [...] Flerleddede underopslagsord er alfabetiseret efter det første led, idet der ses bort fra hovedopslagsordet. (DaEn90)

Der er et par ubekendte i denne formulering, som ikke uden videre skaber klarhed, men det peger i retning af det, man i metaleksikografien kalder et redeordningsprincip. Det betyder i praksis, at man skal finde "selvrelativ adresse" under "adresse", men alfabetisk ordnet inden for artiklens rammer mellem "indirekte adresse" og "symbolsk adresse". Derimod skal man finde "adresserbart punkt" som selvstændigt (hoved)lemma, placeret mellem "adresse" og "adressere". Mod-sætningen til dette princip er det strengt alfabetiske, som Kaufmann/Bergenholtz (1992) beskriver:

Alle opslagsord er ordnet i en **strengt alfabetisk rækkefølge** uden hensyntagen til særtegn, fx bindestreg, apostrof eller tal, som kan være

¹¹ Spørgsmålet er imidlertid, om det betyder så forfærdelig meget for muligheden for genfindning. Paradoksalt nok er makrostruktur et rigt udforsket emne inden for den teoretiske metaleksikografi, specielt i den kontinentale strukturalistiske tradition, som er stærkt repræsenteret i fx WöDi89–91. Inkongruensen mellem det, som henholdsvis praktikere og teoretikere finder (u)væsentligt, kunne tyde på, at forskerne har et forklaringsproblem.

en del af ordet. Dog anføres ord indeholdende særtegn efter ord uden særtegn. (Kaufmann/Bergenholtz 1992:15)

– eksemplificeret i rækkefølgen

G
G1
G2
galle
gamet
gap
gapped duplex
gaskromatografi
G-bånd
GC
G+C
GC-box.

Ved flerordstermer bruges den såkaldte "ord-for-ord"-variant af den strenge alfabetisering. "Tegn-for-tegn"-princippet ville give rækkefølgen

embryo
embryo cell
embryogenesis
embryo transfer

– mens ord-for-ord-princippet giver

embryo
embryo cell
embryo transfer
embryogenesis

Jacobsen/Toke (1955) har ønsket at spare på pladsen for at mindske den tid, det tager at slå op i bogen. Det har de gjort dels ved en stram selektion, dels ved "en uortodoks forkortning, der ikke følger stavel-sedeling og ordgrupper" (Jacobsen/Toke 1955:Forord). I realiteten er der tale om det, som man i dag kalder nicheordningsprincippet. I stedet for at opføre fx

accept
acceptant
acceptbog
acceptkredit
acceptprovision

acceptvægning

som selvstændige lemmata anbringer man dem alle i samme artikel med "accept" som hovedlemma med resten i alfabetisk orden

accept [...]

-ant [...], -bog [...], -kredit [...], -provision [...], -vægning [...].

Princippet følges dog ikke fuldt ud, idet verbet "acceptere" ikke puttes ind mellem "-bog" og "-kredit", men anføres som selvstændigt lemma, alfabetisk placeret mellem "accept" og "accidenstrykker", muligvis fordi der er temmelig mange oplysninger knyttet til "acceptere".

Termordbøger tager typisk udgangspunkt i en systematisk orden. Samtidig kan der være internationale hensyn at tage. Det er fx tilfældet i Eriksen/Helms/Rømer (1979), som udelader nogle forældede termer fra det engelske forlæg, men iøvrigt bevarer samme løbenummerering som forlægget. På denne måde opretholdes forbindelsen til dette, men den danske ordbogs makrostruktur bliver hullet pga. udeladelserne, ligesom den oprindelige emneordning mistes! Dertil kommer, at helt nye tilføjede afsnit er ordnet alfabetisk!

Mikrostruktur

Det kan være vanskeligt at opretholde en skarp grænse mellem mikrostruktur og sprogafsnittet, specielt dets oplysningsafdeling (se ovenfor). De enkelte sproglige kategorier indgår nemlig som **substans** i mikrostrukturens forskellige **fremtrædelsesformer**:

- 1 elementer
 - 11 lemma
 - 111 komprimering
 - 12 angivelse
 - 121 komprimering
 - 13 kvalifikator
 - 131 glosse
 - 1311 komprimering
 - 132 markering
 - 1321 komprimering
 - 133 kotekst
 - 1331 komprimering
- 2 strukturer
 - 21 program
 - 22 adressering
 - 23 gruppering
 - 231 prioritering
 - 24 polysemi/homonymi
- 3 strukturmarkører
 - 31 typografi

Skemaet fortolker selvforståelsesoplysningerne fra de i alt 83 ordbøger med værdi 1-omtekster. Overordnet skelner jeg mellem stoffet i form af de formaliserede elementer, strukturerne mellem elementerne og de markører, der viser de strukturelle relationer.

Elementerne (1) kan videreinddeles i det element, lemmaet (11), man slår op for at få yderligere oplysninger, de oplysninger man søger: angivelserne (12) samt de informationer, som supplerer 11 og 12 i en eller anden henseende ved at fortolke dem (131), relatere dem til en encyklopædisk/faglig kontekst (132) eller sætte dem ind i en tekstuel, sproglig sammenhæng (133). Alle elementerne kan udtrykkes på forskellig vis: fra fuldforn til maksimalt komprimeret form.

Strukturerne (2) kan inddeles i et mikrostrukturelt program (21), som er et sæt regler for, hvordan man behandler forskellige typer af elementer; en adressering (22), som er en relation mellem beskrevet og beskrivende; en gruppering (23) af elementer af samme slags samt polysemi/homonymi (24), som er en relation mellem to elementer, der er identiske i én henseende, men forskellige i en anden.

Endelig er der de udtryksformer, som strukturerne gøres synlige i: strukturmarkørerne (3).

Lemmaet indgår som element i både makro- og mikrostruktur. Lemmaet er det strukturerede opslagsord: det som man vælger skal være indgang til informationerne i ordbogsartiklen. Struktureringen består i, at man vælger en bestemt form af de enkelte ordklasser (når der er tale om ord med flere mulige former), fx nominativ singularis af

substantiver og infinitiv af verber, og lader denne form repræsentere hele ordet. Mulighederne for at komprimere lemmaets udtryksform har vi allerede set et eksempel på ovenfor i form af den alfabetiske nichestruktur.

Angivelserne kan bestemmes som de sproglige oplysninger, der er tænkt som den information, brugeren søger. Substansen kan være alle de kategorier, der er nævnt som den sproglige dimensions oplysninger, og der kan være tale om forskellige former for komprimering af udtrykket. Mest oplagt er det måske i den grammatiske angivelse, hvor en enkelt kode kan indplacere ordet entydigt i det grammatiske system. Vil vi vise, hvordan et ord bøjes, er det under normale omstændigheder for voldsomt at nævne hver eneste form (selv om sådanne ordbøger også findes!). I stedet kan man med et enkelt bogstavs eller et enkelt tals reference til en grammatisk oversigt slutte sig til samtlige ordets former.

Kvalifikatorerne er oplysninger, som præsenteres som uselvstændige informationer. Deres eksistensberettigelse er at præcisere det lemma eller den angivelse, som det knytter sig til (: er adresseret til). Præciseringen kan ske som henholdsvis glosse, markering eller kotekst, og også her er der mulighed for komprimerede udtryksformer.

Programstrukturen er det mønster/de mønstre, som ordbogsartiklerne følger, hvis de udarbejdes på en velovervejet og konsekvent måde. Adressering er en relation, som siger, hvilke informationer der hænger sammen. Oftest er størrelser, der er placeret ved siden af hinanden, også knyttet sammen vha. en sådan relation; men adresseringsrelationer kan også forekomme mellem rumligt adskilte del, jf. hoved+hale-strukturen, som man finder hos fx Gubba (1987):

De fleste mere komplekse opslagsord indledes med en kort oversigt med angivelse af de forskellige betydninger; derefter følger en samling eksempler, som skal tjene til at lette brugeren korrekt anvendelse af opslagsordet på målsproget. (Gubba 1987: Kort vejledning)

Gruppering er der fx tale om, når der er flere ækvivalenter til samme lemma. Grupperingen kan være uspecificeret/udifferentieret eller der kan bestå en prioriteringsrelation, fx i form af en regel, der siger, at den førstnævnte ækvivalent er den "bedste" i en bestemt henseende og dermed den, man bør vælge, hvis der ivotrigt ikke er noget, der taler imod det.

Polysemi- og homonymiproblemet findes både i lemma og angivelse. Ved lemmatisering er det vigtigt at beslutte sig for, hvordan man vil behandle/præsentere homonymer og polysemer: som selvstændige

lemmata eller som ét lemma, hvis videre afklaring er et spørgsmål om monosemering (entydiggørelse) vha. kvalifikatorer.

Strukturmarkørerne bruges fx til at adskille enkeltbetydningerne til et polysemt lemma (fx ved nummerering af disse) – eller til at markere homonyme lemmata (fx ved et lille hævet ciffer, der viser, at der er flere ord med samme lemmatiske form.

Blandt de 25 værdi 2-omtekster vil jeg nævne nogle enkelte. Jacobsen/Toke (1955) gør en del ud af at forklare parentesens funktion som komprimeringsmiddel:

Parentes i engelsk tekst betyder enten: at det, der står i parentesen, kan udelades:

form a (proper) quorum

læses: form a quorum *eller* form a proper quorum

bjærgning salvage (operation)

læses: **bjærgning** salvalge *eller* **bjærgning** salvage operation

eller: at parentesen viser en variant:

order (form) note

læses: order form *eller:* order note

acetate (sugar) of lead

læses: acetate of lead *eller:* sugar of lead

keen (severe) competition

læses: keen competition *eller:* severe competition (Jacobsen/Toke 1955)

Synonymer undgås ofte bevidst i termordbøger, og her er Eriksen/Helms/Rømer (1979) ingen undtagelse. Når de undtagelsesvist alligevel forekommer er de klart nedprioriteret:

Som det vil fremgå, er der i enkelte tilfælde anført danske synonymer. I sådanne tilfælde anbefales det synonym, der står først. Af overskuelighedshensyn er synonymer i øvrigt kun anført ved hovedtermen og udeladt i alle sammensætninger. Engelske synonymer er medtaget som anført i IFIP-ICC ordbogen. (Eriksen/Helms/Rømer 1979:8)

Hasselbalch/Neal/Hasselbalch (1984) ønsker tilsyneladende ligestilling, når der er flere ækvivalenter: de vælger at opføre dem i alfabetisk orden!

Poulsen (1987) er meget omhyggelig med at redegøre for strukturmarkører i et særligt afsnit "De anvendte typers og enkelte tegns betydning". I et andet afsnit af vejledningen "Forklaring af en ordbogsartikel" dissekerer han en typisk ordbogsartikel for at vise, hvor mange informationer der kan hentes ud af den, hvis man sætter sig ind i forkortelser (= komprimerede udtryksformer) og typografiske konventioner (= strukturmarkører).

I de tekniske ordbøger fra Fonden for Fagsproglig Leksikografi bruges fagmarkeringerne også som strukturmarkører, eftersom artikler for polyseme lemmata ordnes netop efter fagområde, idet evt. generelle betydninger dog kommer allerførst.

Kaufmann/Bergenholtz (1992) gør en del ud af, at centrale oplysninger i artiklerne fremtræder tydeligt. Fx bruges en prik til at markere, hvor eksemplerne begynder, mens en trekant signaliserer kollokation, et lighedstegn synonym og et ulighedstegn antonym.

Mediostruktur

Som sidste tilgængelighedsindikator vil jeg kort omtale mediostrukturen eller "henvisningsstrukturen". I forhold til flere af de andre indikatorer er det dog sparsomt, hvad der er af selvforståelsesoplysninger at hente.

- 1 retning
 - 11 fra
 - 111 lemma
 - 112 angivelse
 - 113 omtekst
 - 12 til
 - 121 lemma
 - 122 angivelse
 - 123 omtekst
 - 124 eksternt tekst
- 2 type
 - 21 se
 - 22 se også

De vigtigste oplysninger i de 21 værdi 1-omtekster går på henholdsvis henvisningens retning (1) og dens type (2). 1 kan underinddeles efter den lokalitet, der henvises fra (11), henholdsvis til (12). Kombinerer man fra og til, vil man se, at der for det første henvises **internt** i hoveddelen til et andet lemma eller en anden angivelse i en artikel, og for det andet **eksternt** fra hoveddel til omtekst eller vice versa, eller for det tredje eksternt i betydning ud af ordbogen, fx til en kildesamling.

Efter type skelnes der mellem de rene se-henvisninger (21) og de supplerende se også-henvisninger (22). Førstnævnte er den, man finder ved henvisningslemmata, der kendetegnes ved, at den ikke er tilordnet nogen angivelse. Det er også denne type, der optræder i de særlige henvisningsomtekster: registrene. Se også-henvisningerne bruges fx fra en angivelse til en omtekst, hvor der kan hentes yderligere oplysninger – eller fra et ord, der forekommer i en definitionsangivelse, og som selv er lemmatiseret i ordbogen.

I termordbøger er det almindeligt at henvide fra varianter og synonymer til den primære term. Kun dér vil termen fx blive defineret. Blandt de 8 værdi 2-omtekster er Jørgensen (1983) et sådant tilfælde.

I Kaufmann/Bergenholtz (1992) er brugen af henvisninger så hyppig, at de i omteksten kan skrive:

Ordbogsartiklerne afsluttes som regel med **henvisninger** til andre ordbogsartikler og/eller til den faglige indledning. (Kaufmann/Bergenholtz 1992:11)

og videre fastsættes principperne anvendelsen af henvisninger i et særligt afsnit:

Henvisninger findes kun i den dansk-engelske ordbogsdel. Henvisninger indledes med en pil -> som symbol, og der adskilles vha semikolon [når der er flere henvisninger, HKM] [...] Der gives tildels henvisninger i de tilfælde, hvor vi har de såkaldte henvisningsopslagsord, som kun er anført for at hjælpe brugeren hen til det mere brugte fagord. I andre tilfælde henvises der også til særligt relevante fagudtryk, der har betydning for forståelsen af det pågældende opslagsord. Endelig finder der henvisninger til særligt relevante paragraffer i den faglige indledning. [...] Der anføres højst fem henvisninger til andre opslagsord. De bliver anført på følgende måde: den vigtigste henvisning angives først, derefter følger de øvrige henvisninger i alfabetisk rækkefølge, til sidst står dog altid henvisningen/henvisningerne til den faglige indledning. (Kaufmann/ Bergenholtz 1992:23)

Afslutning

Hensigten med at se nærmere på omteksterne i de 155 bilingvale fagordbøger var at undersøge de skriftligt nedfældede overvejelser, som leksikograferne har gjort om processen fra idé, koncept og udarbejdelse frem til det færdige ordbogsprodukt. I dette empiriske materiale finder man eksplicite udtryk for den teoretiske selvforståelse, som hver enkelt ordbog bygger på. Dette materiale er værdifuldt i to henseender: For det første er det et godt udgangspunkt for afdækning af den samlede selvforståelse, der ligger bag den enkelte ordbog. Selvforståelsen, den eksplicite plus den implicite, er vigtig, hvis man vil evaluere en ordbog på dens egne præmisser. For det andet – og det er det, jeg har fokuseret på i denne artikel – kan den samlede eksplicite selvforståelse fra de 155 ordbøger ses som en repræsentation af leksikografernes viden om, hvad der skal til for at fremstille en god ordbog. De ni undersøgte kategorier er valgt ud, fordi jeg betragter dem som leksikografernes vigtigste bud på, hvordan ordbøger bliver anvendelige,

pålidelige og tilgængelige – tre krav, som vi som brugere med rimelighed kan stille til de ordbøger, vi bruger vores penge og tid på. Den afdækkede leksikografiske viden kan tjene som fundament og sammenligningsgrundlag, når man skal vurdere konkrete ordbøgers kvalitet. Endvidere kan den bruges som inspiration, når man på grundlag af et mangelfuldt eksplicit selvforståelsesmateriale skal uddrage den implicitte for at kunne danne sig et billede af den samlede selvforståelse.

Ordbogsmaterialet med tildelte points

Rækkefølgen af oplysninger følgende:

id * brugssituation (a) * brugerprofil (b) * afsenderintention (c) * fag (d) * sprog (e) * megastruktur (f) * makrostruktur (g) * mikrostruktur (h) * mediostruktur (i).

	a	b	c	d	e	f	g	h	i
Abrahamsen 1987	1	1	1	1	2	1	1	1	1
Am/Am 1984	1	1	0	0	1	0	0	1	0
AmAu60	0	0	0	0	0	0	0	0	0
Arguimbau 1992	1	1	0	1	1	0	1	1	0
AtOr72	0	1	0	1	0	0	0	0	0
Bach/Florant 1987	1	0	0	0	0	0	0	1	0
Bailey 1972	1	0	0	1	1	1	1	1	1
Bakker 1958	1	1	0	1	0	1	1	0	0
Balahang H. 1989	0	1	0	0	0	0	0	0	0
Balahang N. 1989	0	1	0	0	0	0	0	0	0
Bech 1969	0	0	1	1	0	0	0	0	0
Behrens 1972	1	1	0	1	1	1	1	1	1
Bolt 1988	1	1	0	1	0	0	1	1	2
Bondy/Jensen/Jørgensen/Keiding/ Kristensen 1982	1	1	1	2	1	0	1	1	0
Bredsdorff 1960	0	0	0	0	0	0	0	0	0
ByAn86	2	1	1	1	0	0	0	0	0
Chiffre 1976	0	0	0	0	0	0	0	0	0
Christensen/Sibbernsen 1988	1	1	1	1	1	1	0	1	0
Christensen 1981	0	0	0	0	0	0	0	0	0
Claudi 1931	1	1	0	0	2	0	0	0	0

Conradsen 1984	1	1	0	0	1	0	1	0	0
Crewe/Mogensen 1992	1	1	0	1	2	1	1	2	0
Cristofoli/Kock 1983	1	1	0	0	0	0	0	1	0
DaDa85	1	1	2	1	2	1	1	1	1
DaEn67	1	1	0	1	1	0	0	0	0
DaEn79	1	1	0	1	0	0	0	1	0
DaEn82a	1	0	1	1	0	0	1	1	0
DaEn83	0	0	1	1	0	0	1	1	1
DaEn90	1	0	0	2	1	0	2	2	1
DaFr89	0	0	0	2	0	0	2	2	0
DaLa77	1	1	0	0	0	1	0	0	0
Dalhoff/Grell 1990	1	1	0	1	1	1	1	1	1
Dang 1985	0	0	0	0	0	0	0	1	0
DaTy90	1	0	0	2	1	0	2	2	0
DsEl69	0	0	1	1	0	1	1	2	0
DsIn72	0	0	0	0	0	0	1	1	0
Einersen 1966	1	0	0	0	0	0	0	0	0
EnDa60	1	1	1	0	0	0	0	0	0
EnDa69	0	0	0	0	0	0	0	0	0
EnDa72	1	1	0	1	0	0	0	0	0
EnDa78	0	1	0	0	0	0	0	0	0
EnDa79	1	1	1	2	0	0	1	1	0
EnDa83	0	0	2	1	1	0	1	1	1
EnDa85	0	0	0	0	0	0	0	0	0
EnDa88	1	1	0	1	1	1	1	1	0
EnDa91	1	1	1	1	1	1	1	1	0
EnDa91a	1	0	1	2	1	0	2	2	0
EnOr70	0	0	0	0	0	0	0	0	0
Eriksen/Helms/Rømer 1979	1	1	2	2	1	1	2	2	2
Eskildsen 1931–32	1	0	0	0	1	0	0	0	0
FaDa92	1	1	1	1	1	0	0	1	0
FaDa92a	1	1	1	1	1	0	0	1	0
Forbert/Fredensborg/Lemire 1980	1	1	1	1	1	1	1	1	1
Frank 1984	1	1	0	1	1	0	0	2	2
FrDa90	1	0	0	2	1	0	2	2	0
Fryd 1974	1	0	1	1	0	1	0	1	0
Fryd 1977	1	0	0	1	1	1	0	1	0
Fryd 1990	0	0	0	1	0	0	0	0	0
Galindo 1979	1	1	0	1	1	0	1	2	0
Galindo 1989	1	1	1	1	1	0	0	1	0
Galindo 1990	1	1	1	1	1	0	0	1	0
GjDa23	0	0	0	0	1	1	0	1	0
Glahn 1990	1	1	0	1	0	1	1	1	1
GrDe91	1	1	0	1	1	1	0	1	0
Gubba 1984	0	1	0	1	0	0	1	1	1
Gubba 1987	1	1	0	1	1	0	1	1	1

Haffner 1976	1	0	0	1	0	0	0	1	0
HaKo23	0	1	1	0	0	0	1	0	0
Hannibal 1958	0	1	0	1	1	0	0	0	0
Hansen 1988	1	1	1	0	2	1	2	1	1
Hansen/Hinrichsen 1968	0	1	1	1	0	0	0	1	0
Hansen 1974	0	0	0	0	0	1	1	0	0
Hansen 1985	0	0	0	1	0	0	0	1	1
Hansen 1957	0	1	1	1	1	1	1	2	1
Hasselbalch/Neal/Hasselbalch 1984	1	1	0	1	2	0	1	2	0
Hauschildt 1988	0	1	1	1	2	0	1	2	0
Henriksen 1988	0	0	0	1	0	0	0	1	0
Henriksen 1991	2	1	1	2	1	1	1	2	2
Hjort 1977	1	0	0	1	0	1	0	0	0
Hubbard 1970	1	1	2	0	2	0	0	2	0
ImSt87	1	1	1	1	2	0	0	1	0
InFi90	0	0	1	0	1	0	0	1	2
InFo83	0	0	0	1	0	0	0	1	0
InOr91	1	0	1	1	0	0	1	1	1
InRo70	0	0	0	0	0	0	1	0	0
InTo72	0	0	0	0	0	0	0	0	0
Jacobsen/Toke 1955	1	1	2	1	2	0	2	2	0
Jensen 1923	1	1	1	1	2	0	1	2	2
Jochumsen 1983	0	0	0	0	0	0	0	1	0
Jonsson 1988	0	0	0	0	1	0	0	1	0
JoVe89	1	1	0	0	1	1	2	2	1
Jørgensen/Jordal/Andersen/Smith 1986	1	2	1	2	2	2	2	2	0
Jørgensen 1983	1	0	0	1	2	1	1	1	2
Kaufmann/Bergenholtz 1992	2	2	2	2	2	2	2	2	2
Koch 1982	1	0	1	1	1	0	0	1	0
Krogh-Hansen 1985	1	1	1	1	1	1	0	1	1
KuFø13	1	0	0	1	0	0	1	0	0
Kønig/Ottesen 1901	1	0	0	0	1	0	0	1	0
LaLe91	1	1	0	1	1	0	0	1	0
Lange 1941	0	0	0	0	0	0	1	2	0
Larsen 1914	0	0	0	1	0	0	1	1	0
Lassen 1964	1	1	0	1	0	0	1	0	0
Linder 1965	0	0	1	1	1	0	0	1	0
Lohmann 1956	1	1	0	1	0	1	1	1	0
Ludvigsen 1913	2	1	0	2	2	2	0	1	0
Ludvigsen 1919	2	1	0	2	2	2	1	1	0
Maul 1944	2	2	0	1	1	0	0	1	1
MaVæ91	1	1	0	1	1	0	0	1	0
Merrild 1941	1	1	0	1	0	0	0	1	0
Messerschmidt 1968	1	1	0	1	0	1	1	1	0
MiDa87	1	1	0	1	0	0	0	0	0
Mohamad 1989	1	0	0	0	0	0	0	0	0

Moth-Lund A. 1933	1	1	1	1	0	1	0	1	0
Moth-Lund D. 1933	1	1	1	1	0	1	0	1	0
Møllmann 1990	1	1	1	1	0	1	1	1	0
Nellinge 1974	1	0	0	0	0	0	0	0	0
Nielsen 1979	0	0	0	1	1	1	1	1	0
Nielsen 1975	1	1	1	0	1	0	0	1	0
NyLi86	0	0	0	0	0	0	0	0	0
Nørager 1987	1	0	0	1	1	1	0	1	0
Oppenheim 1933	1	1	0	0	1	1	1	1	0
Oppenheim 1954	0	0	0	0	0	0	0	0	0
OrEc77	1	1	0	1	1	0	1	1	1
Pedersen 1940	0	0	0	0	0	1	0	0	0
Porsild 1974	1	1	0	1	0	0	1	1	0
Porsild 1986	0	0	0	0	0	0	0	0	0
Poulsen 1987	1	1	1	1	2	1	1	2	0
Rosbach/Thornfield 1981	1	1	1	0	1	0	0	1	0
Rousing 1976	0	0	0	0	2	1	0	0	0
Scheller 1913	0	0	0	1	0	0	0	0	0
SkDa65	1	1	1	0	0	0	0	0	0
SkDa68	0	1	0	0	0	0	0	0	0
Skibsted 1971/1980	1	1	1	1	0	0	1	2	0
Skjerk 1974	1	1	0	1	0	1	0	0	0
Skjerk 1991	1	1	1	0	0	1	1	1	0
Stürup/Eilertzen 1977	1	1	0	1	0	1	0	1	0
Søndergaard 1987	0	0	0	1	1	0	1	1	0
TeLe68-77	1	1	1	1	0	0	0	0	0
TeOr87	0	0	0	1	1	0	1	2	0
TeRe50	0	0	2	1	0	1	0	1	1
Terkelsen 1977	1	0	0	1	0	0	0	1	0
Thagesen 1985	0	1	1	0	0	1	0	1	0
Thordal-Christensen 1991	1	1	0	1	2	1	0	1	0
TiPI77	0	0	0	0	0	1	0	0	0
TyDa69	0	0	0	0	0	0	0	0	0
TyDa91	1	0	1	2	0	0	2	2	0
Unnerup-Madsen 1989	1	1	0	1	1	0	1	1	0
Unnerup-Madsen 1991	1	1	0	1	1	1	0	1	0
Vesterli 1988	1	1	1	1	0	0	0	1	0
Warrern 1983	0	0	0	1	1	0	1	1	0
Warrern 1984	0	0	0	1	1	0	0	1	0
Warrern 1986	0	0	0	1	1	0	1	1	0
Warrern 1988	0	0	0	1	1	0	1	1	0
Ærsøe 1926	0	0	0	1	0	0	0	0	0
ØkOr86	1	1	0	1	0	0	1	1	0

Analyserede fagordbøger

- Abrahamsen 1987 = John Abrahamsen: *Gads edb ordbog*. København: Gad.
- Am/Am 1984 = Marie-Jehanne Am/Willi Am: *Mad i Frankrig. Mad-ordbog, fransk/dansk, dansk/fransk*. Brabrand: Eget forlag.
- AmAu60 = *950 amerikanske-engelske automobil-fagudtryk oversat til dansk*. General Motors International A/S, Service-Afdelingen 1960 [ca.].
- Arguimbau 1992 = Damián Arguimbau: *Forlagsordbog. Dansk-engelsk, engelsk-dansk*. København: Forening for Forlagsfolk.
- AtOr72 = *Atomordbog*. Red.: T. Bjerge. Risø: Atomenergikommissionen. Forsøgsanlæg Risø 1972.
- Bach/Florant 1987 = Helmut Bach/Jørgen Florant: *Luftfartsteknisk ordbog: engelsk-dansk (2.udg.)*. København: Jernindustriens Forlag.
- Bailey 1972 = Ib Emil Bailey: *Dansk-engelsk handels- og fagordbog for erhvervslivet, administrationen og forvaltningen*. København: Schønberg.
- Bakker 1958 = André F. Bakker: *Here you are (teknisk lommeordbog eng.-da. og da.-eng.)*. København: Teknologisk institut.
- Balahang H. 1989 = Hani Balahang: *Dansk-persisk medicinsk ordliste*. Århus: Forskel lighed.
- Balahang N. 1989 = Noori Balahang: *Dansk-arabisk medicinsk ordliste*. Århus: Forskel lighed.
- Bech 1969 = Per Otto Bech: *Regnskabs- og revisionsudtryk. Engelsk-dansk, dansk-engelsk*. København: Gyldendal.
- Behrens 1972 = Thorkild Behrens: *Media-leksikon. Dansk-engelsk, engelsk-dansk*. København: Chr. Erichsen.
- Bolt 1988 = Dieter Bolt: *Min edb-ordbog (1. udg., 1. opl.)*. Red.: Poul Erik Jensen. København: Munksgaard.
- Bondy/Jensen/Jørgensen/Keiding/Kristensen 1982 = Vivian Bondy/Hans Jensen/Kurt Jørgensen/Jens Th. Keiding/Bente Kristensen: "Dansk-engelsk ergonomi-glosar" (1. udg.). I: *DS Informationsblad* 16. Hellerup: Dansk Standardiseringsråd.
- Bredsdorff 1960 = Ruth Bredsdorff: *Bygge-ordbog. Dansk-engelsk, engelsk-dansk*. København: Nyt Nordisk Forlag, Arnold Busck.
- ByAn86 = *Bygge- og anlægsteknisk ordbog, 1-4 (1-2: Dansk-engelsk, 3-4: Engelsk-dansk)*. Red.: Jørgen Høedt. København: IFF 1986.
- Chiffre de 1976 = Leonardo de Chiffre: *Gloseliste i teknisk sprog til brug for EA/ED-studiet. Dansk-italiensk, italiensk-dansk*. København: Handelshøjskolen.
- Christensen/Sibbernsen 1988 = Benny Winther Christensen/Bente Sibbernsen: *Handelsfaglig ordbog. Dansk-engelsk*. Horsens: Åløkke.

- Christensen 1981 = Flemming Christensen: *Definitioner og ordforklaringer til farmakologi* (Nyt opl.). Århus: Aarhus Odontologiske Boghandel.
- Claudi 1931 = Alb. Claudi: *Kortfattet Dansk-Engelsk Ordsamling, udarbejdet til Brug ved Forberedelsen til den udvidede Maskinisteksamen og tekniske Skoler*. København: Saxtorph-Beck.
- Conradsen 1984 = Inger Conradsen: *Økonomisk ordbog. Dansk-italiensk, italiensk-dansk* (1. udg.). Hørsholm.
- Crewe/Mogensen 1992 = Inger Maria Crewe/Anne Dorte Mogensen: *Handelsfaglig ordbog, dansk-spansk* (1. udg.). Horsens: Åløkke.
- Cristofoli/Kock 1983 = Mirella Cristofoli/Merete Kock: *Glosar i kommercielt sprog. Dansk-italiensk*. København: HHK.
- DaDa85 = *Dansk dataordbog*. Red.: Carlos Mora-Jensen. Ølstykke: Medicinsk forlag 1985.
- DaEn67 = *Dansk-engelsk civilforsvarsordbog*. København: Civilforsvarsstyrelsen 1967.
- DaEn79 = *Dansk-engelsk ordbog for civilforsvaret og det civile beredskab* : Civilforsvarsstyrelsen 1979.
- DaEn82a = *Dansk-engelsk uddannelsesterminologiliste*. København: Undervisningsministeriet. Det internationale kontor 1982.
- DaEn83 = "Dansk-engelsk elektroteknisk ordbog. Danish-English electrotechnical vocabulary". I: *DS Informationsblad* 20.1. København: Dansk Elektroteknisk Komité. Dansk Standardiseringsråd 1983.
- DaEn90 = *Dansk-engelsk teknisk ordbog* (1. udg.). Red.: Dagny Dahl Christensen et al. København: Fonden for Fagsproglig Leksikografi, HHK/Grafisk Forlag, Clausen 1990.
- DaFr89 = *Dansk-fransk teknisk ordbog* (1. udg.). København: Fonden for Fagsproglig Leksikografi, Handelshøjskolen i København. Aschehoug Dansk Forlag, Clausen Bøger 1989.
- DaLa77 = *Danske og latinske navne: kulturplanter, ukrudtsplanter, træer og buske, plantesygdomme, skadedyr*. 3. oplag. Red.: Niels Pedersen. Viby J.: Landskontoret for Planteavl 1977.
- Dalhoff/Grell 1990 = Jørgen Dalhoff/Henrik Grell: *Samfundsøkonomisk minilex* (2. udg., 2. opl.). København: Gyldendal.
- Dang 1985 = Dang Van Nham: *Dansk-vietnamesisk medicinsk billedordbog* (1. udg., 1. opl.). Tåstrup.
- DaTy90 = *Dansk-tysk teknisk ordbog* (1. udg.). København: Fonden for Fagsproglig Leksikografi, HHK. Grafisk Forlag 1990.
- DsEl69 = DS 5004.08: Elektroakustik. I: *Elektroteknisk ordliste. Dansk elektroteknisk norm. Udarb. og vedtaget af Dansk elektroteknisk Komité (Dansk standard, DS 5004, 5006, 5007)* 1969.

- DsIn72 = DS 5004.45: International belyningsordbog (1. udg.). I: *Elektroteknisk ordliste. Dansk elektroteknisk norm. Udarb. og vedtaget af Dansk elektroteknisk Komité (Dansk standard, DS 5004, 5006, 5007)*. Dansk Standardiseringsråd 1972.
- Einersen 1966 = Henning Einersen: *Tysk-dansk ordliste over landbrugsfaglige gloser*. København: Det kgl. danske Landhusholdningsselskab.
- EnDa60 = *Engelsk-dansk, dansk-engelsk reklameordbog*. Red.: Gunnar K. Jørgensen. Charlottenlund: Kirschbaum-Jørgensen 1960.
- EnDa69 = Engelsk-dansk ordliste over skatteretlige udtryk. I: *Meddelelser fra Skattedepartementet 10*. København 1969.
- EnDa72 = *Engelsk-dansk ordliste til nationaløkonomiske fagudtryk*: Redigeret af Institut for Nationaløkonomi ved Handelshøjskolen i København. København: Samfundslitteratur 1972.
- EnDa78 = Engelsk-dansk ordnøgle for servicefolk. Oversættelser til dansk af almindeligt forekommende ord og forkortelser i engelsksprogede servicevejledninger. I: *Bedre lyd og billede 2*. Hellerup: Radiobranchens Fællesråd, Teknisk Sekretariat 1978.
- EnDa79 = *Engelsk-dansk økonomisk ordliste*. (2. udgave, 1. oplag). Red.: Vagn Andersen/Jørgen Steen Hansen. Brabrand: Andersen og Hansen 1979.
- EnDa83 = Engelsk-dansk elektroteknisk ordbog. English-Danish electrotechnical vocabulary. Udarbejdet og vedtaget af Dansk Elektroteknisk Komite, DEK. (1. udg.). I: *DS Informationsblad 20.2*. København: Dansk Standardiseringsråd 1983.
- EnDa85 = *Engelsk-Dansk bilteknisk ordbog*. Ebeltoft: Bjarne Nielsen 1985.
- EnDa88 = *Engelsk-dansk merkantil/økonomisk ordbog* (1. udg., 1. opl.). Red.: Lise Damsgaard, Annemette Lyng Svensson. København: Samfundslitteratur 1988.
- EnDa91 = *Engelsk-dansk økonomisk ordbog* (2. udg.). Red.: Annemette Lyng Svensson. Frederiksberg: Samfundslitteratur 1991.
- EnDa91a = *Engelsk-dansk teknisk ordbog* (1. udg.). Red.: Dagny Dahl Christensen. København: Fonden for Fagsproglig Leksikografi, HHK. Grafisk Forlag, Clausen 1991.
- EnOr70 = *Engelsk-dansk ordbog for auto-reservedelspersonale*. København: Reservedelsafdelingen, General Motors International A/S 1970.
- Eriksen/Helms/Rømer 1979 = Bjørn Eriksen/Hans Jørgen Helms/Mogens D. Rømer: *Edb-ordbog. Dansk standard; DS 2049*. 1. udgave, 4. oplag. København: Gjellerup.

- Eskildsen 1931–32 = Martin Peter Eskildsen: "Fortegnelse over luftfartstekniske Udtryk og Betegnelser". I: *Flyv.* København: Udg. paa Foranledning af Det kongelige danske aeronautiske Selskab – Aug. Bang.
- FaDa92 = *Fagordbog, dansk-engelsk, engelsk-dansk. 1–2.* (Ny gennemrevideret udgave). Red.: Jørgen Høedt, Charlotte Langkilde, Karen Wolf-Frederiksen. København: L&H Ordbøger 1992.
- FaDa92a = *Fagordbog, dansk-tysk, tysk-dansk. 1–2.* Red.: Thomas Nielsen. København: L&H Ordbøger 1992.
- Forbert/Fredensborg/Lemire 1980 = Maria M. Forbert/Jette Fredensborg/Jean-François Lemire: *Dansk-fransk teknisk ordbog: værkstedsteknik = Dictionnaire danois-français des fabrications mécaniques.* Tåstrup: Teknologisk Instituts Forlag, Udlandsafdeling.
- Frank 1984 = Lars Frank: *Edb-ordbogen.* (1. udg., 1. opl.). København: Kommunetryk.
- FrDa90 = *Fransk-dansk teknisk ordbog* (1. udg.) Red.: Hanne Blaaberg, Gitte Grønbech, Birgitte Lauterbach. København: Fonden for Fagsproglig Leksikografi, HHK. Grafisk Forlag 1990.
- Fryd 1974 = Ejnar Fryd: *Tysk-dansk, dansk-tysk specialordbog indenfor revision, regnskabsvæsen m.v.* København: Foreningen af Statsautoriserede Revisorer.
- Fryd 1977 = Ejnar Fryd: *Fransk-dansk, dansk-fransk specialordbog inden for revision, regnskabsvæsen m.v.* København: Foreningen af statsautoriserede revisorer.
- Fryd 1990 = Ejnar Fryd: *Anglo/amerikansk-dansk, dansk-anglo/amerikansk specialordbog inden for revision, regnskabsvæsen m.v.* (3.udg. / revideret og ajourført af Per Otto Bech, 2. opl.). København: FSRs Forlag.
- Galindo 1979 = Orlando Galindo: *Dansk spansk fagordbog = Diccionario especializado danés-español.* København: Teknisk forlag.
- Galindo 1989 = Orlando Galindo: *Erhvervs- og økonomiordbog. Diccionario Industrial y Económico. Dansk-spansk.* København: GTO.
- Galindo 1990 = Orlando Galindo: *Erhvervs- og økonomiordbog. Dictionnaire économique et de l'entreprise. Dansk-fransk. Danois-Français.* København: GTO.
- GjDa23 = *Gjellerups Dansk-Tyske Handelsordbog.* Red.: Vilh. Ludvigsen. København: Gjellerup 1923.
- Glahn 1990 = Marianne Stürup Glahn: *Det franske menukort – en ordbog fra a til z.* København: P. Haase & Sønns Forlag.
- GrDe91 = *Grafisk/desktop ordbog, dansk-engelsk, engelsk-dansk.* Red.: Karen Wolf-Frederiksen/Erik M. Skovfoged. Herlev: L&H 1991.

- Gubba 1984 = Wilhelm Gubba: *Juridisk ordbog: 1: Dansk-tysk* (2. udg.). Mårslet: W. Gubba.
- Gubba 1987 = Wilhelm Gubba: *Juridisk ordbog, 2: Tysk-dansk* (2. udg.). Mårslet: W. Gubbas Forlag.
- Haffner 1976 = Charlotte von Haffner: *Juridisk ordliste, dansk-spansk*. København: Handelshøjskolen. Institut for spansk.
- HaKo23 = *Handelskorrespondenten. Dansk-Engelsk. Skrevet af en praktisk Købmand [N. Lund] til Brug for praktiske Købmænd*. Aarhus: Forlaget Dania 1923.
- Hannibal 1958 = A. Hannibal: *Automobilordbog. Engelsk. Amerikansk. Dansk*. København: Ivar.
- Hansen 1988 = Børge Hansen: *Hjælpebog for engelske korrespondenter. Dansk-engelsk erhvervsproglig synonymordbog*. (4. rev. udg. ved Jørgen Alsø). København: Schønberg.
- Hansen/Hinrichsen 1968 = Helge Hansen/Hans Hinrichsen: *Engelsk-dansk teknisk ordbog*. København: Schønberg.
- Hansen 1974 = Jens E. Hansen: *Tysk/dansk bilteknisk ordbog*. Ringsted: Lentz og Jenssen.
- Hansen 1985 = Mogens Herman Hansen: *Liste over græske fagudtryk. Supplement til Det athenske demokrati i 4. årh. f. Kr.* København: Museum Tusulanum.
- Hansen 1957 = Otto Hansen: *Handelshåndbogen. Politikens handelsleksikon*. Red.: Jan Kobbarnagel. København: Politikens Forlag.
- Hasselbalch/Neal/Hasselbalch 1984 = Gunver Hasselbalch/Alan Neal/Ole Hasselbalch: *Arbejdsretlig terminologi dansk-engelsk*. København: Dansk Arbejdsgiverforening.
- Hauschildt 1988 = Gurli Hauschildt: *Handelsfaglig ordbog: dansk-fransk*. Horsens: Ålørkke.
- Henriksen 1988 = Torben Henriksen: *Glosario de terminología jurídica español-danés*. (Ed. ampl. y rev.). Århus: HHÅ, Spansk Institut.
- Henriksen 1991 = Torben Henriksen: *Juridisk ordbog spansk-dansk*. København: G.E.C. Gads Forlag.
- Hjort 1977 = Øystein Hjort: *Ordbog til byzantinsk kunst*. København: Institut for kunsthistorie, Københavns Universitet.
- Hubbard 1970 = L. Ron Hubbard: *Den lille scientology ordbog*. Overs. ved Inger Budtz-Jørgensen. Rev.: Bene Kjellerup/Niels Kjellerup/Gøte Mathiesen. København: Scientology Kirken, Danmark.
- ImSt87 = *IMO's Standard Marine Navigational Vocabulary. = Maritime standardudtryk*. København: Industriministeriet 1987.
- InFi90 = *Internationale finansielle udtryk og forkortelser* (2. udg.). København: Den Danske Bank 1990.

- InFo83 = *Internationale forkortelser*. København: Udenrigsministeriets Handelsafdeling 1983.
- InOr91 = *Industriens ordbog. Tekniske termer. 1–2. 1: Dansk/engelsk. 2: Engelsk/dansk (1. udg.). I: DS Informationsblad 62:1–2*. Hellerup: Dansk Standardiseringsråd 1991.
- InRo70 = *International road research documentation (IRRD): Alfabetisk dansk-engelsk liste over nøgleord*. København: Statens Vejlaboratorium 1970.
- InTo72 = *Indkøbs-tolk. Shopping-guide. Kupovni tumaã. Alisveris Tercümani. Kharidari ka Rahnuma*. Pakistansk tolk: Masood Razak. Tyrkisk tolk: Selma Jacobsen. Jugoslavisk tolk: Srdjan Vidović. Tegninger: Karin Bartholm. København: Statens Husholdningsråd 1972.
- Jacobsen/Toke 1955 = Aage Jacobsen/Egil Toke: *Dansk-engelsk handelsordbog* (1. opl.). København: Schønberg.
- Jenssen 1923 = J. Jenssen: *Ordbog for Gartnere og Botanikere, med Vejledning til Forstaaelse af de ved Planternes Betegnelse forekommende græske og latinske Slægts-, Arts- og Varietetsnavne samt botaniske Kunstudtryk* (3. Udg. med et Tillæg). København: G.E.C. Gads forlag.
- Jochumsen 1983 = Svend Jochumsen: *Teknisk ordbog. Dansk-spansk*. København: Gyldendal/Dansk historisk Håndbogsforlag.
- Jonsson 1988 = Marit Jonsson: *Islandske hestnavne* (2. rev. udg., 3. opl.). Hillerød: Dansk Islandshesteforening.
- JoVe89 = *Joint venture. Engelske og danske termer* (1. udg.). Red.: Rita Lenstrup. København: Samfundslitteratur 1989.
- Jørgensen/Jordal/Andersen/Smith 1986 = Dolly Jørgensen/Tove Jordal/Jan Andersen/Viktor Smith: *Naturgassystemer på russisk og dansk. Terminologisk analyse og opslagsværk*. København: Samfundslitteratur.
- Jørgensen 1983 = Jørgen Jørgensen: *Danske navne på skade- og nytte-dyr på planter (udarbejdet for Dansk plantepatologisk Selskab)*. København: DSR Forlag.
- Kaufmann/Bergenholtz 1992 = Uwe Kaufmann/Henning Bergenholtz: *Genteknologisk ordbog. Dansk-engelsk/engelsk-dansk molekylærbiologi og DNA-teknologi*. København: Gad.
- Koch 1982 = Hans Christian Dahlerup Koch: *Offshoreordbog. Et opslagsværk om oliebranchens sprog, talemåder og slang* (1. opl.). Esbjerg: RB-Kommunikation. Varde Bank/Offshore.
- Krogh-Hansen 1985 = Niels Krogh-Hansen: *Fransk-dansk ordbog til oversættelser på formuerettens område med tilhørende dansk-fransk ordliste* (1. udg.). København: HHK.

- KuFø13 = *Kulinær fører i Paris. Hvorledes man oversætter den franske menu. Hvor man spiser godt og billigt i Paris.* København: E. Jespersen 1913.
- Kønig/Ottesen 1901 = G.A.F. Kønig/Chr. Ottesen: *Marine-teknisk Ordsamling. Engelsk-dansk. Dansk-engelsk.* København: Gad.
- LaLe91 = *Landbrug/levnedsmiddel ordbog, dansk-engelsk, engelsk-dansk.* Red.: Charlotte Langkilde/Thomas Nielsen/Karen Wolf-Frederiksen. Herlev: L&H 1991.
- Lange J. 1941 = Johan Lange: *Havebrugslatin og ordliste.* København.
- Larsen 1914 = H.C. Larsen: *Artsnavne. I: Tillæg til General-Register til Tidsskrift for Landbrugets Planteavl 1–20 Bd. og Tidsskriftet om Landbrugets Kulturplanter og dertil hørende Frøavl.* København: Nordisk.
- Lassen 1964 = H.C.A. Lassen: *Medicinsk ordliste. Hvad hedder det på fransk?* København, Léopoldville: Det danske undervisningshospital i Léopoldville, Congo (Dansk Røde Kors).
- Linder 1965 = Bernhard Linder: *Svensk dansk teknisk ordbog.* 2. opl. København: NKI-Skolen i samarbejde med Munksgaard.
- Lohmann 1956 = Chr. Lohmann: *Dansk-spansk sætningsbog.* København: Hirschsprung.
- Ludvigsen 1913 = Vilhelm Ludvigsen: *Tysk-Dansk og Dansk-Tysk Handelsordbog. Bd. 1–2.* København: Gjellerup.
- Ludvigsen 1919 = Vilhelm Ludvigsen: *Engelsk-Dansk og Dansk-Engelsk Handelsordbog.* København: Hasselbalch.
- Maul 1944 = Magnus Maul: *Engelsk-dansk Automobilordbog.* København: Branner.
- MaVæ91 = *Maskin- og værktøjsteknisk ordbog, dansk-engelsk, engelsk-dansk.* Red.: Thomas Nielsen/Karen Wolf-Frederiksen. Herlev: L&H 1991.
- Merrild 1941 = Karl Merrild: *Ballet-Ordbog. Franske Ord og Gloser, der fortrinsvis benyttes til Ballettrin og Balletinstudering. Samlet og arrangeret til mine Elever.* (Trykt som Manuskript). København.
- Messerschmidt 1968 = B. Messerschmidt: *Lærebog i militært fransk for Hærens Officersskole + Ordliste.* København: Hærens Officersskoles Forlag.
- MiDa87 = *Miljøordbog. Dansk engelsk, engelsk dansk, 1–2.* Red.: Charlotte Langkilde/Charlotte Riis-Olsen/Kathleen Grave. København: IFF Ordbøger.
- Mohamad 1989 = Sabri Khaled Mohamad: *Dansk-arabisk teknisk ordbog* (under medvirken af Iver Larsen). Århus: Dan-Arab Forskelighed.

- Moth-Lund A. 1933 = Aug. Moth-Lund: *Dansk-fransk Sætningsbog*. København: Hirschsprung.
- Moth-Lund D. 1933 = D. Moth-Lund: *Dansk-tysk Sætningsbog*. København: Hirschsprung.
- Møllmann 1990 = Anne Møllmann: *Dansk-fransk marketing ordbog*. København: Udgivervirksomheden FUHU (Foreningen til Unge Handelsmænds Uddannelse).
- Nellinge 1974 = Solveig Nellinge: *Engelsk/amerikansk-dansk forlægsordbog*. (Oversættelse fra svensk og bearbejdelse og udvidelse ved Erik C. Lindgren). København: Lindhardt og Ringhof.
- Nielsen 1979 = Ellen Würtz Nielsen: *Dansk-engelsk ordliste specielt relateret til det danske sundhedsvæsen*. København.
- Nielsen 1975 = Feodor Nielsen: *Anglo-amerikansk/dansk retsordbog*. København: Erhvervsøkonomisk Forlag (Nyt Nordisk Forlag).
- NyLi86 = "Den nye lille grønne". *Husdyrbrugerens opslagsbog. Sygdomsforebyggende behandlinger* (7. udg., gennemarbejdet og moderniseret af "L.v.K."s veterinære planlægningsudvalg). Hobro: Landbrugets veterinære Konsulenttjeneste (L.v.K.). 1986.
- Nørager 1987 = Per Nørager: *Miniordbog i finansieringsengelsk. Engelsk-dansk/dansk-engelsk* (1. udg., 1. opl.). København: Danmarks Sparekasseforening og Samfundslitteratur.
- Oppenheim 1933 = Harriet Oppenheim: *Dansk-engelsk Sætningsbog*. København: Hirschsprung.
- Oppenheim 1954 = H. Oppenheim: *Dansk-Engelsk, Engelsk-Dansk turist-ordliste. English-Danish, Danish-English Tourist Vocabulary*. København: Hirschsprung.
- OrEc77 = *Ordlister - economics. Engelsk-dansk ordliste over økonomiske fagudtryk*. (3. udg.) Red. af Institut for Nationaløkonomi ved Handelshøjskolen i København. København: Samfundslitteratur 1977.
- Pedersen 1940 = Erna Pedersen: *Filatelistisk Ordbog*. København: Helios (Reitzel).
- Porsild 1974 = Per Porsild: *Dansk / engelsk - engelsk / dansk landmilitær ordliste*. (Udarbejdet til brug ved undervisningen i engelsk på Hærens Officersskoles B-Linie). Frederiksberg Slot: Officersskolens forlag.
- Porsild 1986 = Per Porsild: *Hunde engelsk* (1. opl.). Virum: Peto.
- Poulsen 1987 = Sven-Olaf Poulsen: *Dansk-tysk handelsordbog* (4. udg.). Harlev: A.M. Kjærs Forlag.
- Rosbach/Thornfield 1981 = P. Rosbach/W. Thornfield: *Opslagsbog for tyske korrespondenter* (2. udg., 2. opl.). København. Schönberg.

- Rousing 1976 = A. Rousing: *Engelske forholdsforbindelser anvendt i handelskorrespondance*. Aarhus: Handelsvidenskabelig Boghandel.
- Scheller 1913 = C.F. Scheller: *Dictionnaire de termes nautiques. Français-danois et danois-français. Marine-Ordbog. Fransk-Dansk og Dansk-Fransk*. København: Gyldendal.
- SkDa65 = *Skibsudtryk på dansk-engelsk, engelsk-dansk*. Nakskov: A/S Nakskov Skibsværft 1965.
- SkDa68 = *Skibsudtryk på dansk-tysk, tysk-dansk*. Nakskov: A/S Nakskov Skibsværft 1968.
- Skibsted 1971/1980 = Sigurd Skibsted: *Teknisk engelsk-dansk ordbog* (Fotografisk genoptryk). Lyngby: Polyteknisk Forlag.
- Skjerk 1974 = Ebbe Skjerk: *Engelsk/Dansk bilteknisk ordbog*. Ringsted: Lentz og Jenssen.
- Skjerk 1991 = Ebbe Skjerk: *Bilteknisk ordbog. Engelsk-Dansk* (2. udg., 1. opl.). København: Teknisk Forlag.
- Stürup/Eilertzen 1977 = Marianne Stürup/Jens G. Eilertzen: *Fransk-dansk menuguide: menuforklaringer og beskrivelser*. Farum: Bent Carlsen.
- Søndergaard 1987 = Margrethe Søndergaard: *Ordliste til brug i CF-fagligt engelsk. Dansk/engelsk udgave*. Bernstorff Slot: Civilforsvarsstyrelsen.
- TeLe68–77 = *Teknisk leksikon. 1–5. 1: Maskinteknik, 2. Trafikmidler. Bygningsingeniørvæsen, 3. Matematik og fysik. Svagstrøm. Stærkstrøm. Automatisering. 4. Kemi. Kemoteknik. Tværfaglige emner. 5. Alfabetisk register med opslagsord, henvisninger, supplerende oplysninger*. Red. Harald W. Møller (1–2), B. Lund Poulsen (1–5), K.O.B. Jørgensen (3–5). København, Oslo: Forlaget for faglitteratur a/s 1968–77.
- TeOr87 = *Teknisk ordbog. Dansk-russisk. Ca. 55000 fagudtryk Datsko-russkij politechniceskij slovar'. Okolo 55000 terminov*. Red.: V.F. Maksimov, A.A. Kondraskov. Moskva/København: Russkij jazyk/Gyldendal: 1987.
- TeRe50 = Terminologi vedrørende reaktionsdrift af luftfartøjer. I: *Terminologicalentralens meddelelser (TC) 1*. København: Terminologicalentralen 1950.
- Terkelsen 1977 = Poul Terkelsen: *Ordforklaring til Statut for Sovjetunionens kommunistiske Parti*. Århus: Slavisk Institut, Aarhus Universitet.
- Thagesen 1985 = B. Thagesen: *PIARC's vejtekniske ordbog. I: Notat. Institut for veje, trafik og byplan*. Lyngby.

- Thordal-Christensen 1991 = Anna Thordal-Christensen: *Dansk-fransk handelsordbog. "Pour de bonnes raisons". Med hovedvægt på forholdsordsforbindelser* (1. udg., 2. opl.). Holstebro: Ventus.
- TiPl77 = *Tillæg til plast-ordliste engelsk-dansk dansk-engelsk (DS 2059)*. København: Dansk Standardiseringsråd 1977.
- TyDa69 = *Tysk-dansk ordbog for auto-reservedelspersonale*. København: Reservedelsafdelingen, General Motors Danmark 1969.
- TyDa91 = *Tysk-dansk teknisk ordbog* (1. udg.) København: Fonden for Fagsproglig Leksikografi, Handelshøjskolen i København/Grafisk 1991.
- Unnerup-Madsen 1989 = Dorthe Unnerup-Madsen: *Dansk-tysk handelsordbog "Aus vielen Gründen" med hovedvægt på forholdsordsforbindelser* (1. udg., 4. opl.). Holstebro: Ventus.
- Unnerup-Madsen 1991 = Dorthe Unnerup-Madsen: *Dansk-engelsk handelsordbog. "For many reasons". Med hovedvægt på forholdsordsforbindelser* (4. udg., 1. opl.). Holstebro: Ventus.
- Vesterli 1988 = Svend Vesterli: *Grafisk fagordbog, engelsk-dansk*. Hellerup: Specialbladsforlaget.
- Warrern 1983 = Allan Warrern: *Engelsk-dansk teknisk ordbog* (9. udg.) København: Clausen Bøger.
- Warrern 1984 = Allan Warrern: *Tysk-dansk teknisk ordbog*. (6. udg.) København: Clausen.
- Warrern 1986 = Allan Warrern: *Dansk-tysk teknisk ordbog* (5. udg., 2. opl.) København: Clausen Bøger.
- Warrern 1988 = Allan Warrern: *Dansk-engelsk teknisk ordbog* (6. udg., 3. opl.) København: Clausen Bøger.
- Ærsøe 1926 = Holger Ærsøe: *Radio-Ordbog. Dansk-Esperanto og Esperanto-Dansk* København: Stud.mag. H. Ærsøe, Godthaabsvej 392, Vanløse.
- ØkOr86 = *Økonomisk ordliste. Engelsk-dansk* (1. udg., 1. opl.) Red.: Annemette Lyng Svensson. København: Samfundslitteratur 1986.

Anden litteratur

- DaOp70–71 = *Danske opslagsværker (1–2)*. Axel Andersen (red.), København: Gad 1970–1971.
- Mikkelsen, Hans Kristian 1994: *Danske fagordbøger. I: Manual i fagleksikografi. Udarbejdelse af fagordbøger. Problemer og løsningsforslag*. Henning Bergenholtz/Sven Tarp (red.). Herning: Systime, 258–305

- Mikkelsen, Hans Kristian 1994a: Lemmasektion til Nordisk Leksikografisk Ordbog. I: *Nordiske studier i leksikografi II. Rapport fra Konference om Leksikografi i Norden 11.–14. maj 1993*. Anna Garde/Pia Jarvad (red.). København: Nordisk forening for leksikografi, 193–203
- Ripfel, Martha 1990: Probleme der Erhebung metalexikographischer Daten. I: *WöDi90*, 1631–1638
- Ripfel, Martha/Herbert Ernst Wiegand 1988: Wörterbuchbenutzungsforschung. Ein kritischer Bericht. I: *Studien zur neuhochdeutschen Lexikographie VI. 2. Teilband (Mit einem Namen- und Sachregister zu den Bänden I–VI sowie einer Bibliographie zur Wörterbuchforschung)*, H.E.Wiegand (red.). Hildesheim/Zürich/New York: Georg Olms Verlag, 491–520.
- Wiegand, Herbert Ernst 1977: Nachdenken über Wörterbücher: Aktuelle Probleme. I: *Nachdenken über Wörterbücher*. Mannheim/Wien/Zürich: Bibliographisches Institut, 51–102.
- Wiegand, Herbert Ernst 1987: Zur handlungstheoretischen Grundlegung der Wörterbuchbenutzungsforschung. I: *Lexicographica* 3, 178–227.
- WöDi89/90/91 = *Wörterbücher – Dictionaries – Dictionnaires. Ein internationales Handbuch der Lexikographie* F.J.Hausmann/O. Reichmann/H.E.Wiegand/L.Zgusta (red.). Berlin/New York: de Gruyter 1989–1991.

Henning Bergenholtz

Norsk Riksmålsordbok. Utgitt av det norske akademi for sprog og litteratur. Utarbeidet av Trygve Knudsen, Alf Sommerfelt og Harald Noreng. Bind V. Første tilleggsbind A – *hogst plass*. Bind VI. Annet tilleggsbind *hohenstauffer – a–y*. Oslo: Kunnskapsforlaget 1995. (2534 spalter)

1. Supplement

De første hefter til NORSK RIKSMÅLSORDBOK blev trykt i 1930erne, bind 4 udkom i 1957. Længe før denne ordbog forelå på tryk, blev der foretaget optegnelser til et tillægsbind, som skulle udfylde "mangler i ordtilfanget i de trykte bind" (NORSK RIKSMÅLSORDBOK, Innledning: VII). Allerede i 1957 fandtes der således 25.000 sedler, som skulle indgå i et tillægsbind. Harald Noreng, som havde været redaktionssekretær (fra 1938–1940) og medredaktør (fra 1940) er den, der har fortsat arbejdet med tillægget efter Alf Sommerfelts død (1965) og Trygve Knudsens død (1968). Fra forlaget og riksmålsvernet blev der i 60erne signaleret, at man hellere så en revideret og udvidet udgave end et tillægsbind. Det var dog, skriver Harald Noreng i indledningen, umuligt at skaffe den nødvendige støtte til en nyudgave. Derfor har man måttet "nøjes" med et tillæg. Udtrykket *tillegg* hhv. det tilsvarende danske *supplement* er dog lidt beskedent i denne sammenhæng, når man ser, at de oprindelige fire bind har fået to tillægsbind på i alt 2534 spalter. Hertil kommer i øvrigt forord og indledning på side I–XVI og 20 ikke nummerede sider med litteraturlister og forkortelser. Supplementet udgør således ikke mindre end 1/3 af det samlede værk. Det er en stor andel i sammenligning med andre tilsvarende supplement, sml. Mogensen (1994) og Duekilde (1995).

Tillægsbindene til NORSK RIKSMÅLSORDBOK er således ikke de første supplementsbind i leksikografiens historie, men måske nogle af de sidste. Man må gå ud fra, at der i fremtiden næppe bliver udgivet mange supplement. Når nye ordbøger er færdige, vil de foreligge i en database eller et tekstbehandlingsprogram, hvori man herefter løbende vil kunne revidere i foreliggende artikler eller tilføje nye artikler. Når markedet eller offentlige og private fonde måtte give mulighed herfor, vil man kunne udsende en ny udgave.

Også i dette tilfælde ville brugeren have haft større fordel af en ny samlet udgave, idet det er meget møjsommeligt at springe mellem to

forskellige lemmalister. Faktisk forstår jeg ikke helt, hvorfor man ikke har scannet de gamle bind ind og så har tilføjet de nye ordbogsartikler hhv. angivelser til foreliggende artikler der, hvor de rettelig hører til.

Men det er vel en opgave, som nu kan påbegyndes. Indledningen afsluttes med et sådant ønske og anslår omfanget af en helt ny samlet udgave til at få det dobbelte omfang af nuværende udgave, dvs. $2+4 \times 2 = 12$ bind. Det vil være en stor opgave, som nok burde afsluttes i løbet af en noget kortere periode end den, der har været brugt til den første udgave.

2. Det ydre indtryk

Der forekommer af og til uenighed om arbejdsfordelingen mellem forlag og leksikografer: Forlaget og særligt forlagets layoutmedarbejdere anser hele layoutprocessen for at være en opgave, som udelukkende og uden indblanding fra leksikografer skal løses af forlaget. På sin vis svarer en sådan holdning til de fleste leksikografers forudsætninger. Det ville være for meget forlangt, hvis leksikografen, som i forvejen skal tendere til at være noget af et universalgeni, også generelt skulle være layout-ekspert. På den anden side er det også sådan, at mange ordbøger har valgt et layout, som ikke i alle henseender er hensigtsmæssig, sml. Almind/Bergenholtz (1991) og Bergenholtz (1995).

Også denne ordbog kunne på enkelte punkter med fordel have valgt andre løsninger. Man kan formode, at forlaget ville have foretrukket et andet layout, men har valgt at lade tillægsbindene fremstå med samme udseende, som de hidtidige bind: "Det var naturlig å gi tilleggsbindene en form som ligger nær firebindsverket når det gjelder format, typografi og utstyr forøvrig." (NORSK RIKSMÅLSORDBOK, Innledning: V). Kritikken er dog mindre graverende end ved en række andre storordbøger. Alt i alt fremstår ordbogen som et smukt værk, der er kendetegnet af en solid indbinding og en behagelig varm mørkeblå farve. Det ville nok have været lettere at finde det rigtige bind, hvis nummerering og angivelse af første og sidste lemma havde været sat med en større skriftstørrelse. Denne indvending må i særlig grad gælde for et værk, hvor brugeren til stadighed vil være nødt til at foretage opslag i både et af bindene i hoveddelen og i supplementet. Når man skal finde et bestemt lemma, vil man på det mest påfaldende sted på siden finde en sideangivelse, fx:

Sidetallet burde enten placeres midt over de to spalter eller for neden på siden, hvorimod lemmata burde få den plads, hvor de for brugerne mindre relevante sideangivelser nu står. I øvrigt burde der være en lidt større skydning mellem den levende kolumne i form af sideangivelser og lemmata og selve tekstblokken med ordbogsartiklerne. Den her tabte plads kunne genvindes ved at bruge lidt af den rigelige tomme flade i bunden af hver ordbogsside.

Ordbogsartiklerne er skrevet med en så lille skriftstørrelse, at brugeren enten skal være under 45 år eller i hvert fald have gode øjne eller evt. særligt stærke briller. Det må indrømmes, at selv en ringe forstørrelse af skriften vil få omfanget af ordbogen til at svulme op. Men én ændring havde ikke kostet yderligere plads, nemlig at rykke selve tekstblokken med de to spalter lidt til venstre på venstresider hhv. til højre på højresider. I den nuværende version skal man næsten øve vold mod ordbogen for at kunne læse dele af de midterste spalter.

3. Norsk Riksmålsordbok som dokumentationsordbog

For sproginteresserede nordmænd er det muligvis ikke svært at holde rede på forskelle og ligheder mellem bokmål, moderat bokmål, riksmål, nynorsk og flere andre varianter af norsk. For ikke-norske brugere af ordbogen bliver der givet en ret udførlig fremstilling i indledningen til ordbogen (side X–XII), men helt klar bliver afgrænsningen ikke. Nogle af de forfattere, hvis værker er blevet excerperet til denne ordbog, ville jeg snarere have forventet inddraget til NORSK ORDBOK, som har den forklarende undertitel "Ordbok over det norske folkemålet og det nynorske skriftmålet". Men set ud fra et brugersynspunkt er det ikke nogen indvending mod ordbogens indhold, at den går ud fra en bred varietetsopfattelse, tværtimod vil brugeren have større chance for at få svar på sine spørgsmål i hoveddelen eller supplementet til denne ordbog.

En stor national ordbog kan dog ikke siges kun at være en brugsordbog, fx som opslagsværk ved (a) tekstafhængige receptionsproblemer, (b) tekstafhængige produktionsproblemer og (c) for tekstuafhængige informationer. Derudover, måske endda som vigtigste funktion, tjener en sådan ordbog som dokumentation for et bestemt sprogs bredde og dybde. Som sådan er den en vigtig del af et lands eller en befolkningsgruppes identificering med en særlig kultur og den hertil hørende kulturhistorie. Man kan udtrykke det på følgende måde: Ordbogen skal tilfredsstille et nationalt og dermed også politisk behov

for at fæstne et kollektiv ved bevarelsen af informationer om dette kollektivs sprog. Set i denne synsvinkel ligger der muligvis en sprogpolitisk strid bag det brede tilvalg af tekster med det dertil hørende manglende fravalg af tekster. Der er således ikke bare excerperet tekster fra mange kendte skønlitterære værker, fx af Kåre Holt, Mikkjel Fønhus, Nils Johan Rud, som indeholder mange dialektale udtryk. Rigsmålsordbogen har også medtaget en god del vestnorsk sprogmateriale, herunder tekster med bymålet i Bergen, der er tekster med den sørtrøndske dialekt, med Trondhjemsdialekten, med nordtrøndsk og med nordlandsk. Denne mulige sprogstrid kan en dansker ikke forholde sig til, kun fastslå, at dokumentationsmaterialet for rigsmål er så bredt, at det ikke er let at forstå, hvilke grænse-
dragninger man i virkeligheden drager mellem de forskellige betegnelser for norsk.

Om denne konflikt vil vi uden tvivl høre nærmere i anden sammenhæng, sml. hertil også Johnsen/Nordgaard (1996). Her kan fastslås, at de to supplementsbind dokumenterer en imponerende flid, som dækker mange huller i 1957-udgaven. Harald Noreng nævner udover inddragelse af værker trykt efter 1957 yderligere indarbejdelse af den klassiske norske rigsmålsliteratur, som fx i relation til Wergelands, Ibsens, Bjørnsons, Lies, Kiellands, Hamsuns og Undsets værker havde en række lemmahuller i de første fire bind. Derudover er der i supplementet tilføjet en række eksempler med kildeangivelser til artikler, som findes i 1957-udgaven.

4. Norsk Riksmålsordbok som brugsordbog

Grundpillerne i hver artikel er en leksikografisk definition med dertil hørende markeringer og eksempler med kildehenvisninger (Thoresen 1995).

De leksikografiske definitioner gives normalt i den såkaldte leksikonstil, dvs. i form af nominalfraser eller grammatisk ikke fuldstændige sætninger. Der forklares gennemgående på et forståeligt sprog og giver til udtryk fra dagligdagen i alle samfundslag et godt førstehåndsindtryk af sammenhængen. Dette gælder også for de fleste udtryk, som kan tilregnes den alment dannedes videnssfære (eller mere specielt den akademisk uddannede humanists brede viden). Man kan mærke, at den leksikografiske tradition, som denne ordbog er opstået i, ikke prøver at "nøjes med" såkaldte semantiske oplysninger, men i høj grad bringer sådanne oplysninger, som oftest kaldes encyklopædiske, sml. følgende uddrag af ordbogsartikler, alle taget fra spalte 414:

- botthuve** en slags kvinnelue som sluttet tett om hodet (kjent i forskjellige variasjoner fra Gudbrandsdalen)
- bottlenosefelt** et felt hvor det er god bottlenosefangst (særl. fangstspr. om område nord for Færøyene)
- botulisme** forgiftning som skyldes pøsegift fremkalt af bakterien *bacillus botulinus*, som kan utvikles særlig i pølser, hermetikk og andre konserverte levnetsmidler og i rakefisk
- bouclé** stoff med knudret overflate (av tykk ull ell. bomull)
- bouffant** sterkt rynket vidde i kjoleliv ell. ermer
- bougainvillea** (oppkalt etter den franske oppdagelsesreisende L.A. de Bougainville 1729–1811) prydbusk med en mengde rødfiolette høyblad og små gulhvite blomster
- bouillabaisse** sterkt krydret suppe tillaget med biter af fisk, skalldyr og forskjellige grønnsaker, særlig alminnelig i Sør-Frankrike

Vi har her et typisk tilfælde, hvor forskjellen mellom en sprogordbog og en encyklopædi snarere består i en forskjell i utførligheten, men ikke prinsipielt i arten af opplysninger. Man kan dog se, at nogle opplysninger står parentetisk. Nogle af disse parentetiske angivelser kan forstås som encyklopædiske tillægsoplysninger. Man kan dog også se, at der ikke er en gjennomført fremgangsmåte, se hertil fx tilføjelsen "særlig alminnelig i Sør-Frankrike" i artiklen **bouillabaisse**.

I det hele taget er det store arbeide kendetegnet af en ringe grad af standardisering. Det gjør ordbogen som dokumentationsordbog mere afvekslende at læse i, men uklar som brugsordbog i nogle sammenhænge. Problemet er, at brukeren ikke kan vide, om forskjellige typer af leksikografiske definitioner til samme lemmatype kun tilfældigvis er forskjellige, eller om forskjellene dækker over en betydningsforskell. Sammenlign hertil følgende eksempler med 'indbyggere' af forskjellig art, hvor man enten kan være *innbygger i*, *fra* eller *av* og det enten før eller efter at være *person fra*:

- genfer** person fra, innbygger i Genf
århusianer person fra, innbygger av Århus
koreaner innbygger av, person fra Korea
nordkoreaner person fra, innbygger av Nord-Korea

En leksikolog ville kunne tilbringe flere år med at fundere på, hvornår man kan kaldes nordmand, dansker eller københavner. Blot fordi man bor i København, er man nu ikke københavner. Men hvis man siger om en person, at han er fra en by eller et land, så er han vel københavner, koreaner eller nordkoreaner. Rækkefølgen af de to udsagn kunne også være viktig. En leksikograf kan ikke bruke lang tid til den slags

overvejelser. Ordbogskritikeren heller ikke. Men alligevel ville det nok have været en fordel, hvis der her havde været en vis standardisering. I andre tilfælde er der kun nævnt den mindre rammende betydning 'indbygger' og netop ikke den centrale betydning 'person fra':

taiwaner innbygger av Taiwan
ålborgenser innbygger av Ålborg

De virkelig interessante angivelser er dog dem, der falder helt uden for den hidtidige ramme: Hvis man er født på en ø, i en by eller i et bestemt land, vil man i særlig grad kunne betegnes som en ålborgenser, taiwaner osv. Men i ordbogen har jeg kun fundet en henvisning hertil én gang:

tahitier (innfødt) innbygger eller beboer av Tahiti

En større grad af standardisering, som i øvrigt ville være pladsbesparende, kan i en vellykket ordbog ikke indebære en mekanisk fremgangsmåde med fuldstændig ensartede angivelser i alle artikler med samme lemmatype. Her har ordbogen gode eksempler på, at leksikograferne har været opmærksomme på særlige sammenhænge:

khmer(er) (person tilhørende en) folkegruppe i Sørøst-Asia, særl. Kambodsja

makedoner 1) person fra, innbygger av det gamle Makedonia 2) slavisk-talende person fra den tidligere jugoslaviske delrepublikk Makedonia, selvstendig fra 1991, eller fra tilstøtende områder av Hellas og Bulgaria med samme slaviske språk

En forudsætning for en vis standardisering er klare leksikografiske instruktioner, som er særligt påkrævede, når der deltager mange leksikografer i samme projekt. I praksis kan man se, at ved et projekt, der løber over så mange år – som det er tilfældet med denne ordbog – ville sådanne instruktioner have været fordelagtige. Man kunne fx have sparet nogle angivelser, som snarere fører til en vis eftertanke end til entydig indsigt. Det gælder i en række tilfælde, hvor der er forsøgt synonymiske parafraseringer således, at den leksikografiske definition består af to dele. En sådan fremgangsmåde har måske den fordel, at brugeren får to muligheder for at forstå angivelsen. Det kan også betyde, at der opstår en unødvendig uklarhed. Hertil vil jeg give nogle eksempler, alle fra spalte 675 i nicheartiklen med indgangslemmaet **elskovsalf**. Dette lemma er ikke forsynet med en leksikografisk definition, det er derimod en lang række af de følgende, hvor man

bliver mindet om Mykles berømte sætning: Behøver man at elske for at elske?

elskovedøve(t) bedøvet av elskov
elskoveduset beruset, henført av elskov
elskovedblikk blikk fylt av elskov, av kjærlyghet
elskovedbånd elskov, kjærlyghet som binder (person til person)
elskovednatt natt fylt av elskov ell. kjærlyghet

Hvis "ell." i sidste eksempler står for det ekskluderende *eller*, kan man konkludere, at Mykles retoriske spørsmål bl.a. ved **elskovednat** kan vendes om til: Kan man elske, når man elsker?

Selvfølgelig er en sådan interpretation af hurtigt skrevne leksikografiske definitioner ikke helt fair. Problemet er det kendte: Kan og skal man forklare sådanne komposita, som er konstrueret regelmæssigt (ud fra det pågældende sprogs regler og med dertil hørende let gennemskuelig betydning)? Jeg ville i sådanne tilfælde foretrække den praksis, som også findes i denne ordbog, men endnu mere regelmæssigt i NORSK ILLUSTRERT ORDBOK: Der gives ingen leksikografisk definition, når derivata og komposita "anses for selvforklarende eller når det er naturligt å søke forklaring under det ordet som kommer umiddelbart foran eller umiddelbart efter" (NORSK ILLUSTRERT ORDBOK, Veiledning, kap. 5). I sådanne tilfælde findes der dog både der og i NORSK RIKSMÅLSORDBOK ofte et eksempel:

elskovedsalf *dog er enhver af dem (roserne) en elskovedsalfes hjem* (Werg. I, 199).
elskovedsharpepill *skyharpens høje elskovedsharpepill* (Werg., Saml. Skr. II, 2, 460).
elskovedskjælen *en rosenfingret, elskovedskjælen due* (Werg. IV, 79).
elskovedslykke *het og kjælen elskovedslykke* (Unds., Ol. Aud. Hest. II, 213).
elskovedsmunn *slig en glød paa elskovedsmunde* (Werg., Saml. Skr. II, 5, 83).

Nogle af eksemplerne er uden tvivl mere informative end andre. Selv ville jeg have foretrukket noget længere eksempler, også selv om det havde betydet færre eksempler, herunder også færre litterære belæg. Den omfattende inddragelse af litterært stof gør på den ene side ordbogen til en værdifuld kilde for litteraturinteresserede. De mange litterære henvisninger er på den anden side mindre vigtig for den ordbogsbruger, der i ordbogen først og fremmest måtte søge oplysninger om ords brug og funktion, idet kildehenvisninger tager megen plads, som kunne have været brugt til flere grammatiske og kollokative

angivelser. Kildehenvisninger er især relevante for ordbogen som dokumentationsordbog, mere end de er for en brugsordbog. De mange litterære eksempler påvirker også vægtningen i forklaringsdelen, hvor markeringen "poet." bliver uforholdsmæssigt ofte brugt, ofte sammen med den yderligere markering "litt.". Det må dog betvivles, om disse markeringer er træffende som beskrivelse af opslagsordet. Markeringerne "litterær" og "poetisk" retter sig i virkeligheden mod det litterære eksempel. En elskovslykke er ikke nødvendigvis korrekt beskrevet med tilordningen til "litt.", en elskovsmund ikke nødvendigvis med markeringen "poet.". Der må anbefales en meget mere tilbageholdende brug af pragmatiske, herunder diatekstuelle markeringer, sml. Wiegand (1981). Det skal ikke betvivles, at det norske rigsmål kan bruges uhyre litterært og poetisk, men dog, at der er så mange ord, der udelukkende bruges i sådanne sammenhænge.

Vi nærmer os her et af de leksikografiske grundproblemer: Hvilken empirisk basis og hvilke interpretationer heraf? Der er tradition for en noget uforsigtig interpretation af de nævnte diatekstuelle markeringer, men også ved andre, særlig ved diafrekventive og diatekniske.

For hyppighedsangivelser gælder, at de ikke kan have et sikkert grundlag, hvis ordbogen bygger på en belægsamling, som det er tilfældet ved NORSK RIKSMÅLSORDBOK. Ved en belægsamling kan man ikke vide, i hvilket omfang hyppige og sjældne ord og ordbrug er medtaget i relation til den virkelige frekvens i hele tekstmaterialet. I praksis består mange belægsamlinger i særlig grad i en samling af alt det sjældne, det påfaldende og særligt interessante. Derimod er alt det regelmæssige forholdsvis meget dårligere repræsenteret. Man kan derfor ikke vide, hvad den ofte brugte markering "sj." (sjælden) virkelig betyder, sml. hertil:

elfenbensbestolpet sj.

elfenape sj.

elfenbensdiadem sj.

elfensfløyte sj.

elfentempel sj.

elgjegermerke sj.

Selvfølgerlige har vi her og i de fleste andre tilfælde med denne diafrekventive markering et ord, som ikke bruges særligt hyppigt. Men i et tekstkorpus vil mængden af ord, som kun bruges en eller to gange, normalt udgøre mere end 50% af alle potentielle lemmata. Så denne markering burde, hvis der hermed menes "et eller to belæg i belægsamlingen", bruges uhyre hyppigt, meget mere, end det allerede

er tilfældet i NORSK RIKSMÅLSORDBOK, sml. hertil følgende lemmata, hvor en tilsvarende markering "mangler":

elgdrev
elgfløyte
elggrå
elghornsskjeft
elghud

Ordbogsartikler som de netop nævnte, både dem med og dem uden diafrekventive angivelser, bliver forklaret fyldestgørende og på en sådan måde, at brugeren kan finde de nødvendige baggrundsoplysninger, sml.

elgdrev gruppe, linje av folk som driver, jager elgen under elgjakt
elgfløyte fløyte ell. sirene med høy skarp lyd, montert foran på tog, brukt for å skremme elg bort fra banelinje
elghudspanser om norrøne forh., panser av elghud
elgjegermerke skyttermerke som skal vise at en har skyteferdighet til å delta aktivt i elgjakt

Vi har i disse tilfælde udtryk fra dagligdagen og/eller fra den vidensfære, som en (norsk) sprogleksikograf ud fra egen baggrund kan være velorienteret om. I de mere faglige sammenhænge er betydningsangivelserne ligeledes stort set overbevisende, hvis det drejer sig om humanistiske sammenhænge. Her har en filologisk uddannet leksikograf i høj grad været på hjemmebane. Alligevel vil jeg opfordre til en noget mere tilbageholdende inddragelse af regulære fagtermer, dvs. af sådanne fagtermer, der ikke bliver brugt i aviser o.l., men kun i tekster for fagfolk eller semifagfolk. Der er lang tradition for at medtage så mange fagtermer som muligt, også fra "den lærde terminologi fra et udvalg fagbøger og fagtidsskrifter" (NORSK RIKSMÅLSORDBOK, Innledning, XI). Med henvisning til Paolo Benis anmeldelse af Accademia della Crusca's ordbog fra 1612 foreligger der dog også en anden tradition, som for en almen ordbog advarer mod at medtage sådanne opslagsord, der udelukkende bruges i en "lærd terminologi". Når sådanne termer medtages, påviser Paolo Beni, vil leksikografen både gå ud over sin egen kompetence og give brugeren en oplysning, som han ikke vil søge i en almen ordbog (se Hausmann 1989).

Man ville selvfølgelig have kunnet inddrage fageksperter, så unøjagtighederne kunne være undgået. Argumentet er dog, at ordbogen

her kunne være mere selektiv og undlade artikler som de følgende med molekylærbiologiske og biokemiske termer:

genommutasjon mutasjon som omfatter hele kromosomer ell. hele kromosomsett

Forklaringen er sådan set god nok. Det drejer sig om strukturelle ændringer i kromosomerne eller om at deres antal afviger fra det normale. Men fagtermen herfor har i mindst 20 år været *kromosom-aberration*. Den her anførte term *genommutasjon* er ganske vist ny i forhold til 1957-udgaven, men allerede forældet i 1995-udgaven. Hvis fagtermen medtages, burde en sådan oplysning også tilføjes.

elementaranalyse elementær analyse

Dette kan heller ikke på rigsmål være korrekt. Der må skelnes mellem en elementar og en elementær analyse. Og det er på ingen måde det samme. Om det så er en almen ordbogs opgave at oplyse om forskellen, er en anden sag.

bakteriofag submikroskopisk organisme, opfattet som virus, som snylter på og til dels opløser og dreper bakterier

En bakteriofag er en virus, den ikke bare opfattes som en virus (hvordan man så også ville kunne gøre det). Bakteriofager har bakterier som værtsceller og angriber disse. Om disse bakterier bliver dræbt eller ej, er afhængig af en lang række faktorer. På lignende vis er det også misvisende at sige, at de opløser bakterier, noget sådant er en kemisk proces, som evt. sættes i gang, men man kan ikke uden videre sige, at det gøres af bakteriofagerne. I korthed kan man sige, at en kort forklaring, fx "virus, der angriber bakterier", havde været mere træffende og måske også lettere forståelig end den noget længere i ordbogen. Det gælder også for det næste eksempel, som kort og godt kan forklares som "ædeceller, som er vigtige for immunforsvaret".

makrofag stor(e) fagocyt(er) ell. etecelle(r) som har vigtige renovasjons- og forsvarsfunktioner i organismen

For de hidtil nævnte molekylærbiologiske termer gælder, at de ikke nødvendigvis skal optages i en almensproglig ordbog. Derimod er andre termer fra samme fagområde centrale udtryk i den samfundsmæssige debat og bør derfor ikke bare medtages, men også forklares korrekt. Men det må medgives, at der stilles store krav til økonomien

for et leksikografisk projekt, idet det ideelt set vil dreje sig om inddragelse af flere hundrede eksterne eksperter, som fx kunne have hjulpet ved følgende termer:

gen celleelement som arveanleggene er knyttet til

Arveanlæg er ikke knyttet til generne eller omvendt. Gener er arveanlæg. At tale om celleelementer kan muligvis forsvares overfor lægfolk. Mere nøjagtigt ville være "DNA-nukleotidsekvens, som koder for et protein eller forskellige RNA-molekyler". Denne forklaring er måske for teknisk, i stedet for kunne man nøjes med 'arveanlæg' som eneste forklaring. Det er både korrekt og let forståeligt.

genteknologi område innenfor moderne genetikk og cellebiologi der en gjennom biologiske teknikker kan påvirke arvestoffet i en organisme, slik at denne utfører bestemte opgaver

Hvis genteknologi skal tilordnes nogen videnskabsgren, må det snarere blive det ikke nævnte område molekylærbiologi. Selvfølgelig spiller genteknologi også en stor rolle inden for genetik, cellebiologi, medicin, veterinærmedicin osv. osv. I virkeligheden drejer det sig ikke om et område inden for en bestemt videnskabsgren, men inden for en lang række. Det kaldes også DNA-teknologi og kan betegnes som anvendelse af laboratoriemetoder, hvorved der dannes rekombinant DNA, dvs. nye kombinationer af DNA-molekyler. Ofte er der tale om en sammenhægtning af DNA fra forskellige kilder. De teknikker, der bruges, kan man indtil en vis grad betegne som biokemiske teknikker, hvis man vil, men næppe biologiske teknikker. Dette ville man snarere kunne gøre, hvis der var tale om bioteknologi. Problemet er således ikke kun at skaffe den nødvendige viden, men også at fremstille den således, at den er forståelig uden derved at miste for meget af den korrekte faglighed.

5. Nationalordbøger: før, nu og i fremtiden

For NORSK RIKSMÅLSORDBOKS pendant, dvs. NORSK ORDBOK, kan der angives følgende tidsramme. Den påbegyndtes i 1930. Første bind kom i 1966, bind 2 i 1978, bind 3 i 1995, bind 4 efter denne hastighed ca. i 2010. "Så det kan virke som om verket i likhet med menneskeheten skrider langsomt frem" (Gundersen 1995:8). Dette udsagn bliver ikke mindre træffende af Gundersens videre oplysninger om, at man ved

udsendelse af bind 1 regnede med 7–8 bind i alt, med udsendelse af bind 2 var tallet udvidet til 10 bind, og i forordet til bind 3 regnes der med 12 bind i alt. Ved udgivelsen af bind 4 må man regne med et omfang på 14, ved bind 5 på 16 osv. Når det omtalte bind 12 udgives, er vi således foreløbig nået til et omfang på 30. Vi har her en ny version af Achilles og skildpadden.

NORSK RIKSMÅLSORDBOK står i denne sammenligning på sin vis stærkere. Den blev påbegyndt ved samme tid som den nynorske ordbog, men foreligger (allerede) afsluttet. Det er ganske vist kun i en udgave med 4+2 bind. Men den er færdig. På mange måder er det et imponerende værk, specielt når man tænker på, at de to sidste bind næsten udelukkende er udarbejdet af én leksikograf. Alligevel må man sige, at denne anmeldelse ikke er nær så positiv som fx Thoresen (1995:18), som uden indskrænkninger giver "stor ros for tilleggsbindene til NRO, og det samme gjelder ham som i all hovedsag har utarbejdet verket: professor Harald Noreng". Forskellen i de to bedømmelser ligger ikke i, at der i denne anmeldelse er påvist en del unøjagtigheder også på grund af den manglende standardisering af de leksikografiske definitioner, men at der her er blevet lagt større vægt på ordbogens brugsfunktion. Det betyder også, at ordbogens dokumentationsfunktion ikke er blevet sat i centrum. Samtidig er der gjort opmærksom på den empiriske unøjagtighed, der nødvendigvis opstår, når materialet er en belægsamling, som ikke uden videre kan siges at udgøre en eksemplarisk stikprøve af de inddragede tekster. Desuden er det ikke uproblematisk at generalisere litteraturbelæg som gældende for hele sproget, som det i høj grad er tilfældet både med denne og med andre tilsvarende ordbøger, fx den danske ODS eller NORSK ORDBOK. Endelig foreligger der som i mange andre totalsproglige ordbøger en større inddragelse af faglige universer, end det er relevant for den normale bruger af en sprogordbog, og som ikke uden for store problemer kan beskrives af en sprogleksikograf.

Hvad så? Der findes ikke nogen let løsning, men måske kan følgende tanker tages op: I sidste århundrede begyndte traditionen med at udarbejde nationale ordbøger med tidsrammer, som gik over mange generationer. Denne tradition lever videre i en række projekter, som efter min mening først og fremmest må forstås ud fra deres historiske forudsætninger: Dokumentation af den nationalsproglige fortid og nutid. I CD-ROM- og Internet-tidsalderen er man, om man kan lide det eller ej, tvunget til at tænke og handle ud fra en meget kortere tidshorisont. Jeg mener, at det er muligt uden helt at opgive nationale projekter af den art, som dokumenteres gennem udarbejdelse af NORSK

RIKSMÅLSORDBOK og NORSK ORDBOK. Men der bør foretages nogle kursændringer.

For det første bør man satse på de nye medier og tilbyde resultater, som indeholder både ordbogen og dennes empiriske basis, dvs. samtlige belæg eller analyserede tekster, for så vidt de foreligger i computerlæsbar form. Under forudsætning af en tilgang til denne basis vil man kunne foretage en slankekur af det leksikografiske koncept, så papirversionen i højere grad bliver fri for de forstyrrende belæghenvisninger og kan koncentrere sig om det virkelige budskab, oplysninger om riksnorsk (eller norsk eller moderat bokmål, hvis man vil kalde det det). Hvis man samtidig foretager en eller anden form for samarbejde med et af de to store norske encyklopædiske projekter, vil man i højere grad kunne undlade at gå ind på en leksikografisk definition af en række fagtermer, som overstiger normale sprogleksikograferes vidensniveau. I stedet kunne oplysningerne i en encyklopædi og evt. også i et udvalg af enkeltfagsleksika forudsættes, så man i en papirudgave kunne koncentrere sig om de almensproglige betydningsangivelser og udtalemæssige og grammatiske oplysninger om alle opslagsord. Men man kunne også vælge at foretage en integration i form af en encyklopædisk ordbog, både til en papirudgave og en CD-ROM-version.

7. Litteratur

- Almind, Richard/Bergenholtz, Henning 1990: Klæder skaber folk: Ordbøgers lay-out. I: *Hermes* 4, 31–47.
- Bergenholtz, Henning 1995: Ordbøgers funktion og æstetik. I: Peder Skyum-Nielsen (udg.): *Sprogets funktion og æstetik*. København: Gad, 55–81.
- Duekilde, Anne 1995: Hovedværk og supplement. I: Ásta Svavarsdóttir/Guðrún Kvaran/Jón Hilmar Jónsson (red.): *Nordiske studier i leksikografi 3. Rapport fra Konference om leksikografi i Norden. Reykjavík 7.–10. juni 1995*. Reykjavík: Nordisk forening for leksikografi 1995, 83–92.
- GENTEKNOLOGISK ORDBOG = Uwe Kaufmann/Henning Bergenholtz: *Genteknologisk ordbog. Dansk-engelsk/engelsk-dansk molekylærbiologi og DNA-teknologi*. København: Gad 1992.
- Gundersen, Dag 1995: Norsk ordbok. I: *Språknytt*, 23. årg., nr. 3, 8–9.
- Hausmann, Franz Josef: *Kleine Weltgeschichte der Metalexikographie*; in: *Wörterbücher in der Diskussion. Vorträge aus dem Heidelberger*

- Lexikographischen Kolloquium*, hrsg. von Herbert Ernst Wiegand. Tübingen: Niemeyer 1989, 75–109.
- Johnsen, Lars G./Torbjørn Nordgård 1996: Korpusbasert leksikonbygging (i dette tidsskrift)
- Mogensen, Jens Erik 1994: [Anmeldelse af Ordbog over det danske Sprog. Første bind. København 1992]. I: *Hermes* 12, 217–225.
- Noreng, Harald 1995: Om tilleggsbindene til Norsk Riksmålsordbok. I: *Ord om ord*. Årsskrift for avdeling for leksikografi. Universitetet i Oslo, 23–24.
- NORSK ILLUSTRERT ORDBOK = *Norsk illustrert Ordbok. Moderat bokmål og riksmål*. Redaktør: Tor Guttu. Oslo: Kunnskapsforlaget 1993.
- NORSK ORDBOK = *Norsk ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet*. Band III. Flusker–Gigla. Red. av Reidar Bø/Arnbjørg Hageberg/Laurits Killingbertrø/Sigurd Norlie/ Gunnar Pedersen. Oslo: Det Norske Samlaget 1994.
- ODS = ODS: *Ordbog over det Danske Sprog* 1–28, udgivet af Det danske Sprog- og Litteraturselskab. København 1919–55.
- Thoresen, Arnold 1995: Norsk Riksmålsordbok, bindene V og VI – et storverk. I: *Språknytt. Meldingsblad for Norsk Språkråd*, 23/4, 18.
- Wiegand, Herbert Ernst 1981: Pragmatische Informationen in neuhochdeutschen Wörterbüchern. Ein Beitrag zur praktischen Lexikographie. I: *Studien zur neuhochdeutschen Lexikographie I*, hrsg. von H.E.Wiegand. Hildesheim/New York: Olms, 139–271.

Ilse Cantell/Nina Martola

Nationalencyklopedins ordbok

Nationalencyklopedins ordbok 1–2 1995–96. Språkdata, Göteborgs universitet. Bokförlaget Bra Böcker: Höganäs. Pris SEK ca 2.500:-, FIM ca 1.680:-.

Nationalencyklopedins ordbok is a new, large Swedish dictionary in three volumes. The layout of the dictionary is very pleasant and clear and the information is presented in a very systematic way in the articles. The definitions of the headwords are in general adequate. However, the principle of giving a definition for every single headword results in a certain amount of empty or even confusing definitions. The valency information is also given in a systematic way, which is a considerable improvement on earlier Swedish dictionaries. More attention could have been paid to the semantics of the preposition usage.

Svenska ordboksanvändare hade sett fram emot att *Nationalencyklopedins ordbok* (i fortsättningen NEO) skulle utkomma. Det kändes därför högtidligt att hålla det första bandet i sin hand, när det utkom hösten 1995. Våren 1996 kom band två, och hösten 1996 beräknas det tredje och sista bandet utkomma.

Motsvarar då den nya ordboken de högt ställda förväntningarna? Kanske inte helt och hållet. Åtminstone kan en viss besvikelse anas bland många som bläddrat i den eller använt den för att slå upp i. Möjligen kan besvikelsen ha att göra med alltför högt uppskrivade förväntningar.

Vem är ordboken tänkt för?

NEO är en omfattande ordbok, och man frågar sig naturligtvis vem den riktar sig till. Det enda som sägs om den saken i inledningen är: "Vissa av de informationer som ges kan synas självklara för svenskar, men vi har också haft användare med annat modersmål än svenska i åtanke." Det sägs alltså inte om ordboken är tänkt för gemene man eller för folk som är professionella språkbrukare eller om den skall kunna användas av bäge kategorierna. Troligt är väl att ordboken är avsedd för en bred kategori av olika användare.

Oavsett hur det är tänkt, är det i alla fall inte så sannolikt att NEO blir en ordbok som sporadiska ordboksanvändare skaffar sig i någon

större utsträckning. Därtill är den alltför dyr och påkostad. I de finländska bokhandlarna kommer totalpriset att ligga på ca 1.680 FIM. Norstedts stora svenska ordbok, som är en nyupplaga av Svensk ordbok (i fortsättningen SOB), kostar 493 FIM, alltså knappa tredjedelen. Förmodligen är prisrelationen ungefär densamma i Sverige.

Informationen i artiklarna

Den nya ordboken är trevlig att bläddra i. Layouten är tilltalande och tydlig. Texten är visserligen rätt liten men sidorna är spatiöst tryckta, och principen ny information – ny rad gör artiklarna överskådliga och klara. Men samtidigt gör den här glesheten att sidantalet sväller ut. NEO är ingen användarvänlig ordbok från hanterlighetssynpunkt; tre tunga band i formatet 23 x 28 x 3,5 cm. För att undvika ont i handleden bör man ha fri bordsyta när man skall slå i böckerna – och det är något som många flitiga ordboksanvändare har för litet av. En tvådelad version i paperback med kompaktare layout vore en tilltalande tanke för oss vana och hängivna ordboksanvändare.

I artiklarna följer de olika informationskategorierna alltid varandra i en bestämd ordning, vilket gör att man som användare snabbt lär sig hitta i artiklarna. Artikeln **emalj** kan få tjäna som exempel:

Fig. 1. Artikeln **emalj** i NEO

Efter uppslagsordet ges alltid uppgifter om ordklassstillhörighet och böjning. Därefter följer en ordledsangivelse på egen rad och sedan kommer betydelsebeskrivningen. Ordet *emalj* har två huvudbetydelser och dessa markeras med siffror. Om en betydelse avviker från huvudbetydelsen anges det explicit med förklaringen *betydelsenyans(er)*. Språkexemplen ges under den betydelse de hör till, vilket är tydligt och

bra. Att t.ex. sammansättningarna till ett uppslagsord tas upp under respektive betydelsegrupp är mycket värdefull information för användaren, som många gånger kan ha behov av att kontrollera till vilken betydelse av ett ord en sammansättning är bildad. Det är ju en sak som inte ens modersmålstalare alltid är säkra på. Det hade kanske varit mera utrymmesekonomiskt att som språkprov ta upp sammansättningar som inte återfinns som uppslagsord, eftersom ett sammansättningsurval ändå alltid bara blir ett axplock bland alla de sammansättningar som är tänkbara. Möjligheterna att bilda sammansättningar är ju väldigt stora på svenska.

NEO är en s.k. definitionsordbok. Varje ord förklaras alltså med hjälp av en omskrivning, som skall gå att sätta in i en kontext i stället för uppslagsordet. Vissa avsteg görs från denna princip vad gäller adjektiven, men detta påpekas i användardirektiven. Över lag är definitionerna i NEO klargörande och bra. Mest problematiskt blir kravet på definition när uppslagsordet avser all dagliga och välbekanta företeelser. Ordet *byggnad* definieras så här i sin första betydelse: 'fristående, täckt, över marken upphöjd, fast konstruktion'. Frågan är om inte uppslagsordet är förståeligare än definitionen som skall förklara det i ett sådant fall. De finstilta tilläggförklaringarna är mera upplysande, vilket beror på att synonymen *hus* och det "lätta" ordet *rum* uppträder: 'vanl. med ett antal rum e.d.; ofta om större hus avsett för ngt särskilt ändamål, ofta för annat än boende'. I en definitionsordbok får synonymer inte användas för att förklara ett ord. Det är i och för sig en bra princip, eftersom synonymer lätt leder till cirkelgång och till suddigheter, men å andra sidan frågar man sig hur många som blir hjälpta av definitionen på *byggnad* ovan. Och om definitionen inte klargör någonting undrar man vilken mission den har.

Inte heller definitionen på *byxa* är särskilt upplysande: 'plagg som täcker underkroppen och större eller mindre del av benen och har en öppning för vardera benet'. Definitionen av räkneorden ställer man sig också undrande inför; *arton* 'grundtalet som är sju plus ett', *femton* 'grundtalet som är fjorton plus ett'.

Ibland är definitionerna bara tämligen intetsägande; *glödsteka* 'steka på glöd', *gropig* 'som har många gropar', *grönländsk* 'som har att göra med Grönland', *gröpning* 'det att gröpa'.

Den här typen av nedslag går det naturligtvis att göra i vilken definitionsordbok som helst för att lyfta fram definitioner som är mindre lyckade, och det förtjänar att upprepas, att det allmänna intrycket är att NEO över lag har lyckats bra. Att ge vettiga definitioner på uppslagsord som avser välbekanta fenomen måste vara ett av enspråkslexikografens största dilemman, och man kan diskutera om kravet på att det skall finnas en definition för vartenda uppslagsord inte ibland

leder till orimligheter. En dansk systerordbok till NEO, eller kanske snarare till SOB, *Nudansk ordbok*, avstår från att definiera genomskinliga och välbekanta ord, och sparar därmed en hel del utrymme. Å andra sidan är naturligtvis inte alla "välbekanta" ord så lättförståeliga för icke-modersmålstalare. Men för att tillgodose deras behov borde kravet på enkelhet och lättförståelighet skärpas betydligt. NEO:s definitioner är knappast utformade med tanke på användare som inte har svenska som modersmål, därtill är de flesta för formaliserade och teoretiska. Diskussionen om definitionsspråket går säkert vidare.

I ett fall uppträder faktiskt en synonym i stället för definition i artiklarna. Om två synonyma uppslagsord står nära varandra i alfabetet definieras det ena och det andra förklaras med det som har definierats. Principen verkar vara att det som kommer först i alfabetet får definition. Exempel på det här är *glacialräffla* – *glaciärräffla*, *glädja sig* – *glädjas*, *gråsej* – *gråsida* och *oöverlagd* – *oövertänkt*.

Ett utmärkt grepp, som är väl genomfört i NEO, är att ge de allra centralaste synonymerna och motsatserna och att ange närbesläktade ord. Se t.ex. artikeln **emigration**:

Fig. 2. Artikeln **emigration** i NEO

Det här bidrar till att ge en allsidig belysning av språket, och för en professionell språkbrukare som sitter och brottas med formuleringar kan det vara ovärderligt, eftersom man ibland kommer ihåg just t.ex. motsatsen till det ord man söker. Vad gäller artikeln *emigration* undrar man bara om inte *migration* vore att betrakta som ett överordnat begrepp till uppslagsordet, så att en uppåtriktad pil borde ha använts i stället.

Trots att en databas ligger till grund för den tryckta versionen av NEO kan man hitta vissa inkonsekvenser mellan ordartiklarna (i vilken ordbok kan man i och för sig inte det).

För ordet *palt* anges t.ex. att det i sammansättningar kan ha betydelsen 'kräk, stackare' och som exempel ges sammansättningen *harpalt*. För uppslagsordet *harpalt* anges däremot betydelsen '(liten)

hare', vilket är en adekvatare förklaring. Någon vanlig efterled är *palt* definitivt inte i betydelsen 'stackare'. I Språkdatas korpus över Bonniersromaner från 1976–77 återfinns sammansättningen *oxapalt* (finns också i SAOB-korpusen), men den bildningen har en ålderdomlig klang. Moderna sammansättningar till ordet *oxe* saknar *-a-* i fogen. Några andra belägg verkade inte finnas i korpusarna.

Frasen *med benäget bistånd* kommer upp under både *benägen* och *bistånd*. Under *benägen* finns den som exempel på den andra huvudbetydelsen 'som visar välvilligt intresse mest om handling e.d.' Som synonym anges *godhetsfull*. Under *bistånd* är exemplet *för ledningsmålet stod NN med benäget bistånd av ON*, där betydelsen 'godhetsfull' knappast är aktuell.

Ordledsangivelser

NEO markerar konsekvent ordledsgränserna för alla uppslagsord. Huvudledsgräns anges med ett längre streck och övriga ledgränser med ett kortare; *don-at-or-er*, *frukt-ans-värd*. Dessa ledgränserna fungerar samtidigt som gränser vid s.k. morfologisk avstavning. I ordboken används också genomgående denna avstavning, och i förordet sägs det: "Det är inte utan att avstavningar som *gök-ur-et* och *gång-berg-art-en* är mer tilltalande än den strikt mekaniska principen *gö-ku-ret* och *gång-ber-gar-ten*". Men är det inte snarare automatiska avstavningsprogram som ger upphov till avstavningar av det senare slaget och inte reglerna för mekanisk avstavning, där ordgränser veterligt går före enkonsonantsregeln? Det här är ett tämligen kraftigt ställningstagande i en kontroversiell fråga.

Enligt uppgift kommer morfologisk avstavning att införas i nästa upplaga av Svenska Akademiens ordlista, vilket innebär ett ännu kraftigare ställningstagande, eftersom den är normerande. Det är kanske onödigt att gå in i avstavningsdebatten här, men en stor nackdel finns det med den morfologiska avstavningen, och det är att den gör skribenterna ytterst beroende av ordböcker. Reglerna för mekanisk avstavning är tämligen enkla att tillägna sig. Morfologisk avstavning kräver djupare insikter i svenska språket än vad det stora flertalet skribenter besitter eller också källor att slå i.

Första belägg och etymologi

För varje uppslagsord anges när det först är belagt. Det här är en uppgift som i och för sig är intressant och som många användare säkert

gärna tar del av, även om den kanske i allmänhet har mindre praktisk betydelse för professionella användare. I förordet sägs det uttryckligen att det är fråga om första belägg, och det kunde man ha försökt framhålla bättre också i artiklarna, med tanke på att det brukar antas att en ganska stor del av ordboksanvändarna aldrig läser förordet. *Hist. sedan* låter litet väl kategoriskt. Ett ord kan ju ha förekommit tidigare i texter som inte bevarats till eftervärlden eller det kan ha förekommit i texter som inte har excerperats. Angivelsen om första belägg åtföljs i många fall av en etymologisk förklaring, som i *emalj* ovan. Det är en trevlig nyhet eftersom etymologier brukar intressera folk.

Konstruktionsuppgifter

En synnerligen värdefull information som satts i system på ett helt annat sätt i NEO än i föregångaren SOB är konstruktionsuppgifterna. Det här en stor förbättring, särskilt med tanke på användare som inte har svenska som modersmål, men också modersmålstalare tvekar ju ibland och har behov av att kontrollera prepositioner.

Det finns ett stort antal engelska inlärningsordböcker, d.v.s. ordböcker som är avsedda för personer med ett annat modersmål än engelska. Men det finns bara en svensk inlärningsordbok i egentlig mening, nämligen Lexin-seriens *Svenska ord*, som också fått fungera som bas för en rad tvåspråkiga ordböcker med svenska som utgångsspråk. Tyvärr är dessa ordböcker inte speciellt omfattande (23.500 uppslagsord i *Svenska ord*). En stor svensk inlärningsordbok existerar inte. Därför är personer som inte har svenska som modersmål hänvisade till enspråkiga svenska ordböcker som inte är gjorda med tanke på deras speciella behov.

Det är naturligtvis orimligt att kräva att NEO eller någon annan enspråkig ordbok skulle vara gjord med ett annat språk i tankarna. Men det är inte orimligt att kräva att ordboken tar fram det som är speciellt för just svenskan.

Under rubriken *Konstruktioner* ges formaliserade uppgifter om hur uppslagsordet konstureras. (Att ha egna explicita rubriker för dessa och för idiomerna är mycket användarvänligt!) Trots att uppgifter om prepositionerna kan ha förekommit i språkexemplen tidigare i en artikel är den formaliserade informationen viktig, eftersom den är allmängiltigare. En användare kan inte alltid veta hur mycket det går att generalisera prepositionsanvändningen utifrån ett enskilt exempel. Syntaktiska exempel finns redan i SOB, och ofta har samma exempel övertagits i NEO. Idiomerna har i SOB grupperats för sig på samma sätt som i NEO, men däremot är de formaliserade, allmängiltiga kon-

struktionsuppgifterna något som är nytt för NEO. I NEO har redaktionen arbetat vidare med sitt exempelmaterial och fört samman den information som finns i de levande exemplen till ett "dött exempel". Konstruktionsuppgifter ges för verb, substantiv och adjektiv.

Obligatoriska element står utan parentes, fakultativa inom parentes, vilket är ger synnerligen värdefulla upplysningar för den som inte har svenska som modersmål:

betjäna KONSTR.: ~ *ngn* el. *ngt*

betjäning KONSTR.: ~ (*av ngn*)

På dörren till finländska affärer ser man ofta anslaget *vi betjänar kl. 9.00–20.00* e.d. under inflytande av det motsvarande finska verbet *palvella* som kan konstrueras objektslöst.

Ibland saknas tyvärr konstruktionsuppgifterna i NEO i fall då de hade varit på sin plats. Heter det t.ex. *angivelse om* eller *för något*, eller kan ordet inte konstrueras med preposition? Och kan *bemannning* användas med andra prepositioner än *vid*? Den prepositionen framgår implicit av det syntaktiska exemplet *bemanningen vid batteriet* och inte i en konstruktionsuppgift. Men heter det inte *bemanningen på en ubåt*? Eller kan man bara säga *manskapet på en ubåt*? Men för uppslagsordet *manskap* ges heller ingen konstruktionsuppgift. I betydelsen 'det att förse med manskap' (som enligt uppgift inte är så vanlig) bör väl prepositionen *av* kunna användas? Det anges inte heller.

Det man också noterar är att reflexivt pronomen ibland uppträder i konstruktionsuppgifterna; för *kolka* ges ~ (*i sig*) *ngt* och under *bekänna*, betydelse två, förekommer ~ *sig till ngt*. Vanligen tas dock de reflexiva formerna upp som egna uppslagsord, och tanken har tydligen varit att skilja mellan reflexiv användning av ord och lexikaliserade reflexiver, men frågan är om det inte leder till onödiga gränsfall. På vilket vis skiljer sig t.ex. *förlora sig i ngt* från *bekänna sig till ngt* konstruktions- och lexikaliseringsmässigt? *Förlora sig* uppträder som eget uppslagsord.

Konstruktionsuppgifterna och semantiken

För substantivet *lager* anges att det i betydelsen 'lokal för (längre) förvaring av varor' kan konstrueras med prepositionerna *i* eller *på*:

i el. *på* ~

Substantivet kan också få ett prepositionsattribut bildat med *av*:

ett ~ av ngt

Dessa uppgifter ges endast implicit i SOB. Men hur är det vid bildlig användning? Kan båda prepositionerna användas i samband med både den konkreta och den överförda betydelsen av ordet? Kan man säga *har du några historier i lager*? NEO ger inget svar på den frågan och lämnar därmed de användare som inte har svenska som modersmål i sticket. (Till NEO:s försvar bör nämnas att *Svensk handordbok* över huvud inte tar upp den överförda betydelsen av *lager*.)

Konstruktionsuppgifterna i NEO ges ibland mycket kompakt och abstrakt i form av en förkortning för den typ av led som kan ingå i konstruktionen. Följande exempel kommer ur artikeln *göra* (moment 2):

~ ngt, ~ ngn el. ngt PRED, ~ PRED att + SATS, ~ ADVL (i ngt)

(PRED står för predikativ och ADVL för adverbial)

Detta formaliserade sätt att ange konstruktionssättet är allmängiltigt och i de flesta fallen i NEO tydligt och klart. Men hur stort är det grammatiska kunnandet bland vanliga ordboksanvändare? Kan de alltid tillgodogöra sig ett sådant presentationssätt? Exemplet ovan ser nog närmast avskräckande ut. Balansgången mellan det implicita och det explicita tycks vara oerhört svår.

Att konjunktionen *att* får stå som representant för alla bisatsinledare verkar något halsbrytande. Varför inte KONJ här eller en annan förkortning för bisatsinledare?

I fråga om verb görs skillnad mellan transitiva och intransitiva verb. Om ett verb kan få ett direkt objekt eller ett prepositionsobjekt ges prepositionen inom parentes. Så t.ex. i artikeln för verbet *höra* i betydelsen 'inrikta hörselsinnet (mot ngt) för att tillgodogöra sig ngt innehåll' (moment 2):

~ (på) ngn el. ngt

Verbet *lyssna* däremot får alltid ett prepositionsobjekt i de fall då det får ett objekt:

~ (till el. på ngn el. ngt)

Att det finns en betydelseskillnad mellan de två alternativen för verbet *höra* framkommer inte av artikeln. Satserna *jag hörde ett föredrag* och *jag hörde på ett föredrag* är ju inte synonyma.

I NEO ges partikelverben som egna lemman, så också i fråga om partikelverb på *höra*, t.ex. *höra av sig*, *höra hemma* osv. Däremot har verbet *höra på* (synonymt med det objektlösa *lyssna*) inte upptagits som ett eget lemma, vilket det kanske borde, utan det finns med som ett exempel för olika betydelsenysanser av verbet i moment 2.

I NEO görs som nämnt skillnad mellan obligatoriska och fakultativa led och här ges dessutom alternativa prepositioner som kanske inte alls kommit fram i de syntaktiska exemplen. I moment 7 i artikeln *köra* (betydelsen 'bli underkänd i tentamen e.d.')

 ges i SOB och NEO samma syntaktiska exempel

han körde på tentan igen

I artikeln i NEO ingår dessutom en konstruktionsuppgift

~ (*i* el. *på* *ngt*)

som alltså skall läsas så här: "I den här betydelsen kan verbet stå i en sats utan eller med adverbial. Det eventuella adverbialet är en prepositionsfras med antingen *i* eller *på*." Däremot framgår det inte av konstruktionsuppgiften hur de två prepositionerna distribueras i kontext. Kan man säga *han körde i tentan igen*? Och vilken preposition skall man välja om ordet *tenta* byts ut mot *prov*?

Den explicita konstruktionsuppgiften ger visserligen användaren mycket mera information än vad som kan utläsas implicit ur de syntaktiska exemplen. Men det kunde kanske vara bra med en rad tillämpningar av konstruktionsuppgifterna då dessa en gång bjuder på alternativ. Varför inte ha konstruktionsuppgifterna först som en allmän regel och sedan en rad levande exempel eller redaktionsexempel? I det problematiska fallet med konstruktionsangivelsen för verbet *göra* ovan kunde de olika konstruktionsalternativen ges separat med ett belysande exempel efter vart och ett av dem. Detta tar naturligtvis mycket plats, kanske för mycket i en ordbok som NEO, som inte uttryckligen är en konstruktionsordbok.

Men det att konstruktionsuppgifterna ändå är ofullständiga återkommer på andra ställen. Den formaliserade konstruktionen koncentrerar sig bara på formen men anger inte restriktioner som gäller dess tillämpning. Dessa restriktioner är för det mesta semantiska och mer eller mindre diffusa, men de är restriktioner som finns i lexikonet och som därför kunde vara bra att få beskrivna i ordartiklarna, särskilt med tanke på de användare som inte har svenska som modersmål. Här ett exempel till:

Verbet *längta* kan konstrueras med ett komplement som är en prepositionsfras med antingen *efter* eller *till*. Valet av preposition styrs av en semantisk regel som NEO inte tar upp. Konstruktionsuppgiften i NEO ser ut så här:

~ (*efter* el. *till* ngn el. ngt el. att + INF)

Uppgiften är dock inte tillräcklig. Av den framgår inte hur den alternativa prepositionerna används. Kan båda prepositionerna användas t.ex. tillsammans med infinitiv? Bland de syntaktiska exemplen finns bara ett exempel med konstruktionstypen *längta efter + att + INF*. Men kan man lika väl säga *hon längtar till att få träffa sin son*? Svaret är kanske självklart för en modersmålstalare. Det ser ut att finnas en selektionsregel som säger att *till* i samband med verbet *längta* förutsätter en ort eller en person (på annan ort dit talaren vill resa). Prepositionen *efter* används mera generellt och neutralt och också t.ex. i samband med abstraktare ting, händelser o.d. Denna regel har man lyckats formulera så här i *Svensk handordbok*:

~ *intensivt, ivrigt efter ngn el. ngt*

~ *efter att få göra ngt*

~ *till ngn, till en plats.*

Det är klart att man kan plocka fram fel och brister i ett så omfattande material som NEO och den kritik mot NEO:s konstruktionsuppgifter som här kommit fram bör väl närmast ses som en önskelista för "den fullständiga svenska konstruktionsordboken". De formaliserade konstruktionsuppgifterna i NEO är ett jättekliv från SOB:s autentiska exempel (som enligt förordet i SOB bara i viss mån bearbetats av redaktionen). En okommenterad listning av autentiska exempel gör ingen skillnad mellan det allmänna och det enskilda.

Verbavledning

För verben som uppslagsord ges alltid uppgifter om vilka verbal-substantiv som kan bildas till ordet. Det här är i och för sig nyttiga upplysningar, men ibland verkar formerna ganska teoretiska, så man undrar hur väl belagda de är i de texter den lexikaliska databasen bygger på. Bildningar som *bitning*, *droppning*, *operering* och *försovande* fanns åtminstone inte i de av Språkdatas textkorpusar över löpande text som vi på Forskningscentralen för de inhemska språken kommer in i. En stor brist är också att bildningarna inte förklaras på

något vis. När man arbetar kontrastivt med svenska och något annat språk har man många gånger behov av att kontrollera ett verbalsubstantivs "aspekt", alltså om det avser en punktuell eller en iterativ eller durativ handling eller om det möjligen kan avse både och.

NEO och dess finska "syster"

Motsvarigheten till NEO på finskt språkområde är *Suomen kielen perussanakirja* ('Finsk basordbok', recenserad av Krista Varantola i LexicoNordica 1, i fortsättningen PS). Det kan vara intressant att göra en jämförelse mellan de två ordböckerna.

Det är i båda fallen fråga om ett ordboksprojekt som genomförts vid ett vetenskapligt forskningsinstitut, Språkdata vid Göteborgs universitet respektive Forskningscentralen för de inhemska språken i Helsingfors.

De två ordböckerna innehåller i stort sett samma informationskategorier och anknyter till en internationell tradition av enspråkiga ordböcker avsedda för en rätt så bred användarskara. Några skillnader finns det dock och dessa beror dels på skillnader mellan de två språken, dels på de respektive redaktionernas medvetna val.

På grund av finskans rika böjningssystem har man inte i ordartiklarna i PS kunnat ge explicita böjningsuppgifter. PS har i stället i början av ordboken ett böjningsschema där en modell ges för alla 78 böjningstyper för substantiv, adjektiv, pronomen, räkneord och verb. Indexsiffror efter uppslagsorden hänvisar till tabellerna.

PS innehåller nästan inga uttalsuppgifter. Det enda undantaget är nyinlånade (engelska) lånord som har kvar sitt ursprungliga uttal, t.ex. *high tech*. Inte ens betoningssuppgifter behövs här, i finskan uttalas t.o.m. franska lånord med betoningen på första stavelsen.

För PS har man inte arbetat fram en sådan definitionsvokabulär som tillämpas i NEO (och SOB). Definitionerna får tyvärr ibland ersättas av synonymer. Detta har dock den fördelen att PS i hög grad kan användas som synonymordbok. NEO är rätt så restriktiv i sin angivelse av synonymer. I stället kan dessa då ses som fullständiga synonymer (om sådana kan anses finnas).

Exempelaterialet i PS är rikare än i NEO och bidrar också för sin del till att ringa in uppslagsordens betydelse. Detta sker visserligen indirekt, men är ofta belysande och pedagogiskt i fall då en definition skulle vara klumpig eller absurd.

Redaktionen för PS har medvetet gått in för att undvika grammatiska termer och grammatisk information. Till och med ordklassuppgifterna saknas för ord som kan få olika ordklassstillhörighet.

Så är det inte i NEO, där ordklassstillhörigheten i vissa fall fungerar som grund för homonymsepareringen, t.ex. för de två homonymerna *av*. Exempelen är sedan grupperade under "sin" homonym. NEO:s användare får på detta sätt adekvat grammatisk information, som inte kräver mycket utrymme, och genom exemplen en förklaring på vad den innebär. Enkelt och pedagogiskt – och ganska självklart.

PS är inte lika klar. Ordet *päällä* kan vara både postposition ('på, ovanför') och adverb ('på, ovanför', ofta 'påslagen, påkopplad'). Ordklassuppgiften saknas och ordartikeln är momentindelad helt och hållet på semantiska grunder. Inte ens exempelsatserna är ordnade enligt ordklassstillhörighet hos uppslagsordet. Delvis kan detta kanske också förklaras med hjälp av skiljaktigheter mellan de två språken: finskan saknar kategorin verbpartikel, som kanske ställer till med förvirring i gränsdragningen mellan prepositioner och adverb i svenskan. I analysen av en finsk sats stöter man inte på detta "störande" element och för en finskpråkig person som kan lite grammatik, är ordets två ordklassstillhörigheter självklara. Men trots detta verkar lösningen i PS som ett överdrivet undvikande av en informationskategori, ett undvikande som leder till förvirring, åtminstone om man tänker på en icke-finsktalande användare av ordboken. Här är NEO klart bättre.

NEO i jämförelse med Svensk ordbok

Eftersom den stora skillnaden i pris mellan NEO och Norstedts stora svenska ordbok (Svensk ordbok) påpekades i början kan det vara intressant att jämföra de båda ordböckerna litet, för att se om de innehållsrika skillnaderna står i proportion till prisskillnaderna.

I förordet till NEO sägs det att den lexikaliska databasen, som de två ordböckerna bygger på, har utökats med ett avsevärt antal artiklar, och att nya informationskategorier tillkommit så att databasen är ny till hälften. SOB omfattar cirka 100.000 ord och fraser enligt texten bak på omslaget. I NEO definieras enligt förordet 137.000 språkliga enheter.

En del stickprov i uppslagsordlistan på olika avsnitt av alfabetet visade, att det tillkommit en hel del moderna ord, som knappast varit aktuella när SOB utarbetades, men också en del annat. På ett avsnitt på *da-* har vi *dark horse* och *darren* och, inte oväntat, en hel del ord på *data-*: *datagrafi*, *dataintrång*, *datalingvistik*, *datalogi*, *datanät*, *dataträff*, *datavirus*. Nyttillkomna är också *daterbar* och *datja*. På avsnittet på *gl-* finns det tre nyttillkomna ord; *glottal*, *glädjebägare* och *glädjefattig*. I gengäld har två ord fallit bort; *glossematik* och *glosögon*. Det senare ordet har markeringen *mindre brukligt* redan i SOB. Nya sammansättningar på *ljus-* är *ljusbärare*, *ljusskylt*, *ljussax* och *ljussvag*.

Artiklarna i NEO har en betydligt överskådligare uppställning än i SOB och nya informationskategorier har tillkommit; ordledsangivelserna, konstruktionsuppgifterna (de explicita) och uppgifterna om första belägg och etymologi. Uppgifterna om besläktade ord och motsatser är också nya. I SOB anges endast synonymer. Det är alltså en hel del. Däremot verkar inte definitionerna och språkexemplen ha ändrats i särskilt stor utsträckning för de uppslagsord som är gemensamma.

Slutord

För den professionella språkbrukaren, som har behov av att följa den nyaste utvecklingen inom svenska språket och som är beredd att betala för detta behov är NEO säkert ett ganska självklart tillskott i bokhyllan. Ordboken är ett gediget arbete, trots att man som i alla ordböcker kan hitta vissa fel och brister när man synar den i sömmarna. De tydligt strukturerade artiklarna är dess absolut största förtjänst. Tillförlitligheten verkar också god vad gäller den information som finns i artiklarna. Däremot letade vi under de sista månadernas arbete med manuskriptet till Finsk-svensk ordbok förhållandevis ofta efter information i NEO utan att hitta den där. Framhållas skall i rättvisans namn, att det då rörde sig om problem som varit extra besvärliga och därför kvarstod med frågetecken i manuskriptet.

En recension gör sällan en stor ordbokssatsning rättvisa. Det blir ohjälpligen så att vissa aspekter som ligger recensenten/recensenterna nära lyfts fram, medan betydelsefulla saker som ordboksredaktionen satsat mycket tid, möda och resurser på att lösa passerar obemärkta. I denna recension är det på grund av skribenternas bakgrund många kontrastiva aspekter som behandlas. En NEO-recensent med enspråkig bakgrund hade troligtvis haft helt andra synpunkter.

Litteratur

Ordböcker

- Finsk-svensk storordbok* Forskningscentralen för de inhemska språken. WSOY. Under tryckning.
- Nudansk ordbok* 1988. København: Politikens Forlag.
- PS = *Suomen kielen perussanakirja 1–3*. ('Finsk basordbok') 1990–1994. Helsingfors: Painatuskeskus Oy.
- SOB = *Svensk ordbok* 1986. Stockholm: Esselte.

Svenska ord = *Lexin. Svenska ord. Med uttal och förklaringar*. 1994 (2. uppl.). Stockholm: Norstedts.
Svensk handordbok 1966. Svenska språknämnden. Uppsala: Esselte.

Övrigt

Varantola, Krista: Suomen kielen perussanakirja – Finsk basordbok.
Recension i *LexicoNordica 1* 1994, 309–317.
Språkbanken vid Göteborgs universitet.

Karl Johan Danell

Jeannie Enwall och Ruth Lötmarker: *Fransk-svensk ordbok*. Stockholm: Natur och Kultur 1995 (XIV + 1007 sid., 55.500 uppslagsord och 34.500 fraser enligt förord).

Allmänt

En fransk-svensk ordbok är ett generellt instrument som kan användas på mycket olika sätt av olika målgrupper. Grundskoleelever, gymnasister, universitetsstuderande, lärare, översättare och andra specialiserade kategorier har varierande behov. Även en individuell användare utnyttjar den på helt olika sätt om han läser en Maigret-roman jämfört med om han ska sätta sig in i ett franskt affärskontrakt.

Vad efterfrågar då användaren? Olika användare, och även samma användare vid olika tillfällen, kan ha ganska motsägande önskemål: Å ena sidan bör ett fransk-svenskt lexikon vara innehållsrikt, i idealfallet innehålla alla ord man söker, samt ge alla polysemiska och fraseologiska valörer som annars skulle vara svåra att lista ut. Å andra sidan finns det nackdelar med ett mycket omfångsrikt lexikon – det är dyrare, mer besvärligt att ta med till skolan eller på resan, och mer tidsödande att konsultera.

I stort är det ändå inte alltför kontroversiellt hur ett generellt lexikon av ett givet omfång bör utformas – uppslagsorden och frasemen väljs rimligen efter sin frekvens, kanske i någon mening intuitivt vägd mellan normalprosa (ungefär tidningsspråk med sina olika subgenrer), vardagsspråk/talspråk och vetenskapligt och litterärt språk. Eftersom vägningen aldrig kan bli vetenskapligt "objektiv", så ter det sig ganska ofruktbart att mer exakt diskutera om det ena eller andra gränsfallsordet ska stå med eller inte; lexikonet är inte en rent vetenskaplig produkt utan ett verktyg för praktiskt bruk.

I Sverige har vi mycket länge saknat moderna tvåspråkiga franska lexika – de båda klassiska verken Hammar: *Svensk-fransk ordbok* och Vising: *Fransk-svensk ordbok*, som nytryckts ett otal gånger, härrör från 30-talet och är naturligtvis idag hopplöst föråldrade. Att de ändå användes åtminstone in på 80-talet visade bara den pinsamma bristen på nyare ersättare.

Det är desto mer glädjande att äntligen sådana börjar komma. För några år sedan kom den enligt min mening förträffliga Franska Ordboken från Esselte i ett behändigt format: både svensk-fransk och fransk-svensk del i en 500 sidors volym med totalt ca 80.000 ord och

fraser. Den fransk-svenska ordbok från Natur och Kultur, som här ska recenserar, är en åtskilligt mer omfattande volym. Vid en direkt jämförelse med *Petit Robert* (1993) visar den sig innehålla i det närmaste lika många uppslagsord, men kortare artiklar och färre fraser.

En första viktig fråga är om den ska åtföljas av en svensk-fransk ordbok, vilket vore ytterst tacknämligt (ingenting härom avslöjas i förordet). Det är uppenbart att var och en som på någon nivå frekventerar ett främmande språk (som turist, som studerande eller lärare, som specialiserad översättare) mycket snabbt kommer att behöva lexikon i båda riktningarna, och ett aldrig så välgjort verk i bara en riktning kommer att kännas som en torso.

Om även en svensk-fransk del är planerad, så skulle vi alltså, efter femtio års väntan, äga lyxen att kunna välja mellan ett mindre och ett större komplett och modernt alternativ.

Att utbudet tidigare har varit magert framgår också av den här aktuella bokens bibliografi, omfattande en sida, där den nyssnämnda Franska Ordboken är det enda citerade fransk-svenska eller svensk-franska lexikonet.

Uppläggning och disposition

Ordbokens uppläggning är i stort sett traditionell, vilket säkerligen är vad användarna önskar. Enligt förordet täcker boken

det allmänna ordförrådet i dagens franska, inklusive många nytillkomna ord inom kultur och samhällsliv. Det äldre språket, speciellt det litterära, är emellertid också väl representerat. Dessutom finns starka inslag av fackspråk från områden som ekonomi, juridik, medicin, teknik, naturvetenskap och humaniora. Bland speciella ordkategorier vill vi därutöver nämna egennamn (historiska, geografiska), titlar på litterära och musikaliska verk, ord från språkområden utanför Frankrike och anglicismer.

Ett ansevärt antal specialister har anlåtats för att granska resp. fackord, en värdefull säkerhetsåtgärd.

Till de sedvanliga avsnitten med bruksanvisning, förkortningar, uttalsbeteckningar och själva ordboken har fogats några nya inslag: franska oregelbundna verb (säkerligen något som många användare gärna ser), specialartiklar (se nedan), och vidare gastronomiskt ABC, vintermer, druvsorter, EU-termer och ordspråk. Alla dessa tycker jag försvarar sin plats genom att ge lite krydda åt ordboken.

Bokens målgrupp är uppenbarligen svenskar eller personer med mycket goda kunskaper i svenska. Man kan förmoda att franskspråkiga personer som studerar svenska (en mindre, men inte försumbar grupp)

också är intresserade. Det finns dock inga avsnitt särskilt avsedda för att underlätta användningen för dem (t.ex. bruksanvisning på franska, lista på svenska oregelbundna verb).

Bruksanvisning

Liksom förordet är bruksanvisningen mycket kort och koncis. Det blir överskådligt för den praktiske användaren. Dock får man inga svar på mer principiella frågor, t.ex. om hur urval gjorts inom den svåra och viktiga fraseologin.

Artiklarnas struktur

Artiklarna är traditionellt utformade, så som normalanvändaren säkerligen önskar, med uppslagsord, uttal, ordklassangivelse och därefter, i numrerad följd, de olika översättningarna till svenska ord och frasem. En nyttig detalj är att oregelbundna ordstammar, som *eut*, *fut*, *ir-* står som uppslagsord med hänvisning till grundformen.

Utformningen av typografi och layout är tydlig och läsarvänlig, enligt min icke professionella bedömning.

En liten parentes: Artikeln *aigle* har en något underlig struktur; i ordningsföljd ges betydelseerna 1. 'pulpet', 2. 'havsgös' 3. 'pappersformat', och först därefter, med markeringen SB *m, f* (substantiv mask och fem) 'örn, örnhona'. Skälet till denna ordning tycks vara att de ord som endast är maskulina ges först (men hur är det med havsgöshonor?). Men för praktiskt bruk är naturligtvis betydelsen 'örn' den dominerande och borde komma först.

Specialartiklar

En nyhet är att man lagt in 30 specialartiklar om "grammatiska" ord som, enligt förordets formulering, "har en rikedom på betydelser och fraseologi som brukar resultera i att vanliga ordboksartiklar blir oöverskådliga". Det är en bra och lovvärd idé, och dessa artiklar har blivit mer överskådliga genom typografin – de inleds med en liten innehållsförteckning, som sedan återtas med siffror eller tydliga rubriker. Själva innehållet skiljer sig sedan inte så mycket från en vanlig artikel, med uppräknings av olika betydelsenyanser och frasem, men en ansats till syntaxregler finns.

Jag granskade närmare specialartiklarna *aller* och *en*. Artikeln *aller* är på det hela taget välgjord. Man kan alltid ha olika åsikter om hur grupperingen av valörer ska göras. Både den syntaktisk-semantiska klassificeringen och urvalet av valörer och frasem är nödvändigtvis subjektiva och besvärliga. Så t.ex. ges *tout va bien* under betydelsen ' fungera, gå', medan den närliggande nyansen av *aller* i *aller de mal en pis* återfinns under 'bildlig betydelse'. Det sistnämnda uttrycket är f.ö. kanske inte det vanligaste för att uttrycka betydelsen ifråga: enligt mitt corpusmaterial (tidningstexter, skönlitteratur m.m.) och även min intuition är *de plus en plus mal* vanligare. Vidare blandas semantisk och syntaktisk klassificering: efter 10 avsnitt med olika semantiska valörer kommer 11 'konstruktioner med **en**', 12 'med gérondif' och 13 'konstruktioner med **y**'. Eftersom ordbokens mål i första hand är praktiskt kan man dock inte kritisera denna blandning av olika typer av kriterier, om den är till hjälp för brukaren.

Specialartikeln om *en* har däremot blivit ett olycksfall i arbetet. De två homonymerna (prepositionen och pronominaladverbet) behandlas i en enda artikel, och genom typografiska korrekturmissar (för det kan väl inte vara avsikten) hamnar pronominaladverbet som en underavdelning av prepositionen, och därtill utan rubrik i artikeln, vilket rimligen måste förvirra användaren.

Även i denna artikel kan man naturligtvis diskutera urval och klassificering: frasen *en l'an 2000* ges, men inte den långt mer frekventa typen *en 1996, en 2010* – även *en 2000* förekommer i min corpus, och mitt tips är att denna kortare variant kommer att slå igenom helt, av "ekonomiska" skäl, nu när vi hastigt närmar oss sekelskiftet.

Jag skulle ha föredragit att man inte som skett buntade ihop det mycket klart definierade fallet där prepositionen *en* anger 'material' (*ce bracelet est en or*) med mer svårklassificerade och disparata fall som *la même valise en plus grand* och t.o.m. *ils correspondent en français*.

När det gäller specialartiklarna inställer sig frågan mer generellt hur mycket syntax ett lexikon ska innehålla (liksom grammatiker kan fråga sig hur mycket ordkunskap och fraseologi en skolgrammatik bör ta upp). Gränsen mellan syntax, fraseologi och vokabulär är suddig, och placeras på olika sätt av olika lingvistiska och psykolingvistiska skolor, vilket gör att inget klart svar kan ges. Det är välkänt att grammatikens kapitel om t.ex. prepositioner och adverb brukar vara sorgebarn, eftersom det i hög grad är fråga om vokabulär och fraseologi, och grammatikern låter mer eller mindre implicit förstå att detta måste läras in genom textstudier, eller i lexikon. Lexikografen å sin sida har spegelvända bekymmer, eftersom prepositioner och adverb också i hög grad ingår i syntaxen. Svårigheten kan illustreras med en av de definitioner som ges för pronominaladverbet *en* i specialartikeln:

13 som ersättningsord för **de** + substantiv, infinitiv eller hel sats (motsvaras i svenskan av olika prepositioner och pronomen; ibland saknas dock svensk motsvarighet) med (på, över) det (den, dem)

Man kan inte säga att denna "syntaxregel" är felaktig, men den kan knappast vara begriplig för den som inte redan inhämtat kunskaper om *en* i en betydligt fylligare grammatisk framställning.

Trots svårigheterna är själva idén med specialartiklar en välkommen innovation. De är en förbättring i åskådlighet jämfört med andra lexikon – det är inte författarnas fel att frekventa grammatikord svårligen låter sig hanteras lexikografiskt.

Särskilda problem

Egennamn. Enligt förordet har man lagt vikt vid egennamn, men ingenting sägs om vilka principer som styr urvalet, som rimligen måste vara mycket restriktivt med tanke på mängden egennamn. Man kan gissa att det särskilt gäller kända eller frekventa namn som skiljer sig påtagligt mellan språken (*Aristote, Charlemagne, Gothenbourg* och även *Picsou!*). Däremot förstår jag inte meningen med att ta upp ett namn som *Guadeloupe*, som ju har samma form på svenska.

Titlar. I förordet poängteras att man lagt in titlar på litterära och musikaliska verk. Jag roade mig med att spontant skriva ner de mest berömda titlar som jag subjektivt kom fram till. Av dessa fanns endast *La nausée* och *Le père Goriot* med (under *nausée* resp *père*) medan jag inte hittade *A la recherche du temps perdu, La comédie humaine, Eugénie Grandet, Le misanthrope, Le barbier de Séville, L'étranger* eller *Clair de lune* (det sista kunde ju vara endera Beethoven eller Debussy!). Jag menar dock att författarnas restriktivitet är klok – det är svårt att se vilken användare som kan ha någon större glädje av dessa titlar i lexikonet. Passusen i förordet borde alltså strykas.

Förkortningar. Ett stort antal förkortningar, även mycket moderna, tas upp. Några exempel inom politiken: EAO, CEE, CEEA, CEI, CECA, CE, UEM. De som står i den särskilda listan på EU-termer finns även i den generella ordboken. Detta är tacknämligt, eftersom användaren mycket lätt glömmer bort appendix utanför huvuddelen. Några exempel från universitetsväsendet: CNRS, DEUG, ENA, UFR, UER, UV.

Dataord. För att se hur ordboken klarat ett område med många nya ord har jag letat efter datatermer. Resultatet är i stort positivt: jag hittade t.ex. *curseur, didacticiel, disque, disquette, imprimante, laser* (med

imprimante à -- och *platine --*), *lecteur* (bl.a. 'diskettstation'), *lecteur CD-ROM*, *logiciel*, *modem*, *méga-octet*, *carte à puce*, *puce*, *système d'exploitation* (under *exploitation*), *tableur*, *traitement de texte* (under *texte*, inte under *traitement*).

Däremot saknas *Internet*, *Email*, *CD* och *CD-ROM* (dessa heter ju detsamma på svengelska och vållar därför inga nämnvärda problem) liksom *kilo-octet*. *Octet* följs till att börja med av den underliga översättningen 'data oktett' före den normala termen 'byte'.

Disque dur hittar man inte under *disque*, bara under *dur*. Detta stämmer med vad som sägs i förordet, nämligen att "lexikaliserade ordgrupper" mestadels står under det senare ledet (*ruban adhésif* under *adhésif*), men det är en princip som förefaller mig leda till en del ointuitiva placeringar.

Fraseologi. För att ta ett stickprov på fraseologin i ordboken har jag undersökt artiklarna *jour* och *journée*, två ord som jag tidigare studerat (se Danell 1995). Det visar sig att de flesta frasem som jag hade hittat i min corpus också återfinns bland ordbokens totalt ca 40 frasem (*l'ordre du jour*, *l'autre jour*, *un beau jour*, *à un de ces jours*, o.s.v.). Ordboken ger också flera frasem som inte fanns i min relativt begränsade corpus (*mettre fin à ses jours*, *c'est du tous les jours*). Vissa av dessa kan nog misstänkas vara omoderna (*donner à qn ses huit jours*, *le flambeau du jour*), men behövs vid läsning av äldre litteratur.

I något fall kan översättningen ifrågasättas: *à jour* 'a dato' hjälper knappast en normalbildad svensk.

Några i corpusen relativt frekventa frasem saknas i boken: *tenir à jour*, *d'un jour à l'autre*, *jour et nuit*, *du jour au lendemain*, *A chaque jour suffit sa peine*, *en début (fin) de journée*. Detta kan knytas till två generella problem som ännu så länge saknar klara lösningar inom lingvistik. Hur definieras ett frasem? Och ska en ordbok av detta slag endast ta upp franska frasem som är svåra att genomskåda för en svensk? I så fall är det ju riktigt att utelämna *jour et nuit* som direkt motsvarar 'dag och natt'. Men då borde samma sak gälla *un beau jour* 'en vacker dag'. Om den principen gäller, så skulle författarna av en svensk-fransk ordbok få ha en ganska annorlunda princip: i det fallet bör man sträva efter att ta upp franska frasem som är svåra att aktivt komma på, även om de är transparenta.

Konklusion

Man kan alltid, som jag ovan gjort, diskutera principer för urval av ord och fraser, och man kan peka på enstaka detaljer som kan förbättras. Men sådana petitesser är försumbara jämförda med det faktum att det

nu finns ett länge efterlängtat modernt stort fransk-svenskt lexikon, gediget och väl genomarbetat. Man kan bara lyckönska författarna till att ha slutfört ett mödosamt arbete, och be en stilla bön om att energin och resurserna så småningom ska kunna räcka till även för en svensk-fransk del.

Litteratur

- Danell, K.J. 1995: *Le phénomène de concurrence en français moderne*, Acta universitatis umensis, Umeå Studies in Humanities 123. Umeå.
Franska Ordboken 1989. Solna: Esselte Studium.
Hammar, Th. 1936: *Svensk-fransk ordbok*. Stockholm.
Le nouveau petit Robert 1993: Dictionnaires le Robert. Paris.
Vising, J. 1936: *Fransk-svensk ordbok*. Stockholm.

Ela Dura, Sven-Göran Malmgren & Maria Toporowska-Gronostaj

Jacek Kubicki: *Słownik szwedzko-polski/Svensk-polsk ordbok*. Sthlm: Natur och Kultur 1988.

Jacek Kubicki: *Słownik polsko-szwedzki/Polsk-svensk ordbok*. Sthlm: Natur och Kultur 1995.

Med *Polsk-svensk ordbok* (härefter *PSO*) fullföljer Jacek Kubicki det projekt som han inledde på 1980-talet och som då resulterade i *Svensk-polsk ordbok* (härefter *SPO*). Det är ytterst glädjande att vi därmed äntligen har fått ett komplett svensk-polskt lexikografiskt hjälpmedel i mellanklassen (studieordboksklassen, för att använda en finlandssvensk term), inte minst med tanke på det sedan 1960-talet snabbt ökande antalet potentiella användare av ett sådant hjälpmedel; i dagsläget med en grov uppskattning flera tiotusental polacker, i synnerhet genom blandäktenskap och invandring till Sverige av andra orsaker, och åtskilliga tusen svenskar, till stor del genom blandäktenskap. I själva verket kan det vara så att polska tillhör de 6–8 främmande språk som flest personer med svenska som modersmål har praktisk anledning att lära sig. Något liknande kan gälla för polacker i förhållande till svenska.

Det är sålunda åtta år sedan *SPO* kom ut och det kan tyckas som om tyngdpunkten i recensionen borde ligga på den nyligen utkomna *PSO*. Men de båda ordböckerna bildar på flera sätt en oupplösbar enhet, inte minst med tanke på dikotomin aktiv/passiv (ordbok), och det kan därför inte undvikas att vi också kommer in en hel del på den äldre ordboken. Vi gör först en mer allmän genomgång, utgående från standardkriterier för värdering av tvåspråkiga ordböcker, och därefter genomför vi en mer intensiv kontrolläsning av 12 sidor i vardera ordboken. Dessutom prövar vi till sist den polsk-svenska ordboken praktiskt genom att läsa ett stycke polsk tidningstext med ordboken i handen.¹

1. Storlek, lemmanurval m.m.

¹ Om författarna: ED och MTG har polska som modersmål men är verksamma som lingvister i Sverige sedan länge; SGM har svenska som modersmål och behärskar polska tämligen väl. De avsnitt i det följande, där tonvikten ligger på ettdera av språken, ansvarar i första hand de(n) författare för som har detta språk som modersmål. Vi vill också tacka docent Olof Paulsson för värdefulla synpunkter.

Enligt baksidestexterna innehåller SPO ungefär 46.000 lemmor och PSO ca 37.000; en snabb kontroll tyder på att detta stämmer bra. Den svensk-polska ordboken är dock troligen inte (nämnvärt) större än den polsk-svenska, utan skillnaden i antalet uppslagsord beror huvudsakligen på olikheter i de båda språkens struktur. Svenska sammansättningar – som lemmatiseras – motsvaras i regel av polska fraser, som inte lemmatiseras. Exempelvis motsvaras det svenska lemmat **björntjänst** av den polska frasen *niedêwiedzia przys[]uga* (bokstavligt 'en björns tjänst') som tas upp under lemmat **przys[]uga** i den polsk-svenska ordboken (dock med fetstil i motsats till fraser i den svenska ordboken som ges i kursiv). För nordiska läsare ger därför siffran för den svensk-polska ordboken den bästa uppfattningen om ordböckernas storlek.

Med sina 46.000 lemmor bör SPO täcka ungefär 75% av lemmorbeståndet i Svensk ordbok, som Kubicki också nämner som en av sina källor. Ett viss gallring i mängden av möjliga uppslagsord har alltså gjorts; vi kommer senare något in på frågan hur optimalt Kubickis urval kan anses vara.

Som alltid är täckningen av fraser och syntaktiska konstruktioner minst lika viktig som antalet uppslagsord. Inte minst häri ligger de båda ordböckernas styrka, något man kan få en första antydning om genom att slå på uppslagsord som **gå** eller **ligga** i SPO eller deras motsvarigheter i PSO (här är dock den svensk-polska ordboken betydligt utförligare.) Artikeln **gå** tar t.ex. upp 1 1/2 sida i SPO. Vi återkommer också till denna punkt.

I Sverige har på senare år etablerats en tradition att ange, inte antalet uppslagsord, utan antalet "ord och fraser" i ordböcker. Traditionen är på gott och ont, och vi vill inte utan vidare stödja den, men hade förlaget valt att räkna antalet "ord och fraser" i reklamen, så hade antalet blivit inemot det dubbla jämfört med antalet ord (68.000 ord och fraser i PSO och troligen fler i SPO).

2. Typografi, makrostruktur m.m.

Båda ordböckerna har en tydlig och överskådlig typografi. PSO är något spatiösare genom att den inte har rak högermarginal. Från estetisk synpunkt är dock SPO klart överlägsen, dels tack vare den raka högermarginalen, dels genom att den allmänt tycks vara mer påkostad (papperskvalitet, omslag m.m.). Det är faktiskt något av en njutning att hålla SPO i handen och bläddra i den.

I SPO tillämpas partiell nischning: sammansättningar och avledningar som hör till ett gemensamt grundord samlas under detta grundord, men endast i den mån bokstavsordningen inte bryts. Den konstanta delen av grundordet markeras med hjälp av ett lodrätt streck. Det kan diskuteras om det inte hade varit bättre att avstå från nischningen, i synnerhet med tanke på de polska användarna. Det lär ju vara så att även svenska användare har svårt att hitta rätt i svenska ordböcker som tillämpar nischning, t.ex. SAOL. Det förekommer också en viss överanvändning av lodstrecket i SPO, t.ex. i följande fall, där "efterlederna" inte är särskilt användarvänliga:

Fig. 1. Exempel på nischning i SPO (citat, **pietism**)

För att bevara den alfabetiska ordningen vänder ordboken också någon gång på ett förvirrande sätt på ordningen mellan grundord och sammansättning:

Fig. 2. Artiklarna **bjällerklang** och **bjällra** i SPO.

Det torde inte vara särskilt lätt, allra minst för den polske användaren, att hitta artikeln **bjällra**! Här hade det naturligtvis varit bättre att avstå från nischningen. Så sker också vanligen i analoga fall i SPO, t.ex. vid **tavla** och dess sammansättningar.

I någon mån används lodstreck med samma funktion också i PSO. Även där blir det ibland användarvänligt, t.ex. i artikeln **próg** ('tröskel'; det polska ordet är enmorfemigt):

Fig. 3. Artikeln *próg* i PSO

3. Bruksanvisningar, ordboksgrammatik o.d.

Tyvärr måste man säga att anvisningarna till ordböckerna är alltför knapphändiga. Dels saknar man kortfattade grammatiska upplysningar om de båda språken – mer därom nedan –, dels borde vissa av ordböckernas konventioner, betingade av strukturella skillnader mellan de båda språken, ha utretts. Vi ska här nöja oss med två exempel på det senare.

Polskan har en rik verbböjning, som bl.a. möjliggör att ett pronomen ofta kan utelämnas. Exempelvis motsvaras ett svenskt *jag* eller *du* av en speciell verbändelse i polskan, i preteritum därtill av en ändelse som samtidigt anger genus. Ett svensk verb i preteritum i en sats med pronomenet *jag* som subjekt kan därför motsvaras av två olika polska verbformer, en maskulin och en feminin. Kubicki har valt att genomgående översätta med maskulinformen, och det är i och för sig i sin ordning. Men detta borde förstås ha nämnts i inledningen.

Polskan saknar vidare bestämda artiklar. Ett polskt substantiv i ett exempel kan därför motsvaras av ett svenskt substantiv antingen i obestämd eller bestämd form. Ibland är båda översättningarna möjliga, men i PSO ges normalt bara den ena. Det är också fullt rimligt, men principerna bakom valet borde nog ha utretts något.

4. PSO och SPO som aktiva och passiva ordböcker

Det är uppenbart att PSO har ambitionen att, åtminstone i viss utsträckning, vara både en aktiv ordbok för polacker och en passiv ordbok för svenskar, och att motsvarande gäller för SPO. Baksidestexterna och inledningarna, liksom titlarna, är på båda språken i båda ordböckerna, och i var och en av ordböckerna finns helt klart uppgifter som är avsedda bara för svenskar och uppgifter som är avsedda bara för polacker. Å andra sidan kan man knappast vänta sig att SPO och PSO, tagna var för sig utan hänsyn till systerordboken, ska vara fullständiga både aktiva och passiva ordböcker. Det finns som bekant få exempel på detta i den allmänna tvåspråkiga ordbokslitteraturen, åtminstone i Norden – ett är Valerij Berkovs rysk-norska ordbok (jfr också Berkov 1996). I vårt fall skulle detta innebära exempelvis att varje svensk ekvivalent i den polsk-svenska ordboken vore försedd med fullständiga (eller i varje fall nödvändiga) uttals- och böjningsuppgifter varje gång den

uppträder, något som det än så länge saknas en svensk tradition för och som åtminstone i mellanstora ordböcker kanske vore för utrymmeskrävande. Men om man betraktar SPO+PSO som en enhet skulle de tillsammans kunna fylla alla de fyra funktionerna: både aktivt och passivt ordboksverk för både svenskar och polacker. (Så är t.ex. många av van Dales tvåspråkiga ordböcker med nederländska som ena språk upplagda.) Det är från denna synpunkt vi nu ska granska SPO+PSO, först på ett mer övergripande sätt (detaljer kommer senare). Dock är det naturligen så, att den passiva funktionen i stort sett kan fyllas av en enda ordbok, den polsk-svenska för svenskar och den svensk-polska för polacker.

a) PSO som passiv ordbok för svenskar

För den svenske användaren som konsulterar PSO för att få veta vad en polsk ordform betyder på svenska är inte minst tre saker viktiga – vi förutsätter att ordboken ö.h.t. tar upp det polska lemmat. Han måste hitta fram till den rätta polska grundformen, och om den svenska ekvivalent som ges är flertydig måste det framgå vilken betydelse som är aktuell. Dessutom är det givetvis viktigt att ordboken översätter ett hyggligt antal polska fraser med mindre förutsägbar innebörd. Det sista återkommer vi till längre fram i recensionen.

Den polska böjningsmorfologin är mycket komplicerad, och enligt vår mening är PSO en smula nyckfull, och lite väl sparsam, med hänvisningar från besvärliga (oregelbundna) ordformer till motsvarande grundformer. Man hänvisas visserligen från *bior* 'jag tar' till *braç* 'ta' och från genitivformen *d"du* till grundformen *deszcz* 'regn', men man får ingen hänvisning från *szed* 'han gick' till grundformen *iÊç* 'gå' och ingen hänvisning från *pior* 'jag tvättar' till *praç* 'tvätta'. Inte heller ges några böjningsuppgifter i artikeln **praç** – eller i någon annan verbartikel. Tyvärr saknar både PSO och SPO också förteckningar över oregelbundna – liksom över regelbundna – böjningsmönster. Det senare är f.ö. en av ordböckernas mera svårförståeliga brister; att ge listor åtminstone över starka och oregelbundna verb är ju vanligt i svensk tvåspråkslexikografi. Däremot hänvisar PSO ordentligt enligt gängse konventioner inom slavisk lexikografi från perfektiva verb(varianter) till motsvarande imperfektiva. Men man måste konstatera, att den svensk som ska använda PSO som en passiv ordbok helst bör kunna en hel del polska.

PSO:s system för betydelsedifferentiering fungerar i stort sett väl. Med hjälp av synonymer, översättningar av polska fraser längre fram i artikeln (med svenska sammansättningar som ekvivalenter), fackområdesangivelser m.m. leds den svenske användaren i allmänhet till den

rätta av de möjliga betydelserna. Dock är det från hans synpunkt en nackdel att fackområdesangivelserna ges på polska. Ett litet utdrag ur en artikel ser ut så här:

poziom /.../ bud plan /.../.

För den svenske användaren, som om inte annat kan vara ovan vid polska förkortningar, hade det varit bättre om det hade stått "bygg(nadskonst) plan". Ett utdrag ur en annan artikel ser ut på följande sätt:

pozew /.../ praw stämning /.../

Den polska förkortningen praw skulle motsvaras av den välkända svenska förkortningen jur. Det bör emellertid sägas, att det går att hitta fram till de svenska motsvarigheterna till de polska förkortningarna i den tvåspråkiga förkortningslistan i början av ordboken.

b) SPO som passiv ordbok för polacker

Mycket av det som sades i föregående avsnitt om PSO som passiv ordbok för svenskar gäller om SPO som passiv ordbok för polacker. Men det måste ändå konstateras, att SPO är bättre som passiv ordbok än PSO. Betydelsedifferentieringarna tycks vara ungefär likvärdiga i de båda ordböckerna (fackområdesangivelserna är på svenska i SPO), men hänvisningarna från besvärliga ordformer till motsvarande grundformer tycks vara betydligt fler, och betydligt konsekventare genomförda, i SPO än i PSO. Så får användaren t.ex. hänvisningar från imperfekt- och supinumformerna till grundformerna av alla starka verb. Vidare ger SPO i princip tillräckliga böjningsuppgifter i resp. artiklar för alla svenska substantiv och verb (jfr dock nedan), medan PSO inte ger några böjningsuppgifter för de polska verben och endast genusuppgifter för substantiven.

c) SPO+PSO som aktivt ordboksverk för svenskar

En svensk som använder (i första hand) SPO och (i andra hand) PSO i samband med textproduktion på polska är naturligtvis betjänt inte bara av att få polska ekvivalenter, han behöver också få veta hur dessa – om det är fråga om verb eller substantiv – böjs. Här är uppgifterna i SPO

och PSO tyvärr inte tillräckliga; den enda böjningsmorfologiska uppgift som över huvud taget ges om polska ord är uppgift om substantivens genus. Denna uppgift ges i gengäld i båda ordböckerna, vilket i förbigående sagt förefaller vara en överloppsgärning.

Det ska nu åter understrykas, att den polska böjningsmorfologin är till ytterlighet komplicerad, och att det inte är någon lätt uppgift att ens i en polsk-svensk ordbok ge fullständiga böjningsuppgifter för alla verb och substantiv. Troligen är det fördelaktigast att använda sig av något slag av siffersystem, så som t.ex. görs i en polsk-tysk ordbok av något större omfång än Kubickis båda ordböcker (Bzd´ga, Chodera & Kubica 1973). Ett utdrag ur en verbartikel i den polsk-tyska ordboken jämte motsvarande ställe i den 50-sidiga översikten över polskans böjningsmorfologi ser ut på följande sätt; hänvisningen görs som synes med den romerska ettan (endast en del av böjningsmönstret återges här):

Fig. 4. Utdrag ur Bzd´ga, Chodera & Kubica (1973)

Förvisso vore det önskvärt att något liknande kunde genomföras också i en polsk-svensk ordbok, och så våldsamt utrymmeskrävande skulle det inte behöva vara (kanske 5–10% av PSO:s nuvarande omfång; översikten i den polsk-tyska ordboken kunde ev. förenklas och förkortas). Man måste dock ha förståelse för att Kubicki och förlaget ännu så länge avstått från att göra denna satsning.

Däremot förtjänar SPO+PSO allvarlig kritik för sättet att lämna uppgifter som har med de polska verbens aspekt att göra. Polskan, liksom de flesta andra slaviska språk, kännetecknas ju av att de flesta verb uppträder "parvis", med en imperfektiv och en perfektiv variant. Ett

svenskt verb motsvaras sålunda redan av denna anledning oftast av två polska verb, ett imperfektivt och ett perfektivt, och för den som använder en svensk-polsk ordbok för textproduktion på polska är det viktigt att få båda dessa verb som (partiella) ekvivalenter. Men i SPO ges oftast bara de imperfektiva polska motsvarigheterna till de svenska verben. Detta hade kunnat accepteras, om det kompletterande perfektiva verbet generellt hade givits i motsvarande artikel i PSO. Men så sker bara i vissa fall, och aldrig eller nästan aldrig, vad vi kan se, när perfektiveringen sker genom prefigering.

Ett enkelt exempel kan åskådliggöra, också för icke-slavister, hur allvarlig den påtalade bristen är. Det svenska verbet *installera* motsvaras i den tekniska betydelsen ('montera' o.d.) av två polska verb, det imperfektiva *instalować* och det perfektiva *zainstalować*. Det finns ingen regel som säger att perfektiveringen just i detta fall åstadkoms med prefixet *za-*, och i många kontexter är det helt nödvändigt att använda det perfektiva verbet, alltså *zainstalować*. Men det finns inget enkelt sätt för den svenske användaren att hitta fram till detta verb; de relevanta artiklarna i SPO (betydelsemoment 2) resp. PSO ser ut så här:

[...]

Fig. 5. Artikeln **installera** i SPO

Fig. 6. Artikeln **instalować** i PSO

Det enda sättet för den svenske användaren att få reda på den perfektiva motsvarigheten till *instalować* är genom trial-and-error, prövning av olika prefigeringsmöjligheter jämte kontroll mot PSO. När han till sist kommer till artikeln **zainstalować (si')**, finner han där en hänvisning som visar att han kommit rätt:

Fig. 7. Artikeln **zainstalować (si')** i PSO

Men detta är självfallet inte tillfredsställande. Aspektinformationen om *(za)instalować* och många, förmodligen flera tusen, andra polska verb borde ha givits på det sätt som är standard i slavisk lexikografi och som tillämpas exempelvis i Bzd'ga, Chodera & Kubica (1973, 1978):

Fig. 8. Artikeln **installieren** i Bzd'ga, Chodera & Kubica (1978)

Fig. 9. Artikeln **instalowaç** i Bzd'ga, Chodera & Kubica (1973)

En komplettering av aspektinformationen i den antydda riktningen, helst både i SPO och PSO men i varje fall i PSO, bör ha högsta prioritet i en eventuell andra upplaga.

Om sålunda den morfologiska informationen om polskan på flera sätt är otillräcklig för den svenske användaren, förtjänar SPO(+PSO) som aktiv ordbok å andra sidan stort beröm för flera andra saker. Betydelsedifferentieringen är, också från den svenske användarens synpunkt (jfr ovan), utmärkt väl genomförd, ofta med stöd av antydningar om synonymer eller om selektionsförhållanden i mindre stil. Se t.ex. följande artiklar:

Fig. 10. Artikeln **lyhörd** i SPO

Fig. 11. Artikeln **lysa** i SPO

Antalet mer eller mindre fasta fraser som tas upp är så stort man kan begära, och ett betydande antal viktiga prepositions-komplement (t.ex. i fraser som *arg på ngn*, *lärare i ngt*) översätts; också annan information av valenskaraktär ges i fullt tillfredsställande omfattning. I dessa avseenden är SPO faktiskt fullt jämförbar med de största svensk-engelska och svensk-tyska ordböckerna (jfr Malmgren 1995).

d) PSO+SPO som aktivt ordboksverk för polacker

Den polske användaren av PSO+SPO får fullt tillräcklig böjningsinformation om alla svenska verb och substantiv, huvudsakligen dock i den svensk-polska ordboken. De enda morfologiska uppgifter om svenska ord som ges i PSO är genusuppgifter för substantiven (med markeringen *n* vid neutrer och utan markering vid utrala substantiv). Eftersom den polske användaren ändå måste gå vidare till SPO för att få fullständiga böjningsuppgifter kan man diskutera hur meningsfullt det är att över huvud taget ge morfologiska uppgifter för svenska ekvivalenter i PSO.

Arrangemanget att huvudsakligen placera böjningsinformationen om de svenska orden i SPO fungerar emellertid bra, bortsett från en liten detalj. Informationen om substantiven och de starka verben ges i klartext, men informationen om de svaga verben ges med en siffra som visar konjugationstillhörighet (första, andra eller tredje konjugationen). Detta må väl kunna accepteras, även om man måste känna till en del regler för att bilda de rätta verbformerna, men bara under förutsättning att användaren kan slå upp böjningsmönstren för varje konjugation i inledningen. Av obegriplig anledning saknas emellertid denna information. Hur många polacker kan förväntas känna till de gängse koderna för det svenska verbböjningssystemet? Inte ens svenska användare torde göra det i någon större utsträckning.

Beträffande betydelsedifferentiering, valens och fraseologi kan vi i stort sett nöja oss med att upprepa berömmet från föregående avsnitt, med en liten reservation. Kubicki har en i och för sig vällovlig ambition att träffa samma stilläge hos ekvivalenterna som hos källspråkets ord och fraser. Men de svenska översättningarna blir inte alltid hundra-percentigt idiomatiska, och ibland känns de inte fullt brukliga. Vi återkommer med exempel i nästa avsnitt.

Som hjälpmedel för textproduktion – i allmänspråklig mening – på svenska är PSO+SPO sålunda mycket användbart, betydligt användbarare än SPO+PSO för textproduktion på polska. Men beklagligt nog fattas en sak för att PSO+SPO ska kunna betraktas som ett fullvärdigt aktivt ordboksverk för polacker: uttalsuppgifter om svenska ord. Här är ytterligare en mycket viktig – men ganska arbetskrävande – komplettering att göra i en eventuell andra upplaga, om resurserna räcker till, och liksom böjningsuppgifterna bör uttalsuppgifterna i så fall placeras i anslutning till de lemmatiserade svenska orden, alltså i SPO. (De polska

orden behöver däremot i allmänhet inga uttalsuppgifter eftersom deras uttal med få undantag är förutsägbara på grundval av stavningen; däremot behövs kortfattade polska uttalsregler.)

5. En närläsning av några utdrag ur PSO och SPO

Vi ska nu redovisa resultatet av en närläsning av 12 sidor (ca 2%) i vardera ordboken: s. 50–52, 180–182, 350–352 och 500–502 i SPO och s. 50–52, 200–202, 400–402 och 600–602 i PSO. Vi tar främst upp uppslagsord och fraser som ordböckerna ev. kunde kompletteras med resp. ev. överflödiga uppslagsord och fraser, ev. felaktiga eller icke-idiomatiska översättningar eller andra lingvistiska fel samt rena tryckfel.

a) SPO

Uppslagsord och fraser. Några tämligen överflödiga uppslagsord på de 12 sidorna är *begabba* (mycket ålderdomligt, vilket f.ö. inte anges), *guldvaskare* och *trumslagarpojke*; flera andra sammansättningar kunde nämnas. Några nu saknade ord som med betydligt större rätt skulle försvara en plats är *befryndad*, *begåvningsreserv*, *guide* (!), *gummiparagraf*, *guru* (bl.a. i frasen *ekonomisk guru*), *ostridig*, *osportslig*, *trofé* (!), *trollkrets* och *trubadur* (!). Åtminstone tre av dessa ord måste hänföras till kategorin otvetydiga missar. Helt optimalt tycks alltså inte lemmaturvalet i SPO vara, givet antalet 46.000, men antalet förbisedda goda uppslagsord tycks ändå inte vara alltför stort.

Mot uppsättningen av fraser och exempel finns inte mycket att anmärka. Möjligen saknar man fraserna *gå i skolan* och *gå runt* i den ekonomiska betydelsen i den imponerande artikeln **gå**, och frasen *jag är i begrepp att gå* kunde med fördel bytas ut mot *jag står i begrepp att gå*. Vidare är det tveksamt om man kan säga *gyckla över ngn*, och frasen *gå i otakt* hade försvarat en plats i artikeln **otakt**. Frasen *begränsade resurser* är upptagen under lemmat **begränsa** trots att lemmat **begränsad** finns. Men detta är relativa petitesseer; på det hela taget är behandlingen av fraseologi och syntaktiska konstruktioner imponerande i SPO.

Tryckfel o.d. Det måste nog sägas att ordboken hade mått väl av ytterligare en korrekturläsning; det är lite för många tryckfel. Oftast tycks det dock vara fråga om skönhetsfel, t.ex. typografiska fel, som inte behöver leda läsaren vilse. Här är några artiklar på de 12 sidorna där fel

stilsort och/eller grad i större eller mindre utsträckning insmugit sig: **beljuga** (en av bokstäverna i den polska ekvivalenten har råkat bli kursiv), **gynna** (*m el. f.* i st. *f. m el. f.*), **gå**, bet.moment 15, **ornat** (fel grad på fackområdesmarkeringen). Vi ska inte trötta läsarna mera med detta.

b) PSO

Uppslagsord, betydelser och fraser. Några saknade uppslagsord på de 12 sidorna som ev. skulle försvara en plats är *cichobieźny* ('tystgående'), *listewnik* ('ett slags brevlåda'), *pryncypa* ('chef'), *prywatyzować* ('privatisera') och *przebarwienie* ('för stark färgning'). Ett inte omistligt ord är *cichodajka* ('kvinna som då och då prostituerar sig'). I artikeln **ciemnieć** (huvudbet. 'mörkna') saknas betydelsen 'anta intensivare färg'. I artikeln **przebiec** saknar man kanske frasen *szmer przebieg po sali* ('det sorlade i salen'). Några andra fraser som kunde övervägas är *prywatna inicjatywa* ('privat initiativ'), *prywatna własność* ('privat egendom', ett viktigt begrepp i postkommunismens Polen!), *ciemny typ* ('en skum typ') och *cieszyć się na widok kogoś* ('vara glad att träffa ngn'). Överflödig är kanske frasen *lubić tańczyć* ('tycka om att dansa').

Översättningar. De svenska översättningarna är säkert så gott som alltid korrekta, men det händer att de får ett egendomligt stilvärde eller att de inte känns helt idiomatiska. Kubicki har bl.a. ambitionen att översätta polska ordspråk med svenska ordspråk också när dessa är tämligen obrukliga; ibland får man intrycket att han själv har skapat ordspråksliknande vändningar. Här måste man anbefalla den polske användaren en viss försiktighet. Det skulle göra ett komiskt intryck om en polack presterade svenska vändningar som *ett bugat huvud blir sällan avhugget, nyfikenhet driver fågeln i snaran* eller *ur ett tomt kärlek kan inte ens Salomo ösa*. Det är möjligt att dessa översättningar mest är avsedda för den svenske användaren, men hur ska den polske användaren veta att det inte är fritt fram att använda dem? Stor försiktighet är vidare av nöden vid användandet av de svenska könsord (*mus, kussimurra, månta* m.fl.) som PSO erbjuder en rik provkarta på, ibland med stilmarkeringen "posp" ('vardagligt') som klart otillräcklig varningsflagga. Några exempel på stilistiskt inte fullt lyckade översättningar är vidare *han är inte korkad* i st. *f. han är inte dum* samt *jag kan inte stå ut med sådana människor* i st. *f. jag står inte ut ...*

6. PSO som hjälpmedel vid läsning av en polsk dagstidning

Vi redovisar slutligen resultatet av läsning av tre längre artiklar – en sportartikel, en politisk artikel och en kulturartikel – i den stora dagliga tidningen *Gazeta Wyborcza* (31/7 1996) med PSO som hjälpmedel. Vi tar endast upp ord eller delbetydelser av ord som förekommer i artiklarna men saknas som uppslagsord i ordboken.

Det enda ordet i sportartikeln (s. 1 i tidningen) som saknades i PSO var *czasówka* ('tempolopp (på cykel)'). Lemmat *tempolopp* saknas f.ö. även i SPO. Dessutom förekommer i tidningsartikeln kollokationen *osada kajakarzy* ('kanotlag, kanotbesättning'), medan ordboken endast ger översättningen 'roddarlag' till *osada*.

Tre ord av betydelse i den politiska artikeln (s. 5) saknas i PSO: *legislacja* ('lagstiftning'; det polska ordet finns som ekvivalent till detta ord i SPO), *internetowy* ('hörande till Internet') och *minimalizować* ('minimera'; detta ord saknas som uppslagsord i SPO, vilket är en betydande miss).

Även kulturartikeln (s. 8) innehåller tre ord av viss betydelse som saknas i PSO: *akademizm* ('akademism'), *egzystencjalny* ('existentiell'; även om *egzystencjalizm* finns är betydelsen av det polska adjektivet inte helt härledbar från substantivet, lika lite som betydelsen av *existentiell* är härledbar från *existentialism*) samt *perfekcyjny* ('perfektionistisk' – detta ord liksom *perfektionist* saknas även i SPO, vilket är en klar miss).

Sammantaget gav läsningen av de tre artiklarna sju ord (och en kollokation) av viss betydelse som saknas i SPO. Det är nog en täckningsgrad som är ungefär vad man kan vänta sig av en ordbok med 37.000 uppslagsord. Möjligen bortsett från *legislacja* och *perfekcyjny* vill vi inte utan vidare hävda att något av de sju orden hade platsat i PSO med större rätt än de ord som nu är upptagna. Men slutsatsen av det lilla testet kan möjligen vara att det finns skäl att överväga att ta in ytterligare några tusen ord i nästa upplaga av PSO.

7. Sammanfattning

Kubickis ordboksverk innebär ett jättekliv framåt för den svensk-polska lexikografin. Att det är lätt att påvisa brister, delvis av rätt allvarligt slag, är egentligen bara naturligt med tanke på att ingen tillnärmelsevis jämförbar svensk-polsk eller polsk-svensk ordbok tidigare existerat. Sin stora styrka har ordboksverket i den relativt sett mycket grundliga behandlingen av fraseologi och valens. Också betydelsedifferentieringen är mycket väl genomförd. Lemmaurvalet är inte optimalt men acceptabelt, och det hade varit önskvärt med ett något större lemmabestånd. De övriga förbättringar som föreslagits i recensionen bör ha följande

prioritet i den andra upplaga, som förhoppningsvis kan komma till stånd så småningom: 1) Förbättring av behandlingen av de polska verbens aspekt 2) Förbättring av bruksanvisning och ordboksgrammatik 3–4) Införande av uppgifter a) om böjningen av polska substantiv och verb b) om uttalet av svenska ord 5) Makrostrukturen, bl.a. vissa egendomliga nischningar, bör ses över.

Referenser

- Berkov, Valerij 1987: *Russisk-norsk ordbok*. Moskva: Russkij jazyk.
- Berkov, Valerij 1996: Passive vs. Active Dictionary. A revision. I: Martin Gellerstam et al (eds.), *Euralex '96. Proceedings I–II*. Göteborg: Göteborg University, Department of Swedish, 547–550.
- Bzd'ga, Andrzej, Jan Chodera & Stefan Kubica 1973: *Podr'czny s[łownik polsko-niemiecki/Handwörterbuch Polnisch-Deutsch*. Warszawa: Wiedza Powszechna.
- Bzd'ga, Andrzej, Jan Chodera & Stefan Kubica 1978: *Podr'czny s[łownik niemiecki-polski/Handwörterbuch Deutsch-Polnisch*. Warszawa: Wiedza Powszechna.
- Malmgren, Sven-Göran 1995: Prepositionsfraser i tvåspråkiga ordböcker med svenska som källspråk: en undersökning och några förslag. I: *LexicoNordica* 2, 79–88.
- Svensk ordbok*. Utarbetad vid Språkdata, Göteborgs universitet. Sthlm: Esselte 1986.

Birger Lohse

Kåre Nilsson: *Norsk-portugisisk ordbok*. Oslo: Universitetsforlaget, 1994. X+594 s.. Pris 575,- n. kr.

The work under review, the revised and enlarged edition of Kåre Nilsson's Norwegian-Portuguese Dictionary, is the most comprehensive dictionary from a Scandinavian language to Portuguese and, probably, also one of the biggest encoding dictionaries to this language so far produced anywhere. It puts emphasis on the structural differences between the two languages, being particularly informative in the field of phraseology and idiomatic expressions. On all linguistic levels European Portuguese takes priority over the Brazilian variant.

Indledende bemærkninger

Den her anmeldte ordbog, Kåre Nilssons *Norsk-portugisisk ordbok* (herefter NPO), udgør en revideret og stærkt forøget udgave af forfatterens værk med samme titel fra 1979. Det var dengang et pionerarbejde i form af en offsettrykt, maskinskreven ordbog, der kom til at udfylde et længe følt behov for en seriøs ordbog til portugisisk, også i Danmark. Det produkt, som brugeren nu får i hænde, er det imponerende resultat af en erfaren leksikografs fortsatte virke. I elegant indbinding og med smuk klassisk typografi lægges der et meget overbevisende værk frem, ydedygtigt og lettilgængeligt på én gang.

Omslagsteksten oplyser, at NPO indeholder ca. 45.000 opslagsord, og at den rummer en mængde eksempler og idiomatiske udtryk. Dette sidste må, i betragtning af bogens helt overvældende rigdom inden for fraseologien, siges at være en pæn underdrivelse fra forlagets side. I omfang placerer denne anden udgave af NPO sig som den hidtil største ordbog mellem portugisisk og et skandinavisk sprog. Da det drejer sig om en aktiv ordbog, en indkodningsordbog, med kildesproget (norsk) som eksklusivt metasprog, er den formentlig også blandt de mest omfangsrige af denne ordbogstype til portugisisk i det hele taget, eftersom de store ordbøger fra hovedsprogene oftest sigter mod en portugisisk-talende målgruppe.

Anmeldelsen deler sig tematisk i to. Først gives en generel karakteristik af NPO som tosproglig ordbog, dels m.h.t. mere overordnede forhold, dels vedrørende diverse detailspørgsmål. På baggrund af NPO's prioritering af den europæiske variant af portugisisk behandler anmeldelsen derefter særlig de problemer, som fænomenet regional variation afstedkommer for den leksikografiske beskrivelse.

Generelle karakteristika

Makrostrukturelt er NPO enkelt opbygget, idet den, ud over ordbogsartiklerne på i alt 594 sider, blot omfatter en brugervejledning på knap fem sider. Der er således ingen grammatisk oversigt eller andre udenværker, hvortil der henvises på artikelniveau. Brugervejledningen består af lister over tegn og forkortelser, en gennemgang af principperne for opstillingen af det fraseologiske stof og diverse andre oplysninger. Efter almindelig leksikografisk konvention skelnes der i artiklerne typografisk mellem normal skrift, der bruges i målsprogets oversættelsesækvivalenter, og halvfed type, som er reserveret for lemma og for (i kursiv) kildesprogets fraseologi. Alle andre informationer, herunder kildesprogs-elementer, der optræder i forklarende parenteser, står i normal kursiv.

De seks anvendte typografiske tegn omfatter tilde, binde-, skille- og skråstreg, der tjener til afgrænsning og erstatning af ord eller orddele (fx "**afrikan**|er *m*, -**sk**"), samt parentes og henvisningspil. I oversættelsesækvivalenterne benyttes bindestreg uden forudgående afgrænsning vha. skillestreg, således at det overlades til brugeren at foretage den korrekte indsættelse af den udeladte orddele, fx "africano (*m*), -na (*f*)", "**adoptivbarn** *n* filho *m* adoptivo, -lha *f* -va", "**livstidsfange** *m* prisioneiro *m* perpétuo, -ra *f* -ua", men "**livstid**, **på** ~ perpétuo (fx *prisão p-ua*), *vitalício* (fx *renda v-ia*)", i enkelte tilfælde uden hjælp af bindestreg: "**ventilsete** *n* sede *f* da válvula (*pl* des das las)".

I stedet for distraherende koder, specialtegn, symbolske ikoner o.l. (en blomst for botanik og et tandhjul for teknik, bogstaver i grotesk for særlige sproglige registre etc.) benytter NPO udelukkende forkortelser, i alt ca. 140, skrevet i kursiv. Størstedelen vedrører specificering af fagsprogligt tilhørsforhold og semantisk-stilistiske præciseringer eller står for grammatiske termer. Disse er alle norske undtagen én ("*def*" for *pretérito definido*), som, en passant bemærket, ikke mere tilhører portugisisk grammatisk nomenklatur. Der er inkluderet forkortelser for tre regionale varianter af portugisisk: afrikansk, brasiliansk og europæisk (lusitansk) portugisisk (hhv. "*afr*", "*br*" og "*lus*"), og for et par målsprogsudtryk ("*ac*", "*alg*" og "*esp*" for hhv. *alguma coisa* 'noget', *alguém* 'nogen' og *especialmente* 'særlig'). I artiklerne forekommer derudover forkortelser af oversættelsesækvivalenter, der genoptages i skikkelse af forbogstav fulgt af punktum (fx "**abort** [...] **selvbestemt/ulovlig** ~ a. voluntário/clandestino").

Til forskel fra 1979-udgaven anvendes der ikke omløbende redaktion i form af underopslagsord, idet hvert lemma optræder med selvstændig indgang, bortset fra tilfælde som det allerede citerede "**afrikan|er** *m, -sk*", hvor opslagsordene umiddelbart følger hinanden. Det er en pladskrævende, men særdeles brugervenlig fremgangsmåde, der understreger bogens generelle præg af klarhed og overskuelighed. Til gengæld skal man, i hvert fald ud fra dansk kutyme, vænne sig til, at visse flerordsudtryk har egne opslag, selv om leddene skrives adskilt, alfabetiserede efter det første led, fx "**av gårde**" (med henvisning fra "gård"), "**De forente nasjoner**" (og "**forente nasjoner, De**"), "**etter som**" ('efterhånden som', der også figurerer under "**etter 2**"), "**i stand**" (med henvisning fra "stand", hvor der er andre udtryk med "i" og "stand"), "**så godt som**" (også under "godt"), "**så nær som**" (uden specifik henvisning).

Der synes kun at forekomme ganske få trykfejl i NPO. Af disse og andre formelle småfejl findes bl.a. følgende. Under "**adgang**" er i udtrykket (*med*) ~ *kjøkken* også badeværelset kommet med i oversættelsen: "com serventia de cozinha/de quarto de banho". Kursiv mangler ved genusangivelse af "**adressat**". Der er stavfejl i "instrucções" (under "**bruksanvisning**") og "práctica" (i "**bruksverdi**"), der skrives uden *c*, både i Portugal og i Brasilien. Tilde til genoptagelse af lemma mangler i "**bry** [...] *få/ha med*". I "**gjest**" er ("*fig*") trykt med fed kursiv. Ved at blive skrevet med stort sættes i artiklen "**gud 1**" *diabo* ('djevæl') på linie med *Deus* ('Gud'). I "**forbryte**" mangler det initiale *f*-. I "**se** [...] *vi får* ~ (*på det*) [...] vamos a ver" skal der ikke være præposition ("*a*") i det portugisiske udtryk. I oversættelsen til "**sentrallås**", "fecho centralizado" er *-a-* foran "*-do*" smuttet væk.

Grammatiske og fonetiske oplysninger

De grammatiske oplysninger vedrører primært genus og numerus, der meddeles både for norsk og portugisisk, herunder også særlige pluralis- og femininumformer i portugisisk, fx "*carácter m (pl caracteres)*" og "*filho m adoptivo, -lha f -va*" og tilsvarende eksempler nævnt ovenfor. Et specialtilfælde som *cais* ('kaj'), der har samme form i pluralis, er uomtalt under "**kai**", men anføres under "**brygge 1**" som ("*pl = sg*").

NPO informerer også om modusbrug. Det gælder ikke blot oplysninger om konjunktiv, hvis udbredte brug i portugisisk udgør en af de mærkbare strukturelle forskelle til de skandinaviske sprog, men også om indikativ, hvad der er noget mere usædvanligt i ordbogssammenhæng. Det er således en prisværdig service, når det fx om kompletivsætninger efter ækvivalenter til "**forekomme**" oplyses, at de får indi-

kativ: "*det ~r meg att* parece-me/afigura-se-me que (+ *ind*)", og når det ved "**kanskje**" ('måske') gøres klart, at modusbrugen veksler ikke blot afhængigt af, hvilket udtryk der anvendes, men også (ved *talvez*) af, hvilken position verbet indtager:

talvez (+ *konj el ind* + *talvez*), *se calhar* (+ *ind*), *pode ser que* (+ *konj*)

Tilsvarende hjælpes brugeren godt på vej i artiklen "**dem**", der klart søger at forebygge den fælde, som kan ligge i semantisk parallelle konstruktioner, der har divergerende syntaktiske konsekvenser: "*det er ~ som* (= *det finnes folk som*) *há quem* (+ *konj*), *há pessoas que* (+ *ind*)".

Det siger dog sig selv, at de grammatiske oplysninger vanskeligt kan være deskriptivt udtømmende i en standardordbog som NPO. Modusangivelserne vedrørende konjunktionen *se* i artiklen "**hvis**" må således forudsætte, at brugeren har et sådant kendskab til portugisisk, at han/hun ikke tolker informationen "(+ *fut konj om eventualitet*)", som om der nødvendigvis skal forekomme futurum konjunktiv efter *se* i denne situation. Den implicitte pointe er derimod, at det ikke går an at anvende præsens konjunktiv (der ellers benyttes efter andre betingelseskonjunktioner som *caso* og *a não ser que*), men netop kun futurum konjunktiv, også i nutidskontekster. I datidige sammenhænge er futurum konjunktiv til gengæld umulig efter *se*, og i stedet optræder imperfektum konjunktiv. Vildledende bliver det imidlertid, når det i samme artikel meddeles, at *se não* ('hvis ikke') tager indikativ. Denne modus kan udmærket forekomme efter *se*, men hvis den omtalte "eventualitet" gør sig gældende, dukker også her futurum (eller imperfektum) konjunktiv op, hvad enten der følger *não* eller ikke.

Oplysninger om målsprogets udtale finder almindeligvis ikke plads i indkodningsordbøger, men NPO indeholder dog i ny og næ information om portugisisk i så henseende. Det sker ikke gennem egentlig fonetisk transskription, men ved anvendelse af den ortografiske form af bogstavernes navne på portugisisk, fx i "*abc (utt á-bê-cê)*" og i artiklen om bogstavet A: "*a (utt á)*", hvor "*utt*" står for 'udtales'. Sjovt nok benyttes samme fremgangsmåde ikke ved E og O, hvor den portugisiske ækvivalent kun angives med bogstavets navn "*é*" og "*ó*". (Apropos alfabetet: de portugisiske bogstaver A, C og D er ikke femininum, men maskulinum som alle de øvrige; navnene på L, M og N skrives ikke med accent aigu; bogstavet Q bør tages med, selv om det ikke bruges i norsk, således at brugeren får adgang til den ikke helt uvigtige oplysning, at dette bogstavs navn er *quê*, udtalt [ke], og ikke [ku], der på portugisisk betegner en vis tendentielt tabuiseret legemsdel.)

Oversættelsesdelen

Der opereres med en ganske simpel semantisk hierarkisering blandt oversættelsesækvivalenterne. Den generelle betydning gives først, og diverse særbetydninger er derefter opført sideordnet, adskilt fra hinanden og fra den generelle (umarkerede) ækvivalent ved hjælp af semikolon. De forskellige betydninger karakteriseres ved et forklarende synonym eller anden form for præcisering, ligesom de af og til er forsynet med illustrerende eksempler. Samme oversættelsesækvivalent kan selvsagt forekomme i flere tilfælde, således som i følgende repræsentative eksempel:

gå ir, andar, caminhar (*til fots a pé*); (~ *sin vei*) ir-se embora, partir; (*forløpe*) (de)correr, passar; (*fungere, virke*) funcionar; (*ha avgang*) partir, sair; (*om tid*) passar(-se) (*fx passaram(-se) alguns dias*); (*om motor etc = være i gang*) estar em andamento; (*være i drift*) estar em serviço; (*være i trafikk*) circular; (*være mulig*) ser possível, poder-se fazer; (~ *bra/godt*) dar/bater certo; (*slukne*) apagar-se, extinguir-se; (*svikte, om lyspære/sikring*) ir-se abaixo, fundir-se; (*bli solgt*) vender-se, ter saída

De foreslåede oversættelsesækvivalenter af de norske ord og udtryk er altovervejende netop egentlige oversættelser, sjældent parafraserende forklaringer, og de fremviser et genuint idiomatisk præg. Dette kommer ikke mindst til syne i den righoldige fraseologi. Forklaringer eller forklarende tilføjelser, som det vil være kluntet at indsætte i en given tekstsammenhæng, finder man dog i enkelte tilfælde. I udtrykket "possessivo (falando-se de pais)", svarende til 'besidderglad (om forældre)', der gives som ækvivalent for "**overbeskyttende**", tjener det i parentesen inkluderende således snarere som metaleksikografisk kommentar for portugisisksprogede brugere af NPO end som del af en aktualiserbar oversættelse, og i "atraente de ponto de vista sexual" ('tiltrækkende fra et seksuelt synspunkt'), for "**sexy**", ville *atraente* alene formentlig være tilstrækkeligt til at sikre en adækvat oversættelse. Udtryk som "ficar à espera que as coisas melhorem", for "**bli siddende med hendene i fanget**", og "complicações daí resultantes", for "**komplikasjon [...] tilstøtende ~er**", lyder en anelse konstrueret; det første svarer vel til *ficar a ver navios*, og det andet kunne måske kunne gengives blot med *sequelas*. I "**like [...] ~ lite** tão pouco" med eksemplet "todas eram tão pouco bonitas" ('alle var lige lidt pæne') ville man nok også på portugisisk vælge en positiv udtryksform, så man helt slap for "tão pouco": *todas eram iguais de feias* ('alle var lige grimme'). Som

forslag til oversættelse af "**utålmodighet** [...] *beherske/stagge/tøyle sin* ~" virker "habituar-se a saber esperar" (egentlig 'vænne sig til at kunne vente'), også konstrueret; i øvrigt rammer *ter paciência* (egtl. 'have tålmodighed') vel bedre end "saber esperar" det norske udtryks betydning?

NPO medtager også egennavne, særlig toponymer og institutionsbetegnelser. Det komplicerede spørgsmål om brug af bestemt artikel ved stednavne synes i det store og hel løst korrekt. Fra p. 60 kan således citeres "Castela" ('Castilien') og "Cuba", som anvendes uden artikel, når der ikke følger attributivt led, men "a Catalunha" ('Catalonien'), "o Chile", "a Colômbia", "o Coliseu" ('Colosseum'), "a Córsega" ('Corsica') og "a Costa Rica", hvor den bestemte artikel under (næsten) alle omstændigheder skal med. Hvad angår de store syd- og vesteuropæiske lande, England, Frankrig, Italien og Spanien, hvis navne driller ved – i Portugal – at kunne benyttes både med og uden bestemt artikel, giver NPO "a Inglaterra" og "a França" uden yderligere kommentarer, d.v.s. som om de tilhørte gruppen af landenavne, der altid skal have bestemt artikel. Derimod er "a Itália" udbygget med "*fra/i* ~ de/em Itália", hvilket lader forstå, at artiklen ikke bruges efter præposition, hvad der ikke nødvendigvis er tilfældet, da *da/na Itália* også kan siges. Ved Spanien fås den mest præcise oplysning ved, at den bestemte artikel er sat i parentes: "(a) Espanha", samtidig med, at det eksplicit specificeres, at den oftest udelades efter præposition.

Adjektiver og adjektiviske afledninger til stednavne er et endeløst område. NPO bringer mange sådanne adjektiver til lande- og provinsnavne, men – naturligt nok – forholdsvis få navne på byer og disses adjektiver. Landet Rumænien er nævnt ("**Romania** a Romênia"), men et adjektiv svarende til dansk *rumænsk* (portugisisk *romeno*) synes at mangle, idet "**romansk**" ud fra oversættelsen ("românico") svarer til det tilsvarende ord på dansk. Af bynavne og adjektiver hertil finder man fx "**Lisboa**; *fra* ~ lisboeta, de Lisboa, alfacinha" (hvor det sidste, der egentlig betyder 'salatæder', burde forsynes med en stilistisk markering af uformelt sprog) og "**London** Londres; *fra* ~ londrino, de Londres". Fraværende er derimod *Porto*, med adjektiverne *portuense* og *tripeiro* (uformelt, 'indvoldsæder'), *Rio de Janeiro*, med *carioca* (fra byen Rio de Janeiro) og *fluminense* (fra delstaten af samme navn), og *São Paulo*, med *paulistano* (fra byen) og *paulista* (fra delstaten).

Repræsentativitet

Hvad angår ikke medtagne ord og udtryk, er man som dansksproget anmelder nødt til udvise forsigtighed over for en norsk ordbog. Selv om

norsk og dansk har mange lighedspunkter, findes den dansk-norske ordbog desværre ikke, der skal til for med sikkerhed at konstatere, om en givet glose eller betydning, som en dansktalende måtte savne, faktisk ikke er medtaget, fordi udtryk eller indhold er anderledes på norsk end på dansk. Det legendariske eksempel på norsk-dansk betydningsuoverensstemmelse, "**rar**", giver i NPO (selvfølgelig) oversættelserne "esquisito, estranho, extraordinário" ('mærkelig', 'sær' etc.), og ikke fx "simpático", der ville svare til ordets danske betydning. Termen "**russ**" oversættes med "finalista do liceu", der dækker den norske betydning af gymnasiast, der er ved at tage eller har taget studentereksamen, men ikke det danske *rus* ('studiestarter'), på portugisisk *calouro* (eller *caloiro*). Således har NPO "**barsel**" og tre sammensætninger hermed, men ikke fx *barselsorlov* (eller *orlov*), der står for et begreb, som givetvis må have et udtryk på norsk, men hvis tilstedeværelse anmelderen ikke umiddelbart er i stand til at kontrollere. Blandt (formentlig) konstaterbare lakuner o.l. kan dog nok udpeges følgende:

Under ordet "**asfalt**" mangler *alcatrão*, der bruges i Portugal, ligesom det afledte verbum "alcatroar", der anføres ved "**asfaltere**". I "**borgerlig** civil; (*middelklasse*~) burguês" er den politiske betydning, hvis portugisiske ækvivalent ville være *de direita* ('højre-'), ikke dækket. Ved "**brosje**" bør anføres *alfinete* og *pregadeira*, ud over "broche", der er efterhånden er tabuiseret. For "**elevator**" benyttes i Portugal *ascensor* ved siden af "elevador" (der i Lissabon også står for 'kabelsporvogn'), og "**flaskeåpner**" kan også hedde *tira-cápsulas*. Termen *pescado* ('fiskefangst') er ikke at finde under "**fisk**" eller "**fangst**". Ved "**godtgjøre**" og "**godtgjørelse**" mangler *ressarcir* hhv. *ressarcimento*, der ikke er ualmindelige. For "**hushjelp**" er *mulher a dias* (nævnt under "**rengjøringsassistent**") en god oversættelse i Portugal, svarende til *diarista* i Brasilien; for "**høyt teknologi**" kan også bruges *tecnologia de ponta*, og til "**håndtere**" *manusear*, parallelt til det heraf afledte substantiv "manuseio", som anføres under "**håndtering**". De to oversættelsesækvivalenter til "**kompromissløss**" er ikke synonyme, og kun "intransigente" dækker ordets danske betydning ('stejl', 'ubøjelig'), medens "sem compromissos" betyder 'ubunden', 'uden aftaler' eller 'uden forbindende'. For "**ligningsvesen**" kunne *repartição de finanças* være en mulig oversættelse i Portugal, og for "**motorsykel**" kan også (mere uformelt) bruges *mota*, i Brasilien også *motoca*. Ved fx idrætsanlæg hedder "**omklædningsrum**" *balneário* i Portugal. Under "**person**" oversættes "**juridisk** ~" med "entidade jurídica", men kan direkte gengives med *pessoa jurídica*, der i Portugal specificeres som *pessoa colectiva*, hvis det drejer sig om et selskab, og *pessoa singular*, når det er et individ. For "**scooter**" gives bl.a.

ækvivalenten "motoreta", der vel snarere svarer til en knallert, på norsk *moped*, hvor NPO også har ordet, som i øvrigt kun bruges i Portugal; i Brasilien siges *motoneta*. Under "**seddelomløp**" er "circulação de notas" ikke synonym med "massa monetária", der står for 'pengemængde' (også *meios de pagamento*). Til "**sekund**" og "**øyeblikk**" kunne under "*på et ~*" og "*på mindre enn et ~*" tilføjes *num piscar de olhos* og vel også *na hora*. For "**selvfinansiering**" er *autofinanciamento* nok en heldigere oversættelse end "autonomia financeira", til "**selvforsyning**" i økonomisk betydning passer *autarcia* godt; under "**sette ut**" mangler *despejar* som den juridiske gløse. For "**tak**" findes, når det specifikt drejer sig om taget på en bil, ordet *tejadilho*, i Brasilien *capota*. Appellativet "**tivoli**" gengives med "parque de diversões"; termen *feira popular* ville for en portugiser nok give de helt rigtige associationer. Ordet "**tjener**" oversættes med "criado", men har brugeren behov for at oversætte vokativ (*Tjener!* – som måske kun er en dansk usus?), kan han/hun ikke ty til *Criado!*, hverken i Portugal, hvor det hedder *Se faz favor!* eller *Faz favor!*, eller i Brasilien, hvor man, hvis det drejer sig om restauranter m.v., siger *garçom*, både i om- og tiltale. Til "**utsalg**" i den i NPO anførte betydning 'billigsalg' svarer vel, i Portugal, også termen *rebaixa*.

Regional variation i portugisisk

Det for portugisisk leksikografi så intrikate og langt fra afklarede problem med geografisk betinget variation må i en bilingval standardordbog begrænses til først og fremmest at angå afvigelserne mellem den europæiske og den brasilianske variant af sproget. Dels stiller spørgsmålet sig om den ortografiske norm, der er forskellig i Portugal og i Brasilien, dels er der afvigelser inden for ordforråd og fraseologi og af syntaktisk og semantisk karakter mellem sprogbrugen i de to lande (herefter benævnt P. og Br.). Ortografien er søgt uniformiseret i den seneste reform, der er på vej til at blive gennemført (og som NPO ikke har kunnet tage højde for), men der vil stadig være forskellig praksis i nogle detaljer. Portugisisk i Afrika følger normen i P., men udgør ellers ikke en så ensartet størrelse, at man i streng forstand kan tale om "afrikansk portugisisk" som andet end en samlebetegnelse for sproglige fænomener i et eller flere af de fem lande (Angola, Moçambique, Kap Verde, Guinea-Bissau og São Tomé e Príncipe), der har portugisisk som officielt sprog.

I NPO er der truffet det fundamentale strategiske valg at give forrang til europæisk portugisisk. Det hedder i brugervejledningen (p. x), at der medtages ord og udtryk typiske for brasiliansk sprogbrug, så-

fremt de tilsvarende i europæisk-portugisisk ikke (eller sjældent) finder anvendelse i Br. Det gælder endvidere, at det brasilianske stof markeres eksplicit som sådant i form af en regional specifikation: "(*br*)". En tilsvarende procedure gennemføres for eventuelle afrikanske varianter. De i P. anvendte ord og udtryk udgør derimod "default"-muligheden, og de fremstår derfor som regionalt umarkerede. NPO meddeler udelukkende europæisk-portugisisk ortografi, undtagen i særbrasilianske ord og udtryk, og brugeren må derfor selv foretage den nødvendige adaptering til brasiliansk norm i de tilfælde, hvor kun stavemåden er forskellig.

Modellen for repræsentation af regional variation i NPO bliver således: Hvis ingen regional markering, så er det givne ord eller udtryk brugbart overalt i den portugisisktalende verden; hvis en regional markering specificerer et senere ord eller udtryk, så anvendes de(t) foregående ikke (eller sjældent) i det pågældende område. Regional markering vil i praksis sige brasiliansk portugisisk, eftersom de afrikanske særformer er ganske fåtallige. Det følgende eksempel, der (uden genusmarkering) gengiver oversættelsesækvivalenterne til "**buss**", er repræsentativt for NPO's metode:

autocarro, camioneta; (*br*) ônibus; (*afr*) machimbombo

De to første og umarkerede ækvivalenter angives som gangbare i europæisk portugisisk (med betydning af 'bybus' hhv. 'rutebil'), men ikke i brasiliansk eller afrikansk portugisisk, hvor i stedet den tredje resp. den fjerde ækvivalent anvendes. Tilsvarende i fx "**buss**holdeplass *m* paragem de autocarros/camionetas; (*br*) parada de ônibus", hvor det første, umarkerede udtryk – gennem markeringen på det næste – vises som ikke (eller mindre) anvendeligt i Br., der i stedet benytter det andet. Som i andre sammensætninger med "buss" nævnes den afrikanske variant ikke, og brugeren må vel så konkludere, at det går an at benytte det umarkerede udtryk i det portugisisktalende Afrika, da det andet er specificeret som særskilt for brasiliansk portugisisk.

Alternativ markering af regionale varianter

I enkelte tilfælde afviges der i NPO fra den vedtagne fremgangsmåde, hvad angår markering af regional variation, således at også det i P. benyttede udtryk markeres eksplicit, ikke kun de i Br. og Afrika forekommende. Der synes dog ikke være nogen substantiel forskel på disse situationer og de to førnævnte model-eksempler. Markering af europæisk portugisisk (med "(*lus*)") finder sted i fx "**etternavn** *n* (*lus*) apelido, (*br*) sobrenome *m*", "**fotgjenger** *m* (*lus*) peão *m*, (*br*) pedestre

m", "**slakterbutikk** *m (lus)* talho, (*br*) açougue *m*", "**vindusutstilling** *m (lus)* montra *f*, escaparate(s) *m(pl)*, (*br*) vitrina" og "**økenavn** *n (lus)* alcunha *f*, (*br*) apelido *m*". Denne alternative, men mere omstændelige metode til præcisering af regional variation, ville nok principielt være at foretrække i alle tilfælde, da den indebærer bedre mulighed for entydig information. Et par eksempler kan illustrere problematikken:

I "**kø** *m* bicha, fila *f* de espera, (*br*) fila (de espera); (*av kjøretøyer*) fila, caravana *f* de veículos" fås en egentlig unødvendig gentagelse af samme glose, "fila", der – som det ses – er brugbar både i P. (også uden "de espera") og i Br. Da den prekære term ikke er "fila", men derimod "bicha", som det ikke går an at benytte i denne betydning i Br., hvor ordet er slang for 'homoseksuel', kunne en enklere løsning på markering af regional variation i dette tilfælde være kun at specificere "bicha", nemlig med "*(lus)*", men at lade "fila (de espera)" stå umarkeret. Tilsvarende i artiklen "**jente**", der har følgende oversættelsesækvivalenter: "rapariga, menina, (*br*) moça, garota *f*; (*fam*) tipa, gaja". Der er ikke noget i vejen for også at anvende "menina" i Br. eller "moça" i P., selv om den præcise aldersmæssige definition kan variere (jf. dansk 'pige'), medens "rapariga" i Br. har fået betydningen 'prostitueret'. Det er derfor ikke mindst dét ord, som det gælder om at få ordbogsbrugeren til ikke at tage i anvendelse i den forkerte sammenhæng. Med en specificering af regional variation, der stiller varianterne lige, ville i stedet "rapariga" blive markeret som kun gangbart i P. (eventuelt netop med restriktionen "kun" som tilføjelse til "*(lus)*"), til forskel fra "menina" og "moça" uden markering og "garota" med "*(br)*". (Af de familiære udtryk er "gaja" eksklusivt for P., medens "tipa" har en mere nedsættende klang i Br.).

Ortografiske og grammatiske divergenser

Som nævnt anfører NPO alene europæisk-portugisisk retskrivning. Den ovenfor omtalte adaptering til brasiliansk skrivemåde kan dreje sig om majuskelbrug (fx i månedernes navne, som i P. skrives med stort, men i Br. med lille), og accenter, hvor en accent aigu foran nasalkonsonant i P. modsvares af en accent circonflexe i Br. (fx *académico* og *agrônomo* vs. *acadêmico* og *agrônomo*). I andre tilfælde er accentueringen mere uforudsigelig: det hedder fx *estômago* ('mave'), *fêmea* ('hun(dyr)') og *café* både i P. og Br., men *bebé* ('baby'), *comité* i P., *bebê* og *comitê* i Br. Forskellig skrivemåde har også ord med *gu* eller *qu* foran *e* eller *i*. Her benyttes trema i Br., men ikke i P., til at angive, at *u* skal udtales, fx i *lingüística* (Br.), *linguística* (P.), udtalt med [-gwi-], til forskel fra den hyppigere ortografiske funktion blot at signalere klusil udtale af *g* og *q*.

fx i *guitarra*, med [gi-]. En anden ortografisk forskel hidrører fra de "stumme", men etymologisk berettigede konsonantgrafemer *c* og *p*, der (endnu) bibeholdes i europæisk portugisisk med den primære ortoepiske funktion at angive "åben" udtale af den foregående ubetonede vokal. Dette er irrelevant for udtalen i Br., hvor disse konsonanter derfor ikke skrives: *abstracção* og *abstracto* i P. vs. *abstracção* og *abstrato* i Br., *adjectivo* (P.) vs. *adjetivo* (Br.), *tecto* ('tag') i P. vs. *teto* i Br., *optimismo* og *ótimo* ('fremragende') i P. vs. *otimismo* og *ótimo* i Br. Desværre kan man ikke blot automatisk slette det "overflødige" bogstav og derved gå fra ortografien i P. til ortografien i Br. Det hedder fx *espectáculo* ('skuespil') i P. og *espetáculo* i Br., parallelt til de foregående eksempler, men *espectador* ('tilskuer') i begge lande, trods forskellig udtale, idet *c*'et, der er "stumt" i P., faktisk udtales i Br., hvor det derfor også skrives.

NPO's valg kun at medtage den euro-portugisiske variants ortografi er forståeligt, først og fremmest af pladshensyn, men det tvinger den bruger, der ønsker at anvende den brasilianske norm, til ekstra opslag i en anden ordbog for at checke skrivemåden, når der kan være tvivl. Og for den mindre øvede kan tvivlen også bestå i netop ikke at kunne overskue, hvornår der måtte være tvivl. Særlig slemme er helt uforudsigelige tilfælde som oversættelserne til "**fukt**", "**fuktig**" etc., hvor der i P. skrives med initialt *h*-, men ikke i Br. (til forskel fra andre ord med "stumt" *h*); her har NPO "*humidade*", ikke *umidade* (etc.). Tilsvarende ækvivalenterne til "**jugoslav**", "**Jugoslavia**" etc., der i P. har *j*-, men i Br. *i*- ("*Jugoslávia*" vs. *Iugoslávia*). I talordene 16, 17 og 19, der har forskellig form i P. og Br., giver NPO kun ved 16 den brasilianske variant, men ikke ved 17 og 19.

På det grammatiske plan sætter NPO's europæisk-portugisiske prioritering sig igennem ved, at oversættelserne af udtryk med 'du' selvfølgelig benytter 2. person singularis. Brugere, der ønsker at anvende brasiliansk variant, kan uden større besvær transformere verbet til 3. person, også når det drejer sig om tiltale. Netop tiltaleformerne i portugisisk udgør et vanskeligt felt at håndtere, også leksikografisk. Mangfoldigheden i europæisk portugisisk, hvad angår høflig (eller bedre: ikke-uformel) tiltale, får man et indblik i gennem artiklen "**De**", der præsenterer brugeren for ækvivalenterne "*o senhor, a senhora*" og "*Vossa Excelência*" samt (med kommentaren "*upersonlig, ofte litt nedlatende*") også "*você*". Det ville sprænge rammerne for en almindelig ordbogsartikel at redegøre for de øvrige mulige paralleller til "**De**" (brug af *o senhor* plus titel, brug af navn, med forskel mellem kønnene m.h.t. for- og efternavn, etc. etc.); men blot mellem *o senhor* og *a senhora* på den ene side og *Vossa Excelência* på den anden findes der et pragmatisk skel, der gør, at de ikke uden videre kan sidestilles.

Hvad angår brasiliansk portugisisk, er situationen simplere, og den leksikografiske opgave lettere at løse: Til "De" svarer *o senhor* (og *a senhora*), til "du" derimod *você*, som i Br. ikke er belastet med de konnotationer, som ordet kan have i P. NPO registrerer under "**du**" imidlertid udelukkende "tu", der bruges i P., men kun i meget begrænset omfang i Br. (og i standardsproget med konnotationer), hvor netop *você* er den oplagte ækvivalent til "du". Samme praksis følges naturligt nok under "**dig**", hvor NPO også lader forholdene i brasiliansk portugisisk uomtalte. Som objekt kan *você* ikke bruges i P., men udmærket i Br., hvor de i artiklen anførte former ("te" og "ti") også benyttes, især i talesproget, som alternativer til *você*. Den oblique form til "De", *Dem*, mangler til gengæld i NPO (det bruges måske ikke i norsk?). Såvel "**de**" som "**dem**" og "**deres**" (med lille) forekommer, og også genitiven "**Deres**" af "De" (med stort). Det ser derfor ud til, at den vigtige oplysning, at man i P. bruger *si*, præpositionsformen af det refleksive *se* 'sig', i tiltale som oplagt oversættelse af "Dem", fx i *isto é para si* 'dette er til Dem', heller ikke er kommet med. Under "**deres**" (med lille) gives for 2. person "vosso" ('jeres'), der bruges i P., men ikke i Br. (bortset fra kirkeligt sprog og andre særlige registre), hvor *de vocês* er den normale udtrykform. Ved "**dere**" er ækvivalenten "vocês" ('I', 'jer'), anvendelig både i P. og Br., og den i store dele af P. stadig anvendte (ældre) form *vós* nævnes ikke.

Brasilianske specialiteter

Som et beskedent supplement til de fænomener inden for regional variation, som NPO registrerer, er her afslutningsvist samlet nogle ord, udtryk og betydninger, hvor den brasilianske variant skiller sig ud fra europæisk portugisisk. Kommentarerne, der forhåbentlig kan være forfatteren til nytte ved den "neste korsvei", som han i forordet ser frem til, er ordnet alfabetisk efter opslagsord og omfatter iagttagelser om stort og småt fra NPO's knap 600 sider:

"**absolutt**" som adverbium ("absolutamente") har bekræftende betydning i P., men nægtende i Br. (= 'absolut ikke', 'bestemt nej' o.l.); "**adamsæble**": også *pomo-de-adão*, i Br. tillige *gogó*; "**AIDS**": "SIDA" bruges i P., *AIDS* i Br.; "**antikvar(bokhandler)**", "**antikvariat**": "alfarrabista" er ukendt i Br., hvor *antiquário de livros* eller, især, *sebo* benyttes; "**asparges**": i Br. foretrækkes formen med *asparago*, til forskel fra "**espargo**" i P.;

"**badedrakt**": i Br. *roupa de banho* eller *maiô* (kun for kvinder); "**barnehage**": "*infantário*" ('vuggestue') og "*jardim-escola*" ('børnehave') bruges ikke i Br., hvor det første også kan hedde (*escola*) *maternal*, og det sidste også *jardim de infância*, ligesom i P. (hvor også (*escola/colégio*) *infantil*); "**billett**": "bilhete" i P., men *entrada* eller *ingresso* i Br. til teater, sportsarrangementer m.v. og *passagem* til transport; "**billettkontor**": "bilheteira" i P., men *bilheteria* i Br.; "**bordlampe**": i Br. *abajur*; "**borgermester**": i Br. *prefeito*; "**bruger**"; i Br. *usuário*; "**brun**": det gængse ord i Br. er "marrom"; "**brunkull**" – "lignite": i Br. foretrækkes i kemiske termer på *-ite* endelse på *-a*, fx *grafita* 'grafit', men for *lignita* (eller *linhita*) er opstået formen *linhito* (femininum);

"**detaljhandel**", "**detaljhandler**": i Br. bruges termen *varejo* og afledninger heraf (*comércio a/no varejo*, *varejista*); "**diesel**": "gasóleo" kun i P., i Br. (*óleo*) *diesel*; "**direkte**", om radio og tv: i Br. *ao vivo*; "**dusj**", "**dusje**" (1 og 2): formen "duche" anvendes ikke i Br., derimod *chuveirada* (*tomar uma chuveirada*) og *banho de chuveiro*;

"**finansdepartement**" (og "**finansminister**"): i Br. *Ministério* (og *Ministro*) *da Fazenda*; "**fotokopi**", "**fotokopiere**": i Br. især *xerox*, *xerocar*; "**fruktsaft**": i Br. *suco*; "**fryseboks**", "**fryser**": i Br. *congelador horizontal*;

"**gode**" i betydningen 'fordel', 'frynsegode': i Br. er *mordomia* den præcise gløse; "**godt**": *den/det hadde du bare ~ av*: i Br. *bem feito!*; "**holde seg**" ~ (om personer): i Br. også *ainda valer meia sola*; "**godtatt**": i Br. er det uregelmæssige participium af verbet *aceitar* med *-o*, *aceito*, til forskel fra "*aceite*" i P.; "**gulvteppe**": "*alcatifa*" ('væg til væg tæppe') hedder i Br. *carpete* (som i P. er et større, løstliggende tæppe); "**gård**" ('bondegård'): i Br. *fazenda* (eller *sítio*, hvis mindre);

"**hallo**" hedder i Br. *alô*, i betydningen 'hej' også *oi* og *oba*; "**hulkapsel**": i Br. *calota*; "**hodepute**" er i Br. kun *travesseiro*, ikke "almofada"; "**håndvask**" (genstanden, ikke processen): i Br. *pia*;

– "**identitetskort**": i Br. *carteira de identidade*; "**innkjøring** [...] ~ **forbudt**": *contra-mão* i Br.; for "**i stykker**": "avariado" og "avariar" (*gå ~*) i normal daglig betydning bruges *quebrar*, *entra em pane* eller *dar o prego* i Br. (hvor "avariar" nærmest svarer til 'havarere');

"**K**": "capa" er kun i P. bogstavets navn; "**kanskje**": "se calhar" er typisk for P. og høres næppe i Br.; "**kartotek**": i Br. *fichário*; et ord for "**kjeller**" karakterisk for Br. er *subsolo* for 'kælleretage'; "**kjøleskap**": i Br. *geladeira*, medens "frigorífico" betyder 'kølehus'; "**kommune**": i Br. betegner "autarquia" mere end en kommune, og "concelho" bruges ikke; "**kontroll**" i Br. *controle*; "**kopp**": "chávena" er i Br. en *tekop*; i "~ **kaffe**" er de anførte udtryk eksklusive for P.: "bica" (i denne betydning) svarer i Br. til *café expresso* eller "cafezinho" (der ikke er

sædvanlig i P.) og "calão" (glas kaffe med mælk) hedder *média* i Br.; "**kran(bil)**": "guindaste" er OK i Br., men ikke "grua", der kun anvendes i P. (hvor det også hedder *reboque*); "**København**": "Copenhaga" bruges udelukkende i P., medens *Copenhagen* er gangbart både i Br. og (stadig mere) i P.;

"**lagerbeholdning**", "**lagre**" og "**lagring**": i Br. *estoque*, *estocar* og *estocagem*; "**lampeskjerm**": i Br. *quebra-luz* ("abajur" betyder 'bordlampe'); "**leilighet**": i Br. kun "apartamento", medens "andar" er 'etage' (som også i P.); "**livredder**": *salva-vidas* i Br.;

"**man**": "uma pessoa" er typisk for P., i Br. evt. *a pessoa* (eller de i NPO omtalte andre udtryksmåder som brug af refleksivpronomet *se*, 'sig', etc.); "**mindre**": "mais pequeno" går ikke an i Br., hvor "menor" anvendes også om konkret størrelse; "**Moskva**": "Moscovo" hedder i Br. *Moscou*; "**muggen**" ('sur', 'vred'): i Br. *enjoado*; "**mun**", *være stor i ~en*: i Br. *ser faroleiro*; "**mynt**", *betale i/med klingende ~*: i Br. *pagar em metálico/numerário*; "**MVA**" ('moms') i Br. *ICMS (= imposto sobre a circulação de mercadorias)*;

"**NATO**": den korrekte portugisiske forkortelse "OTAN" bruges især i Br., medens man i P. anvender den engelske (i femininum, *a NATO*);

"**overdra**", "**overdragelse**": i Br. *traspasar*, *traspasse*;

"**parkere**": "arrumar" og "aparcar" kun i P., "estacionar" også i Br. (og *parquear*); "**pengelens**": i Br. (også *duro*, *liso*, *na lona* (uden "-s")); "**pensjon** [...] (*alders~*)": "reforma" bruges i Br. for militærfolk, ellers *aposentadoria* (som NPO har under "**pensjonsalder**"); "**pepper**": "pimenta" betyder i Br. 'chilipeber', og almindelig sort peber hedder *pimenta-do-reino*; "**politistasjon**": i Br. *delegacia de polícia*, "esquadra" kun i P.; "**pulterkammer**": i Br. *depósito*, *despensa* eller *quarto de guardados*, men "arrecadação" kun i P.; "**polak**", "**polsk**": i Br. *polonês*, ikke "polaco" (og slet ikke i femininum, "-ca", der bl.a. betyder [udenlandsk] prostitueret); "**pose**": i Br. bruges *sacola* i betydningen 'indkøbspose'; "**pølse**": "salsicha", der (med varianten *salchicha*) bruges i P., svarer i Br. bruges til *lingüiça* for samme spise;

"**regning**" (dokumentet) "factura" svarer i Br. til *nota (fiscal)*; "**rutebilstasjon**": standardtermen i Br. er *rodoviária*;

"**selvmål**": *autogolo* i P., *gol contra* i Br., snarere end "golo/(br)gol *m próprio*"; "**seks**": for at undgå forveksling med *três* bruges i Br. *meia(-dúzia)* for "seis" i telefonsprog o.l.; "**servicestasjon**": i Br. *posto de gasolina*; "**sentrum**": "baixa" kun i P.; "**skinke**": "presunto" hedder *presunto cru* i Br. "fiambre" *presunto* ('kogt skinke'); "**skudd**": "chuto" i P., men *chute* i Br.; "**sovevogn**" i Br. *vagão-leito*; "**spotsk**", "**spotte**" (etc.): i Br. (også) *gozador*, *gozar*; "**sted**": "sítio" benyttes i denne

betydning kun i P., men betyder i Br. især '(lille) landejendom/(bonde)gård';

"**tale**", *ikke ~om!* i Br. typisk *de jeito nenhum!*; "**tegniserie**": i Br. bruges kun udtrykket "história em quadrimhos", ikke "história aos quadrimhos"; "telefonsvarer": i Br. *secretária eletrônica*; "**turné**": "digressão" kun i P., men i Br. *turnê*;

"**undergrunnsbane**" i Br. *metrô*, ikke "metro"; "**utstede**" ("**utsteder**" etc.) om pas m.v.: i Br. også *expedir (expedidora)*;

"**ved**": i Br. bruges *ao lado de* i st. f. "ao pé de" (egtl. 'ved foden af'), der er typisk for P.; "**veibom**": i Br. *pedágio* eller *rodágio*; "**veikant**": i betydningen 'rabat' bruges i Br. *acostamento*; "**W**": "vê dobrado" kun i P. (hvor også *duplo vê*), i Br. anglicismen *dáblio*.

Sammenfatning

Leksikografens dont er utaknemmelig. En pernittengrynet anmelder evner altid at lede et eller andet frem at sætte fingeren på i det færdige værk, der hyppigere bliver genstand for ris end for ros. Som sammenfatning af denne anmeldelse er det derfor værd klart at fremhæve *Norsk-Portugisisk Ordbok* som et gennemprofessionelt kvalitetsarbejde, der fundamentalt præges af et indgående og reflekteret forhold til de to beskrevne sprog, og som ikke mindst udmærker sig ved en særdeles omfattende dækning af det fraseologiske stof. Hovedvægten er lagt på europæisk portugisisk, hvorfor en øget inddragelse af den brasilianske variant, hvis relative vægt bliver stadig mere mærkbar, fremstår som det væsentligste indsatspunkt ved en videre udbygning af ordbogen som et effektivt kommunikationsværktøj mellem den skandinaviske og den portugisisksprogede kulturkreds.

Sölve Ohlander

"The Big Four". "Learners' dictionaries" inför 2000-talet

Cambridge International Dictionary of English 1995. Editor-in-chief: Paul Procter. Cambridge University Press.

Collins COBUILD English Dictionary, Second ed. 1995. Editor in Chief: John Sinclair. HarperCollins.

Longman Dictionary of Contemporary English, Third ed. 1995. Director: Della Summers. Longman.

Oxford Advanced Learner's Dictionary of Current English, Fifth ed. 1995. Editor: Jonathan Crowther. Oxford University Press.

This review article discusses certain general tendencies and issues in connection with the "The Big Four", i.e. the 1995 editions of the four English learners' dictionaries now commonly regarded as dominant. One readily noticeable trend is that, on a fiercely competitive market, learners' dictionaries are becoming increasingly bulky, covering more and more words and providing ever new kinds of information. The 1995 dictionaries are all based on enormous computerized corpora of both written and spoken language. While this does not eliminate gaps in coverage, it does mean, among other things, that informal spoken English can be paid special attention to. On the basis of a learners' corpus, one of the dictionaries also provides contrastive information concerning English "false friends" in relation to a number of different languages, an interesting attempt, albeit poorly executed, to bridge the gap between monolingual and bilingual lexicography. Another novelty is the introduction of corpus-based frequency information in two of the dictionaries. With regard to definition, all four dictionaries now employ restricted defining vocabularies, whereas there is no agreement on definition format or style. Semantically, there is a clear tendency towards eliminating, in the presentation of word meanings, the distinction between homonymy and polysemy, a lexicographically convenient strategy which, however, may be pedagogically questionable. The article ends with some reflections on the possible limits, not least from a pedagogical point of view, of traditional, printed learners' dictionaries, especially in comparison with the potential of future electronic dictionaries.

Inledning

1995 var ett stort år för engelska "learners' dictionaries". Som för att lägga sig i läge inför nästa årtusende utkom detta år de tre största i

branschen med nya, i flera avseenden kraftigt omarbetade upplagor: *Collins COBUILD English Dictionary (COBUILD)*, *Longman Dictionary of Contemporary English (LDOCE)*, *Oxford Advanced Learner's Dictionary of Current English (OALD)*. Samma år anmälde sig en fjärde intressent till medlemskap i denna grupp av väletablerade och arga konkurrenter på en lukrativ marknad – en uppkomling från Cambridge University Press: *Cambridge International Dictionary of English (CIDE)*. Den tidigare treklövern har förbytts i "The Big Four".

I den allt hårdare striden om marknadsandelar inom det enorma fält som engelska som andraspråk och främmande språk utgör ligger det i varje ny lexikonupplagas intresse att hävda sitt eget nyhetsvärde så otvetydigt som möjligt; blygsamhet är här en tvivelaktig dygd. Pionjären bland "learners' dictionaries", *OALD* (första upplagan 1948), omnämner sin senaste version som "This exciting new edition" (pärmens baksida); motsvarande karakteristik av den närmast föregående, fjärde upplagan från 1989 var "This comprehensively revised and greatly expanded new edition..." (pärmens baksida).¹ Därefter i senioritet följer *LDOCE* (första upplagan 1978), vars "third, completely new edition" (s. xi) "...reveals for the first time the richness of the English language" (pärmens baksida); andra upplagan av lexikonet (1987) varudeklarerades på sin tid som "A totally revised version..." (pärmens baksida). Nya *COBUILD* presenteras på följande sätt av John Sinclair: "This is a new book, a completely new edition..." (s. viii); om den första upplagan av *COBUILD* (1987) lät det så här (omslaget baksida): "The dictionary sets new standards in the presentation and layout of detailed information about the language."² Jämfört med självsäkerheten hos de redan etablerade lexikonen ter sig nykomlingen i gänget, *CIDE*, närmast blygsam: "The fresh approach that we have taken should appeal to all those who appreciate good lexicography..." (s. viii). Till grund för denna relativt återhållsamma formulering ligger måhända den omständigheten att *CIDE* är Cambridge University Press' första inbrytning bland "mainstream monolingual dictionaries" (s. viii).

Som kan utläsas av de ovannämnda citaten har det i accelererande grad gått inflation i uttryck för självuppskattning, med stark betoning

¹ Som en lätt pikant detalj kan man lägga märke till att den senaste upplagan av *OALD* presenterar sig som "The dictionary that really teaches English" (pärmens baksida), säkerligen en medveten travesti på *COBUILD*s välkända slogan: "Helping learners with real English."

² *COBUILD* har också förkortat sitt namn sedan första upplagan 1987 (*Collins COBUILD English Language Dictionary*), kanke för att markera att det verkligen rör sig om en helt ny upplaga. Detta framgår också av att samtliga 100 000 exemplar i lexikonet "are new for this edition" (pärmens baksida).

på varje ny lexikonupplagas omistlighet för användaren. Men trots allt finns det i 1995 års utgivning ett visst fog för den av förlagen eftertraktade uppmärksamheten. Skälet är att innovationerna i de nya lexikonerna till betydande del är av kvalitativ snarare än blott och bart kvantitativ natur. Det handlar således inte enbart om att lexikonerna innehåller fler och nyare uppslagsord än sina föregångare, utan lika mycket om vilket slags information som ges, hur den är strukturerad etc. Denna artikel avser inte att i detalj granska varje enskilt lexikon; vart och ett av dem har sina egna specialiteter. I stället kommer jag att uppehålla mig vid ett antal någorlunda klart urskiljbara tendenser av mer övergripande natur, särskilt sådana som ger sig till känna i fler än ett av lexikonerna. Detta innebär i huvudsak jämförelser på vissa punkter såväl mellan de olika lexikonerna från 1995 som mellan ny och tidigare upplaga av samma lexikon.³ Även för enstaka lexikon karakteristiska "nischer" kan i en del fall vara värda att uppmärksamma; dagens nischer kan vara morgondagens – eller i alla fall nästa århundrades – trender.

Tätare utgivning, större omfång

En skönjbar tendens, som kan kopplas till den allt hårdare konkurrensen men också till den snabba utvecklingen inom lexikografin, är att det i framtiden förmodligen kommer att bli ännu tätare mellan upplagorna av samma lexikon. Mellan tredje och fjärde upplagan av *OALD* dröjde det femton år (1974–1989), mellan fjärde och femte sex år. Första upplagan av *LDOCE* kom 1978, andra 1987, tredje 1995, det sista intervallet samma som mellan först och andra upplagan av *COBUILD*. Lexikonet som färskvara i en snabbt föränderlig språkvärld kommer säkert att betonas allt mer i marknadsföringen, gärna med argument hämtade från språkpedagogisk forskning kring inlärning och undervisning av andraspråk och främmande språk.

Hand i hand med tendensen mot allt tätare utgivning av nya upplagor går en tydlig tendens mot alltmer omfattande inlärningsordböcker. En jämförelse mellan sidantal i olika upplagor ger en grov indikation:⁴

³ Vid jämförelser med närmast föregående upplagor av lexikonerna kommer följande beteckningar att användas: *OALD* (89) = fjärde uppl. av *OALD*, 1989; *COBUILD* (87) = första uppl. av *COBUILD*, 1987; *LDOCE* (87) = andra uppl. av *LDOCE*, 1987. I vissa fall kan även beteckningar som *OALD* (95) etc. förekomma.

⁴ Sifferuppgifterna avser endast den egentliga, alfabetiskt uppställda lexikondelen i böckerna, alltså inte olika "introductions", tabeller, listor eller andra slags appendix, där det finns stora skillnader mellan lexikonerna. Som jämförelse kan här

LDOCE (87) 1229 sidor – *LDOCE* (95) 1668 sidor, *COBUILD* (87) 1703 sidor – *COBUILD* (95) 1951 sidor. Det nya *CIDE* omfattar 1701 mycket tättryckta sidor. *OALD* utgör dock ett intressant undantag till den allmänna trenden: *OALD* (89) 1492 sidor – *OALD* (95) 1392 sidor (dock med något mindre stil). Även i fråga om sidstorlek avviker *OALD* från de övriga lexikonerna genom sitt mindre format. Huruvida *OALD* på detta sätt medvetet profilerar sig gentemot de andra, mer voluminösa lexikonerna är svårt att veta. Men frågan var gränsen går beträffande informationsmängd – praktiskt/pedagogiskt och även kommersiellt – rycker närmare för varje ny upplaga. Det rör sig inte enbart om antalet uppslagsord, definitioner och exempel, utan också om all den övriga information kring ords användning som inlärningslexikon under senare år visat allt större intresse för. Kanske håller vi på att närma oss den punkt där, för ordboksanvändaren, problemet inte längre är brist utan överflöd på information. Hela denna fråga hänger samman med en annan fråga, nämligen hur "learners' dictionaries" i praktiken används, ett område där vår kunskap är begränsad.⁵

Fler och fler uppslagsord

En bidragande orsak, om än inte den enda eller kanske ens den främsta, till den förhärskande färdriktningen mot allt större lexikon är ambitionen att i varje ny upplaga spegla den snabba utvecklingen inom engelskans ordförråd. Så länge tillägg av nya uppslagsord och nya betydelser inte i motsvarande grad kompenseras genom eliminering av lika många föråldrade ord och betydelser kommer antalet uppslagsord och betydelser i "learners' dictionaries" ständigt att öka. Några sifferuppgifter kan illustrera detta.⁶

Trots relativt oförändrat sidantal omfattar den nya upplagan av *OALD* "63 000 references", att jämföra med "57 100 words and phrases" i *OALD* (89); den nya upplagan uppges innehålla "2 300 new

också nämnas att åttonde upplagan av *Concise Oxford Dictionary* (1990) omfattar 1428 sidor.

⁵ På frågan om hur mycket vi egentligen vet om hur "learners' dictionaries" används ger Kernerman (1996:406) ett koncist svar: "Very little."

⁶ Exakta jämförelser försvåras av bristen på en uniform terminologi vid lexikonens egna uppgifter om antal uppslagsord. Där ett lexikon talar om antalet "words and phrases" talar ett annat – eller en ny upplaga av samma lexikon – i stället om "references"; *Concise Oxford Dictionary* (1990) uppger "120,000 entries", varav "Over 20,000 entries new to this edition". Uppgifter om eliminerade uppslagsord eller betydelser saknas genomgående i nya upplagor.

words and meanings". *LDOCE* (87) redovisar "56,000 words and phrases", den nya upplagan "Over 80,000 words and phrases"; uppgift om antal nya ord eller betydelser saknas. *COBUILD* (87) anger "over 70,000 references", den nya upplagan "over 75,000 references", varav "More than 4,000 new words and meanings". *CIDE*, slutligen, uppger ett omfång av inte mindre än "100,000 words and phrases".⁷

Tendensen mot allt större täckning av ordförrådet, med ett kontinuerligt tillflöde av nya ord och betydelser, är således uppenbar. Det kan i detta sammanhang också vara av intresse att utröna huruvida det finns skillnader i täckning mellan lexikonen beträffande ett antal nyare, numera vanliga ord i engelskan (flera av dem känner vi också igen i dagens svenska). Först några ord som saknas i de tidigare upplagorna men som finns med i samtliga nya upplagor, inklusive *CIDE*: *bimbo*, *e-mail* (med stavningsvarianterna *email*, *E-mail*, *Email*), *multicultural*, *multimedia*, *wannabe*. Så några ord som tas upp av de flesta men inte alla de nya lexikonen. *Internet* och *PC* i betydelsen 'politically correct' är inkluderade i samtliga lexikon utom *OALD*; detsamma gäller substantivet *grunge* (välkänt bl.a. inom rockmusiken). Den i dag högfrekventa förkortningen *IT* redovisas i alla lexikonen utom i *COBUILD*, inte ens under uppslagsordet *information technology*.⁸ Å andra sidan är *COBUILD* det enda lexikonet som har med det inte minst i MTV-sammanhang frekventa *unplugged*, medan *hypertext* förekommer i samtliga lexikon utom *LDOCE*. Ett mer vardagligt IT-ord som *home page* har däremot inte lyckats nästla sig in i något av lexikonen. Låt oss slutligen titta på det i informellt språk mycket vanliga substantivet *nerd*, sedan några år allt vanligare även i svenskan (*nörd*).⁹ Ordet finns upptaget i samtliga nya lexikon utom *OALD* – något förvånande med tanke på att *nerd* är inkluderat i *Oxford Advanced Learner's Encyclopedic Dictionary* från 1992. Ordet saknas i såväl *OALD* (89) som *COBUILD* (87), till skillnad från *LDOCE* (87).

Som framgår av ovanstående exemplifiering företer de fyra lexikonen en ganska splittrad bild av en handfull välkända nyord i engelskan; ingen av ordböckerna redovisar samtliga ord. Detta ger ett kollektivt intryck av osystematisk täckning – olika för varje lexikon – med som det verkar tämligen slumpmässiga bortfall av ord som tveklöst hör hemma i ett inlärningslexikon för det sena nittio-talet. Några exotiska

⁷ Jfr också *CIDE*, s. viii: "*CIDE* provides more entries and more examples than has been possible in other learners' dictionaries."

⁸ Det är i sammanhanget värt att notera att *IT* finns med som uppslagsord redan i *LDOCE* (87), liksom även i *Concise Oxford Dictionary* (1990).

⁹ Ordet saknas dock i *Nationalencyklopedins ordbok* (1996).

specialuttryck från begränsade fackområden är det ju inte fråga om. Det handlar om ord som man stöter på dagligen i vanlig tidningstext.

Korpusbaserings

De ovan påtalade luckorna i ordtäckningen är i sig knappast något att uppröras över; oförklarliga luckor har alltid funnits i ordböcker. Ändå kan man tycka att t.ex. frånvaron av *nerd* i *OALD* eller *IT* i *COBUILD* är förvånande. Anledningen till denna förvåning är att sådana luckor kunnat uppstå trots att de aktuella lexikonerna samtliga är baserade på enorma material av samtida autentisk engelsk text, datoriserade mångmiljonkorpusar där representativitet beträffande olika texttyper i princip är garanterad. Korpusbaserings ligger också, direkt eller indirekt, till grund för många av de trender som är märkbara bland "learners' dictionaries".

Den första engelska inlärningsordbok som konsekvent baserades på en stor datoriserad korpus var *COBUILD* (87).¹⁰ Sedan dess har tendensen varit klar. Inte bara *COBUILD* (95), utan även *OALD* (95), *LDOCE* (95) och *CIDE* är korpusbaserade. Detta framhålls också som ett försäljningsargument, ett slags kvalitetsgaranti för att urvalet av ord som behandlas i lexikonet är representativt och har objektiv täckning i det dagens språk, att upplysningar om betydelser, konstruktionssätt etc. är pålitliga, dvs. väl dokumenterade i autentisk text. Utvecklingen inom korpusområdet har som bekant varit mycket snabb och tendensen har varit entydig de senaste tio åren, mot större och större korpusar. Vad som för något decennium sedan var science fiction är i dag lexikografisk verklighet. Några sifferuppgifter, hämtade från de fyra lexikonerna (förord, inledningar etc.), kan belysa detta förhållande.

Till grund för *COBUILD* (87) låg en korpus på 20 miljoner ord löpande text, medan *COBUILD* (95) "has been written using the evidence of over *two hundred million* words – ten times the corpus made for the original dictionary" (s. viii). *COBUILDs* databas, "The Bank of English", motsvaras för *LDOCE* (95) av en korpus på drygt 130 miljoner ord. Tyndpunkten ligger i "British National Corpus" (100 miljoner ord), som även utgör basen för *OALDs* (95) korpus om c. 140 miljoner ord. *CIDE*, slutligen, bygger på "Cambridge Language Survey", ett material omfattande 100 miljoner ord. Samtliga korpusar innehåller såväl brittisk som amerikansk engelska.

¹⁰ Se *COBUILD* (87), s. xv, och *COBUILD* (95), s. viii. Jfr även *LDOCE* (87), s. F8. Angående relationen mellan lexikon och korpusar, se Calzolari (1996).

Utvecklingen mot användning av gigantiska datoriserade korpusar inom lexikografin är således uppenbar och innebär ett revolutionerande framsteg jämfört med tidigare, mindre systematiska metoder att avtappa autentiskt språkbruk för kodifiering i ordböcker. Trots – eller kanske snarare på grund av – detta inställer sig frågan hur det då kan komma sig att samtliga fyra lexikon, om än i varierande grad, i sin korpusbaserade täckning av dagens ordförråd ändå uppvisar luckor av det slag som exemplifierats ovan.

I förordet till *LDOCE* (95) betonar Randolph Quirk att "readers have to be assured that the words they need to understand and use are included, and that such inclusion reflects up-to-date occurrence in material from a wide range of English-speaking countries and from sources dealing with a wide range of subject matter." Han tillägger: "The advent of computerised corpora enables us to achieve a greatly enhanced coverage." Som redan påpekats är sådana uppfattningar representativa för filosofin bakom samtliga här diskuterade lexikon. En stor datoriserad korpus skall bl.a. borga för att inte några i allmänspråket vanliga och användbara ord skall kunna "slip through the net". Som vi sett tycks det emellertid inte alltid förhålla sig på detta vis. Precis som i tidigare, icke korpusbaserade lexikon finns det svårförklarliga luckor. Vad de beror på i de enskilda fallen är inte lätt att ge något generellt svar på. Att t.ex. hänvisa till en ofrånkomlig eftersläpning vid intäckning av nya ord låter sig knappast göras, eftersom olika lexikon uppvisar olika luckor. En allmän slutsats kan dock dras: jättelika datoriserade korpusar av dagens modell må vara ett aldrig så kraftfullt arbetsredskap för moderna lexikografer, men lika lite som tidigare metoder vid ordboksproduktion tycks de kunna garantera heltäckning inom det för lexikonets syfte relevanta ordförrådet.

I sin "Introduction" (s. 9) till *COBUILD* (95) skriver John Sinclair: "Nowadays the problem is not finding the language..." Det ter sig då än mer gåtfullt hur ett ord som *IT* kan saknas i lexikonet. Att se en lösning i ännu större korpusar verkar inte rimligt. De som nu används förefaller mer än väl ha uppnått kritisk massa; de exemplifierade luckorna måste bero på något annat än otillräcklig korpusstorlek. Att uteslutande förlita sig på en korpus räcker tydligen inte. Det verkar i stället som om lexikografens intuition, baserad på egna språkliga erfarenheter och en upptränad observationsförmåga, måste balansera och i viss mån korrigera en aldrig så representativ korpus även vid själva urvalet av de ord som skall tas upp i lexikonet. Kanske är det i praktiken så att de stora korpusarna haft mer revolutionerande betydelse för andra saker i "learners' dictionaries" (grammatiska mönster, kollokationer, stilnivåer etc.) än för just urvalet av uppslagsord.

Talspråk

Resonemanget i föregående avsnitt är inte avsett som ett argument mot korpusbaserade lexikon, utan som en påminnelse att en aldrig så stor korpus inte automatiskt undanröjer alla problem. Dagens korpusar ställer i själva verket större krav än någonsin på lexikograferna då det gäller att välja och vraka bland all den information som en stor datoriserad korpus ställer till förfogande. Vilka kriterier skall styra urvalet? Vilka typer av information och vilket slags språk skall redovisas i t.ex. en inlärningsordbok?

En klar tendens i dagens "learners' dictionaries" är att det informella talspråket fått större utrymme än tidigare. Delvis beror detta säkert på att den kommunikativt inriktade språkundervisning som anammats i allt fler länder sedan åttiotalet särskilt betonat vikten av att verkligen kunna tala det främmande språket. Det finns således en språkpedagogiskt motiverad lexikonmarknad för engelskt talspråk, även accentuerad av det stora utbud av informell talad engelska som finns tillgängligt genom t.ex. satellit- och kabel-TV. Men det finns också ett annat skäl till att talspråket hamnat i fokus på ett annat sätt än i tidigare inlärningslexikon, nämligen den ökade tillgången på stora talspråkskorpusar.

Samtliga de fyra ordböcker som här diskuteras anger att deras korpusar är uppbyggda av såväl skrivet språk som informellt talspråk. *COBUILD* anger storleken på sin talspråkskorpus, som är en del av "The Bank of English", till "about 15 million words", hämtade från "recordings of everyday casual conversation, meetings, interviews, and discussions" (s. xii). De tre övriga lexikonerna ger ingen uppgift, men för *LDOCE* och *OALD* kan man räkna med en talspråkskorpus om åtminstone 10 miljoner ord, som ingår i "British National Corpus".¹¹ Enligt *LDOCE* (s. xi) handlar det om "natural speech, not radio or TV programmes or language that is in any way scripted."

Bland de fyra lexikonerna är det framför allt *COBUILD* och *LDOCE* som särskilt markerar typiskt talspråkliga ord och uttryck, som en viktig stilistisk/sociolingvistisk upplysning.¹² I båda dessa ordböcker – till

¹¹ Uppgiften är hämtad från *Longman Language Activator* (1993: F8), som i likhet med *LDOCE* och *OALD* är baserad på bl.a. "British National Corpus".

¹² Jfr *LDOCE* (pärmens baksida): "The only dictionary to recognize the importance of spoken English..." – en karakteristik som *COBUILD* säkerligen inte skulle hålla med om. I sitt förord till *LDOCE* skriver Randolph Quirk om

skillnad från de två övriga – anges explicit att t.ex. adverbet *yeah* och frasen *you know* tillhör talspråket. De två övriga lexikonen nöjer sig med att mer traditionellt beteckna dessa uttryck som informella, en stilbeteckning som också används i såväl *LDOCE* som *COBUILD*.

Markören "spoken" i *LDOCE* och *COBUILD* kan i sig hälsas som ett välkommet tillskott till den gängse repertoaren av stil- och registerinformation i inlärningsordböcker, till stor del beroende på tillgången till stora talspråskorpusar.¹³ Dock förefaller användningen av denna markör brista en hel del i konsekvens, inte bara – föga förvånande – mellan de båda lexikonen utan också inom ett och samma lexikon. I synnerhet tycks avgränsningen i förhållande till stilmarkören "informal" ha vållat problem. Ett uttryck som *no way!* räknas i *LDOCE* bland "spoken phrases"; i *COBUILD* betecknas det som "an informal expression". Detsamma gäller t.ex. *no such luck!*. Interjektionen *whoops* anges som "spoken" i *COBUILD*, men har ingen särskild stilmarkör i *LDOCE*. Den lika talspråkliga interjektionen *ouch* får däremot inte någon stilmarkör alls i *COBUILD*. Till skillnad från det jakande *yeah* (se ovan) etiketteras det nekande adverbet *nope* som "informal" i *LDOCE*, som "informal spoken" i *COBUILD*. Exempelen kunde mångfaldigas. I och för sig har kanske inte denna brist på ordning och reda så stor praktisk betydelse – det finns ju ett samband mellan t.ex. informellt språk och talspråk, även om de inte är synonyma begrepp. Men värdet av stil- och registerinformation, särskilt i en inlärningsordbok, minskar naturligtvis om läsaren inte riktigt kan lita på att aktuella markörer används konsekvent. Här finns det således, i båda lexikonen, utrymme för förbättringar.

Kontrastiv information: en ny trend?

Det ligger i enspråkiga lexikons natur att de normalt inte ger kontrastivt differentierad information om ordförrådet, anpassad till vad man erfarenhetsmässigt vet är särskilda svårigheter för olika grupper av inlärare med olika förstaspråk. Exempelvis norska, finska och ryska inlärare har delvis olika problem beträffande engelskans (eller tyskans, franskans etc.) vokabulär, något som språklärare i regel känner väl till. Det är därför intressant att notera att två av de här aktuella lexikonen

talspråskorpusen att den "has been put to prominent use for the first time in this new edition [...]".

¹³ *COBUILD*, till skillnad från *LDOCE*, använder också stilmarkören "written" vid ord som huvudsakligen förekommer i skriftspråket, t.ex. *anew*, där *LDOCE* i stället använder markören "literary".

uttryckligen säger sig ha utnyttjat var sin "learner corpus", dvs. en korpus bestående av "learner English". *LDOCE* anger (pärmens baksida) att man vid sidan av det övriga korpusmaterialet använt sig av "the 5 million word Longman Learner's Corpus". Något ytterligare utöver denna knapphändiga information får vi tyvärr inte reda på, t.ex. i vad mån denna korpus har påverkat lexikonets uppbyggnad.¹⁴ *CIDE* å sin sida anger varken storlek eller några andra konkreta uppgifter om sin korpus. Men man talar om (s. viii) att "learner errors are codified for retrieval and analysis, This allows the learning difficulties of specific language groups to be targeted."

Ett konkret resultat av *CIDEs* intresse för språkspecifika, transferrelaterade problem vid ordinlärning är att lexikonet presenterar ett urval engelska "false friends" – dessa klassiska symptom på lexikal interferens – i förhållande till 16 olika språk.¹⁵ Förutom japanska, koreanska och thai är den stora majoriteten av de kontrasterade språken europeiska. Danska, norska och svenska finns med, däremot inte isländska eller finska. Redovisningen av "false friends" sker på två sätt. Dels ges listor över engelska "false friends" i förhållande till vart och ett av de 16 språken, dels ges vid enskilda uppslagsord, som t.ex. *actual*, information om att ordet är en "falsk vän" i förhållande till bl.a. danska och tyska.

Urvalet "falska vänner" i begränsade listor kan alltid diskuteras. Att kräva fullständighet är omöjligt, bl.a. därför att förrådet inte är konstant. Nya medlemmar tillkommer ständigt, däribland en och annan falsk vän.¹⁶ Vad man däremot kan kräva är att urvalet skall bygga på rimliga principer, så att de ord som tas upp verkligen har karaktären av "falska vänner". I detta avseende uppvisar *CIDE* klara brister. I lexikonet förklaras begreppet "false friend" (s. 502) på följande, tämligen okontroversiella sätt: "a word in a foreign language which looks or sounds similar to a word in your own language but does not have exactly the same meaning." En viktig del i denna definition är naturligtvis den semantiska aspekten ("not...exactly the same meaning"). "Falsk vänskap" kan sägas uppstå genom en konspiration mellan form

¹⁴ Jfr *Longman Language Activator* (1993:F8): "The Longman Learners' Corpus of students' writing has given us insights into the problems shared by students from over 70 different countries [...]". Dock sägs ingenting om eventuella *skillnader* mellan grupper av inlärare med olika förstaspråk; jfr Ohlander 1995). För en mer ingående diskussion, se Granger (1996).

¹⁵ Detta är också en av anledningarna till att lexikonet kallar sig "International"; jfr pärmens baksida, som dock felaktigt uppger antalet olika språk till 14 i stället för 16 (se s. 502).

¹⁶ Jfr S. Ohlander (1991:258f.).

och betydelse i relationen mellan två språk. Enbart ljudlighet eller likartad stavning räcker i regel inte för att kvalificera ett ord, psykologiskt, som "äkta falsk vän". Vanligtvis finns där också en semantisk "frestelse", dvs. något slags betydelselikheter som gör bedrägeriet kontextuellt rimligt; ett minimikrav är samma ordklassstillhörighet.¹⁷ Därför är alla överens om att engelska ord som *actual*, *announce* och *eventual* är "false friends" för inlärare med ett flertal olika förstaspråk, liksom *become* i förhållande till tyska *bekommen*. Dessa ord finns också, tillsammans med en hel del andra klassiker, inkluderade i *CIDEs* listor. Så långt är det lätt att samtycka, även om det också på denna fasta mark finns en del gåtfullheter och inkonsekvenser i form av omotiverade luckor i listorna för vissa språk jämfört med andra. Exempelvis saknas i den svenska listan just det välkända *actual*, medan det redovisas i t.ex. den danska och tyska. Ord som *flesh* och *stool* uppfyller också högt ställda krav på "falsk vänskap" i förhållande till svenskan men finns inte med.

Mer principiellt diskutabelt är att en mängd ord i de olika listorna enbart verkar ha tagits med på grund av ytlig likhet i uttal och/eller skrift, utan någon reell semantisk/kontextuell "frestelse". Några exempel från den svenska listan: *bucket* (jfr sv. *bukett*), *burst* (sv. *borsta*), *gang* (sv. *gång*), *glance* (sv. *glänsa*), *tax* (sv. *tax*; eng. *dachshund*), *under* (sv. *under*; eng. *wonder*). I så fall, varför inte lika gärna t.ex. *barn* (sv. *barn*), *glass* (sv. *glass*), *list* (sv. *list*; eng. *cunning*) *swim* (sv. *svimma*), *utter* (sv. *utter*; eng. *otter*) och en mängd andra liknande fall? Andra listor än den svenska ger rikhaltiga exempel på samma sak, dvs. godtyckligt valda, slumpvisa ytlikheter av den sort som i synnerhet "intermediate and advanced learners" uppfattar som roande snarare än problematiska, sådant som språkstudier säkert i alla tider skämtat om mer än lurats av.

Ett annat problem är att en hel del ord i listorna för olika språk inte finns markerade som "false friends" vid enskilda uppslagsord (fr ovan). Detta gäller bl.a. fall som *announce*, *eventual*, *fabric* och *novel*, vilka finns redovisade i den svenska listan över "false friends" men som inte har någon motsvarande markering vid de enskilda uppslagsorden. Denna brist på intern konsekvens mellan listor och uppslagsord är besvärande.

Huvudintrycket av *CIDEs* presentation av "false friends" är att den skulle må väl av en rejäl uppstädning, både beträffande urval och kon-

¹⁷ "Falsk vän" är inte något väldefinierat begrepp, särskilt inte beträffande "balansen" mellan fonologiska och semantiska likheter. Se Veisbergs (1996); jfr även Hill (1982:iv) och U. Ohlander (1991:68f.).

sekvens. Trots detta ger lexikonet en hel del nyttiga upplysningar inom området. Tanken i sig – att ge språkspecifik information i en enspråkig ordbok – är också intressant. Den kan ses som ett försök till inbrytning på traditionellt tvåspråkiga ordboksdömen. *CIDEs* "false friends" kan också uppfattas som ett lexikografiskt uttryck för det nyväckta intresset för kontrastiv lingvistik under de senaste tio åren, liksom för den därmed sammanhängande, urgamla tanken – numera åter teoretiskt respektabel – att olika förstaspråk på olika sätt påverkar inläringen av ett nytt språk.¹⁸ Möjligen kan vi här se början till en ny trend bland engelska inlärningsordböcker.

Frekvensuppgifter

Att veta om ett ord är vanligt eller ovanligt i ett visst textsammanhang är en viktig del av kunskapen i ett främmande språk. Som regel – om man bortser från språkets allra mest högfrekventa ord – är detta något som kräver lång tid och en myckenhet exponering för det främmande språket, i både tal och skrift. Här finns ytterligare ett område där användningen av stora datoriserade korpusar har möjliggjort mycket mer precisa upplysningar än som tidigare varit möjligt. Särskilt tydligt är detta i de nya upplagorna av *COBUILD* och *LDOCE*. Visserligen anger även *COBUILD* (87:xix) och *LDOCE* (87:F20), liksom *CIDE* (s. ix) och den nya upplagan av *OALD* (s. ix), att frekvens varit en bärande princip för ordningsföljden bland olika betydelser inom samma uppslagsord. Men det är först i och med 1995 års upplagor som *COBUILD* och *LDOCE* gått ett viktigt steg vidare: i båda lexikonerna ges nu även frekvensuppgifter för enskilda uppslagsord, om än på skilda sätt.¹⁹

COBUILDs ordfrekvenser är indelade i fem frekvensband. De olika banden innehåller fler och fler ord alltefter fallande frekvens (s. xiii): det första bandet innehåller c. 700 av språkets vanligaste ord (t.ex. *of*, *go*), det andra c. 1200 ord (t.ex. *bridge*, *obvious*), det tredje c. 1500 ord (t.ex. *gate*, *medicine*), det fjärde c. 3200 ord (t.ex. *artistic*, *puzzle*), det femte c. 8100 ord (t.ex. *denim*, *inconceivable*). Uppslagsord utan frekvensmarkering utgör en sjätte grupp som närmast representerar ett centralt passivt ordförråd, "words which you will probably read or hear rather than words which you will often need to use yourself" (s. xiii).

¹⁸ Se Aijmer *et al.* (1996) och Granger (1996:46); jfr också Swan & Smith (1987).

¹⁹ Jfr *COBUILD* (s. ix): "For the first time in a major dictionary, COBUILD gives information about the frequency of the headwords."

Frekvensuppgifterna i *COBUILD* har tveklöst klar pedagogisk relevans för språkinläraren/lexikonanvändaren, liksom för språkläraren, även om man inte utan vidare kan använda dem för att ställa samman engelskans "centrala ordförråd" i någon absolut mening. Att de också är en intressant innovation i "learners' dictionaries" är uppenbart.

Frekvensredovisningen i *LDOCE* har en delvis annan inriktning, med en klar orientering mot "the 3000 most frequent word in both spoken and written English" (s. xi). *LDOCE* arbetar till skillnad mot *COBUILD* med tre lika stora frekvensband. Varje grupp omfattar tusen ord, med särskilda markeringar för talat respektive skrivet språk. Ett ord som *action* har frekvensmarkörerna "S(poken)1" och "W(ritten)1", vilket innebär att ordet tillhör de tusen vanligaste orden i såväl talspråket som skriftspråket; på samma sätt har verbet *prefer* markörerna "S2" och "W2", dvs. ordet tillhör det andra tusentalet vanligaste ord i tal och skrift. Av särskilt intresse är de många fall där det finns en skillnad mellan tal- och skriftspråksfrekvenser. Substantivet *delay* är t.ex. markerat "S2" och "W3" (tredje vanligaste tusentalet) och är således vanligare i tal än i skrift. För en del uppslagsord finns frekvensmarkör endast för antingen tal eller skrift: *delay* som verb har enbart markören "S3", medan *demand* som verb enbart har "W2", till skillnad mot substantivet *demand*, som i stället har markörerna "S1" och "W1". Det särskilt i talspråket vanliga adverbet *hopefully* har mycket riktigt endast markören "S1".²⁰ Det går med andra ord att utläsa intressanta skillnader mellan talad och skriven engelska i *LDOCE*s frekvensmarkeringar för de tre tusen vanligaste orden.

För somliga uppslagsord i *LDOCE* ges också särskilda grafiska uppställningar över sådana skillnader. Vid verbet *bet* får man t.ex. reda på inte bara hur mycket vanligare detta verb är i talspråket än i skriftspråket, utan också att detta beror på ordets förekomst i en del mycket vanliga talspråksfraser som *I bet* och *You bet!*. Även andra typer av frekvensinformation ges i grafisk form vid vissa ord. Så ges vid verbet *receive* upplysningar om frekvensskillnaderna mellan *receive* och *get* i talspråk och skriftspråk; vid *trousers* finns frekvensinformation om användningen av *trousers* jämfört med *pants* i brittisk och amerikansk engelska, vid verbet *expect* om hur olika grammatiska mönster (*expect to do sth*, *expect sth/sb* etc.) frekvensmässigt skiljer sig åt. Allt detta är av givet intresse när det gäller att mer precist identifiera

²⁰ Däremot ges ej separata frekvensmarkörer för *hopefully* som satsadverb/ attityd-
adverb jämfört med *hopefully* som sättsadverb; säkerligen beror dock den höga tal-
språksfrekvensen nästan uteslutande på den första, av vissa talare ogillade första
typen (*Hopefully we can solve the problem*).

autentiskt språkbruk för pedagogiska ändamål. Det är naturligtvis också, i likhet med *COBUILDs* frekvensangivelser, av lingvistiskt intresse, ett av många exempel på hur korpuslingvistisk metodik kan bidra till ökad kunskap om engelskan och andra språk.

Att vi i framtiden kommer att få se en utökad användning av pedagogiskt relevanta frekvensuppgifter i "learners' dictionaries" är ingen djärv gissning. Frågan är snarast hur mycket som är pedagogiskt lämpligt att pressa in i en inlärningsordbok och på vilka grunder urvalet bör ske.

Definitioner och definitionsordlistor

I en inlärningsordbok är det särskilt viktigt att ordbetydelser definieras på ett sådant sätt och med sådana ord att inte orden i själva definitionen hindrar förståelse av uppslagsordets betydelse. Att i ett sådant lexikon förklara ett ord som *beaver* medelst definitionen "any large amphibious broad-tailed rodent of the genus *Castor*..." (*Concise Oxford Dictionary*) vore inte lämpligt. I stället har man alltid i "learners' dictionaries" strävat efter att använda så enkla definitionsord som möjligt. Den första upplagan av *LDOCE*, från 1978, introducerade en kontrollerad definitionsordlista på c. 2000 ord, varmed i princip betydelsen hos samtliga uppslagsord förklarades.²¹ Även den senaste upplagan av lexikonet använder sig av denna "justly famous *LDOCE* defining vocabulary, now still further refined and improved" (Randolph Quirk, Preface, s. ix). I slutet av lexikonet finns, liksom i de två tidigare upplagorna, en lista över samtliga definitionsord, "carefully chosen to ensure that the definitions are clear and easy to understand, and that the words used in explanations are easier than the words defined" (s. B12).

Bruket av en begränsad uppsättning definitionsord är i och med 1995 års utgåvor ett genomgående drag i engelska inlärningsordböcker. Nykomlingen bland dem, *CIDE*, anger en definitionsordlista på "under 2000 words" (s. 1702), att jämföra med *LDOCEs* "about 2000 common words" (s. xvi). Pionjären *OALD* är å sin sida mån om att framhålla som en viktig nyhet i den senaste upplagan "the introduction of a 3500-word vocabulary within which definitions are written" (s. vi). Som synes är antalet definitionsord i *OALD* betydligt större än i *LDOCE* och *CIDE*.

²¹ Till grund för denna definitionsordlista låg Michael Wests *A General Service List of English Words* (1953); se *LDOCE* (1978:F8). Jfr också Wests *An International Reader's Dictionary* (1965), där ordförklaringarna "are written within a vocabulary of 1,490 words held to be the commonest words in English or the words first learnt by foreigners" (s. vii).

Misstanken kan därmed infinna sig att *OALD*s definitionsvokabulär närmast är att betrakta som en måhända något modifierad inventering av de ord som de facto använts i föregående upplagas definitioner. Mot detta talar att långt ifrån alla definitioner i den senaste upplagan är direkta upprepningar från den närmast föregående. Ett exempel kan illustrera detta. Verbet *allocate* definieras i *OALD* (89) på följande sätt: "allot or assign sth...for a special purpose"; i den nya upplagan däremot: "to distribute sth officially...for a special purpose". Den nya definitionen kräver utan tvivel mindre ordkunskap för att kunna förstås. De mindre frekventa definitionsverben *allot* och *assign* har ersatts av det allmännare *distribute*, vilket till skillnad mot *allot* och *assign* också är inkluderat i *OALD*s definitionsordlista.²²

Redan i sin första upplaga (1987) angav *COBUILD* att man använt sig av en noggrant kontrollerad definitionsvokabulär (omslagets baksida); några ytterligare upplysningar gavs inte. Även den nya upplagan (pärmens baksida) betonar användningen av en "carefully selected defining vocabulary". Till skillnad från de övriga lexikonen ger dock inte *COBUILD* någon separat lista över sina definitionsord utan talar enbart om "a natural defining vocabulary with most words in our definitions being amongst the 2,500 commonest words of English" (s. xviii).²³

Trots skillnader i storlek m.m. mellan de fyra lexikonen är det uppenbart att den traditionella idén om ett begränsat och kontrollerat antal definitionsord inom spannet 2000–3500 ord numera är väl etablerad bland engelska inlärningslexikon. Man skulle kanske kunna tro att en sådan gemensam utgångspunkt skulle göra orddefinitioner tämligen likartade och förutsägbara. Detta är knappast fallet, vilket även en yttlig jämförelse mellan de olika ordböckerna ger vid handen. Som illustration kan vi kort återvända till verbet *allocate* och jämföra de olika definitionerna (för *OALD*, se ovan): "to give (something) to someone as their share of a total amount, for them to use in a particular way" (*CIDE*); "to decide officially that a particular amount of money, time etc or something such as a house or job etc should be used for a particular purpose" (*LDOCE*); "If one item or share of something is **allocated** to a particular person or for a particular purpose, it is given to

²² Det kan noteras att varken *LDOCE* eller *CIDE* använder sig av verbet *distribute* som definitionsord. Enligt *LDOCE*s frekvensangivelser (jfr ovan) tillhör det inte heller de tre tusen vanligaste orden, vilket också stämmer överens med *COBUILD*s frekvensgruppering av ordet, för övrigt inom samma frekvensband som *assign*.

²³ Jfr *COBUILD Student's Dictionary* (1990): "[...] we have avoided the restrictions and awkwardness of a fixed defining vocabulary" (s. v). Detta lexikon innehåller också en lista över de 1860 definitionsord som används tio gånger eller mer.

that person or used for that purpose" (*COBUILD*). Även om det finns gemensamma ord i de olika definitionerna, så är det ändå skillnaderna mellan dem som är mest iögonfallande. Olikheterna beträffande ordval, längd och definitionsstil utgör ett värtaligt bevis på att enigheten bland lexikografer är begränsad i fråga om ett så centralt lexikografiskt område som definitioner. Detta i sin tur beror på svårigheten att på någorlunda fast, vetenskaplig grund etablera en allmänt omfattad och praktiskt användbar uppsättning kriterier på vad som utmärker en god definition t.ex. i en inlärningsordbok. Sådana kriterier måste bygga på omfattande forskning kring hur olika faktorer (ordval, längd, syntaktisk komplexitet etc.) påverkar begripligheten för den avsedda lexikonanvändaren.²⁴

En central fråga i detta sammanhang är om alla uppslagsord i samma lexikon måste följa samma mall, oberoende av exempelvis uppslagsordets frekvens, ordklassstillhörighet, semantiska karaktär (abstrakt - konkret etc.), syntaktiska egenskaper, kollokationer m.m. Är det t.ex. maximalt användarvänligt att som *COBUILD* alltid definiera medelst en fullständig sats, oavsett uppslagsord? Samma fråga kan ställas beträffande det mer traditionella syntaktiska utbytbarhets-kriterium som de övriga lexikonerna ansluter sig till (jfr definitionerna av *allocate* ovan). Är det kanske i stället så att maximal användarvänlighet skulle kunna uppstå vid en mer eklektisk inställning till olika definitionsstilar inom samma lexikon?²⁵ Kan det finnas en konflikt mellan lexikografisk estetik och konsekvens å ena sidan och lexikonets rent praktiska syfte å den andra? Detta är inte rätt plats att mer ingående diskutera dessa frågor. Vi kan nöja oss med att konstatera att *COBUILD* fortsätter på den tidigare inslagna vägen med "full sentence definitions", en innovation som väckte diskussion när den först lanserades men som animerat försvaras av John Sinclair i den nya upplagan (s. xi). Denna definitionsstil är än så länge i minoritet bland engelska inlärlarlexikon. Ett begränsat inflytande från *COBUILD*s definitionsstil kan dock spåras i *CIDE* (jfr även not 25), nämligen i kombination med exempel på vissa bibetydelser, som t.ex. (under *pain*): "If you describe

²⁴ För diskussion kring definitioner och ordförklaringar, se Hanks (1987).

²⁵ Jfr *Longman Language Activator* (1993), där en sådan eklekticism faktiskt tillämpas; se Ohlander (1995). I *LDOCE* är det däremot ont om *COBUILD*-påverkade definitioner; ett av de fåtaliga exemplen är vid uppslagsordet *awry*, där frasen *go awry* definieras så här: "if something goes awry, it does not happen in the way that was planned", att jämföra med *go bananas*: "to become very angry or excited".

a person or a situation as **a pain in the neck** or **a pain**, you think that it or they are very annoying."²⁶

Semantisk differentiering: homonymi, polysemi m.m.

Ett välkänt problem för såväl lexikografer som lexikonanvändare utgör graford som karakteriseras av polysemi och/eller homonymi. Sådana ord tillhör i regel också den högfrekventa delen av ordförrådet, t.ex. starkt polysema verb som *run* och *take*, substantiv som *class* och *form* eller adjektiv som *dry* och *wrong*. Att sådana ord i ett språk som engelska dessutom ofta kan tillhöra mer än en ordklass gör inte saken lättare. Ofta har sådana ord presenterats som flera separata huvudord i stället för under ett enda huvudord. Variationen mellan lexikon har dock varit stor. Någon gemensam strategi inom detta notoriskt problematiska område – semantiskt/lexikologiskt och lexikografiskt – har inte funnits.

Den första upplagan av *COBUILD* introducerade ett för inlärningsordböcker radikalt grepp: såväl polysemi som homonymi sorterades in under ett och samma huvudord, även oavsett ordklassstillhörighet. Alla betydelser och användningar av *run* presenterades under samma uppslagsord med 67 underavdelningar (plus underavdelningar till många av dessa), med frekvens som en av ordningsprinciperna. Detsamma gällde klara fall av homonymi, som *bank* och *can*. I den nya upplagan av lexikonet medger John Sinclair (s. x) att denna "strict policy of 'one word, one entry' [...] occasionally irritated users." För att motverka detta har den nya upplagan infört s.k. "superheadwords", vilka delar upp vissa tidigare mycket långa "entries" i större underavdelningar. Exempelvis har *mean* i den nya upplagan tre ordklassbaserade "superheadwords" ("verb uses", "adjective uses" och "noun use"). Även mer renodlat semantiska grupperingar förekommer: *lie* är indelat i "position or situation" och "things that are not true", *fancy* i "wanting, liking, or thinking" och "elaborate or expensive"; *bank* och *can* har på liknande sätt numera vardera två "superheadwords". Införandet av "superheadwords" gör det onekligen lättare att hitta i *COBUILD* i de fall där de används. De kunde dock ha använts oftare. För exempelvis *run* är situationen i stort sett oförändrad, dvs. utan "superheadwords". Även många klara homonymifall står kvar under

²⁶ Jfr *COBUILD*: "If you think that a person, job, or situation is very annoying or irritating, you can say that they are **a pain** or **a pain in the neck**" (*pain*, 4).

samma huvudord: för *love* finns "tennisbetydelsen" ("fifteen–love") insprängd bland helt andra betydelser.

Även om förändringarna i *COBUILD* således inte är direkt överväldigande innebär "superheadwords" ändå en modifiering i riktning mot ett mer traditionellt, i det här fallet tydligare presentations sätt. Ökad användarvänlighet är också tanken bakom *LDOCEs* och *CIDEs* bruk av ett slags semantiska grovmarkörer för att göra det lättare att hitta den önskade betydelsen hos ett ord. Under verbet *love* anger *LDOCE* de tre första betydelserna med markörerna ROMANTIC ATTRACTION, CARE ABOUT och LIKE/ENJOY i fetstil före respektive definition, så att läsaren direkt skall ledas till rätt betydelse av uppslagsordet. Dessa "signposts" är avsedda som hjälp vid "entries with many definitions" (s. xvi). Detta är dock ej konsekvent genomfört. En sak är att de inte förekommer vid grammatiska småord som *but*, *of* eller *there*. Men de saknas också vid många lexikala uppslagsord med flera olika definitioner, t.ex. verbet *like* och substantivet *accident*. Logiken här är inte helt lätt att genomsåda. Annars fungerar dessa "signposts" väl som praktiska visuella hjälpmedel för lexikonanvändaren. Detsamma gäller även *LDOCEs* användning av introducerande, översiktliga "menus" över de huvudsakliga betydelsegrupperna vid vissa uppslagsord med ett stort antal betydelser och användningar, som *business*, *hot* och *run*.

I *CIDE* har den semantiska differentieringen i många fall gått ännu längre: "When a word has more than one meaning, [...] guide words help you to find the right one quickly" (s. x). Syftet med dessa "guide words" är ungefär detsamma som med *LDOCEs* "signposts". Det finns emellertid viktiga skillnader. I *LDOCE* används "signposts" för att tydliggöra olika betydelser inom *samma* huvudord med flera olika definitioner (jfr ovan). *CIDEs* "guide words" används i stället för att markera *olika* huvudord. Adjektivet *green* har i *LDOCE* bl.a. följande "signposts" under *samma* huvudord: COLOUR, WITHOUT EXPERIENCE, POLITICS; i princip samma semantiska begrepp (plus ytterligare några) används som "guide words" i *CIDE*, som visuella markörer för *olika* huvudord.²⁷ Någon systematisk parallelism av detta slag finns emellertid inte mellan de båda lexikonerna. Ett huvudord i *CIDE* kan innehålla ett antal betydelsenysanser (utan "guide words") som i *LDOCE* är tydligt markerade med var sin "signpost". Så är fallet

²⁷ Ett kuriosum i ett lingvistikhistoriskt perspektiv: ett av exemplen under *green* POLITICAL i *CIDE* är *green ideas*, selektionsregler är inte eviga utan är beroende av ords betydelseutveckling. Jfr dock Pinker (1994:93), där det i en diskussion kring Chomskys berömda exempel sägs att "the probability that *green* is followed by *ideas* [...] is surely zero."

med t.ex. verbet *kill*. Det motsatta förhållandet kan också gälla. I *LDOCE* är exempelvis samtliga betydelser av substantivet *depression* samlade under samma huvudord, men utan "signposts". I *CIDE* är situationen mer komplicerad, även för lexikonanvändaren. Det finns för det första två separata uppslagsord, markerade med WEATHER respektive HOLE. Vid dessa huvudord ges hänvisningar till två andra huvudord, nämligen verben *depress* CAUSE SADNESS och *depress* REDUCE. Under dessa huvudord återfinns sedan den psykologiska respektive den ekonomiska betydelsen hos *depression*. Den semantiska tanken bakom detta arrangemang – *depression* utspritt över fyra olika uppslagsord – må vara begriplig; praktiskt är det knappast. En annan olikhet mellan lexikonerna är att *CIDE* till skillnad från *LDOCE* (jfr ovan) även försöker särskilja betydelserna hos grammatiska funktionsord med hjälp av "guide words". I ett fall som *but*, där huvudbetydelserna är någorlunda klart åtskilda går det relativt bra: DIFFERENCE, EXCEPT, ONLY. I andra fall uppstår oundvikligen problem, t.ex. vid *on*, vars ofta överlappande betydelser och funktioner redovisas under 31 olika huvudord, markerade med "guide words" av de mest skiftande slag (ABOVE, CONNECTED, WRITING, PAIN, TO etc.). Uppgiften är svår, försöket heroiskt, resultatet inte helt övertygande.

Som framgått av ovanstående presentation finns det trots viktiga skillnader även grundläggande likheter mellan *LDOCE* och *CIDE*. Den viktigaste är kanske att båda lexikonerna, i olika grad och på olika sätt, i praktiken avskaffat skillnaden mellan homonymi och polysemi, denna klassiska lexikografiska knäckfråga. I den föregående upplagan av *LDOCE* markeras fall av klar homonymi, som t.ex. vid substantivet *bank*, genom olika uppslagsord, där vart och ett i sin tur kan inrymma polysemi. Den nya upplagan behandlar i stället de olika betydelserna av substantivet *bank* under samma huvudord, utan någon markerad skillnad mellan homonymi och polysemi.²⁸ I *CIDE* markeras visserligen traditionella homonymer i regel med var sitt huvudord, som i fallet *bank*, men detta gäller även många fall av vad som traditionellt skulle betraktas som klar polysemi (jfr exemplet ovan). *COBUILD*, å sin sida, gjorde ju redan i sin första upplaga rent hus med den intuitivt centrala men semantiskt och lexikografiskt ofta svårhanterade distinktionen mellan homonymi och polysemi. Införandet av "superheadwords" kan visserligen ses som ett slags reträtt men är i huvudsak en praktisk

²⁸ Däremot skiljer man fortfarande, genom olika huvudord, mellan *bank* som substantiv och *bank* som verb; jfr *COBUILD*, som visserligen genom olika "superheadwords" skiljer mellan "finance and storage" och "areas and masses" men som under den förra huvudbetydelsen tar upp *bank* som både substantiv och verb.

nödfallsåtgärd mot ett begränsat antal huvudord med särskilt många och svåröverskådliga underavdelningar. Endast *OALD* håller som en sista ropande röst fast vid tanken att trots alla problematiska fall försöka lexikografiskt markera skillnaden mellan homonymi och polysemi. Det skall bli intressant att se om kommande upplagor av lexikonet kan stå emot grupptricket.

Sammanfattningsvis kan man säga att engelska inlärningsordböcker tycks befinna sig i en process av fortlöpande desarmering av den besvärliga homonymi/polysemi-distinktionen, en utplattning av de semantiska hierarkierna om man så vill. Denna strategi, som kan ta sig olika uttryck, är bekväm för lexikografen. Huruvida den också är det för lexikonanvändaren är inte lika självklart. Skillnaden mellan homonymi och polysemi kan, trots mängder av gränsfall, betraktas som en psykologisk realitet. Att arrangera ordbetydelser på ett sådant sätt att gränsen mellan (synkront) besläktade och helt obesläktade betydelser blir oklar, eller helt suddas ut, kan knappast vara befrämjande för ordinläring, där sammanhang mellan ord och betydelser brukar framhävas som något väsentligt. Någon helt idealisk lösning på dilemmat – om det nu finns någon – är ännu inte i sikte. Problemet har alla möjligheter att förbli klassiskt.

Avslutande reflektioner

På senare år har "learners' dictionaries" kommit att spänna över allt vidare fält, i takt med delvis nya perspektiv och betoningar inom såväl lingvistik som språkpedagogik. Sedan länge har inlärningsordböcker innehållit mängder av rent grammatisk information. Vi har under något tiotal år också kunnat bevittna en ökande fokusering på socio-lingvistiska, pragmatiska och textlingvistiska aspekter av språket; det samma gäller kollokationer. Hand i hand med denna kvalitativa utveckling går en kvantitativ: fler och fler uppslagsord, tjockare och tyngre lexikon. Konkurrenssituationen driver också fram något av en hetsjakt på nya "special features", innehållsligt såväl som i fråga om presentation, allt i syfte att förstärka den egna profilen på marknaden.

Jag har i denna artikel velat peka på ett antal någorlunda tydliga tendenser och därmed sammanhängande frågor i engelska "learners' dictionaries"; fler skulle ha kunnat nämnas. Som framgått finns det på viktiga punkter en samstämmighet mellan de olika lexikonerna, även om skillnaderna mellan dem i många fall är lika iögonenfallande. Det är på sätt och vis fascinerande att det trots den kollektiva rörelsen, samsynen i stort om vad en inlärningsordbok bör innehålla, ändå är så lite som är självklart när det gäller realiseringen av målen, alltifrån vilka ord som

skall tas med till lexikonartiklarnas utformning i stort och smått. Frågorna är gemensamma, svaren ofta olika. Något säger detta onekligen om lexikografins och lexikografens situation i en ocean av valmöjligheter, en ocean som inte blivit mindre genom de datoriserade korpusarnas intåg.

Det råder ingen tvekan om att engelska inlärningsordböcker blivit allt bättre med åren. Det är lätt att hålla med Randolph Quirk i hans förord till *LDOCE* (s. 10): "I would have liked to have this dictionary when I was a student." Detta skulle kunna gälla samtliga de här diskuterade lexikonen, trots att inget av dem är invändningsfritt, ens på ett generellt plan.

Mot slutet av sin inledning till *COBUILD* (s. xi) säger John Sinclair: "Users expect more and more from their dictionaries [...]". Hur sant detta verkligen är kan man möjligen tvista om. Oomtvistligt är däremot att i alla fall engelska inlärningslexikon tycks drivas framåt av ambitionen att tillgodose vad man uppfattar som ständigt ökande användarbehov. Mycket gott kommer helt visst av detta. Men det kan också finnas en risk att ansamlingen av mer och mer – även mer och mer olika – information inom ett och samma, allt större lexikon leder till "information overload" för användaren. Om ett lexikon skall innehålla nästan allt riskerar lättillgängligheten och den pedagogiska tydligheten att gå förlorade. Det är inte säkert att all den korpusbaserade information som i princip är möjlig att inkludera också är pedagogiskt motiverad att ha med. Ett lexikon kan ändå aldrig bli en ersättning för egna språkliga erfarenheter, där inte minst textläsning – orden i verkligt autentisk kontext, snarare än i autentiska exempel – spelar en central roll vid uppbyggnaden av ett fungerande ordförråd.

Detta leder vidare till frågan om framtidens "learners' dictionaries". De fyra här diskuterade lexikonen utgör inte någon slutpunkt i utvecklingen; många problem återstår att lösa. Arbetet fortsätter – men i vilken riktning? Kanske börjar det bli dags att ställa frågan om vi i själva verket håller på att närma oss gränsen för den konventionella formen för inlärningslexikon, dvs. en bok. Även om det förmodligen alltid kommer att finnas behov av lexikon i bokform kommer kanske inlärningsordböckernas framtida utveckling huvudsakligen att äga rum inom andra medier och i andra format. Det handlar då inte främst om CD-ROM-versioner av redan existerande lexikon, snarare om helt andra sätt att utnyttja den nya teknikens potential, inte minst möjligheten för användaren att själv, efter egna behov, "skräddarsy" sitt

eget elektroniska lexikon med olika databaser som grund. Men vi är inte där än.²⁹

För en mer överskådlig framtid kan vi nog trots allt vänta oss en fortsatt intensiv utveckling inom ramen för mer konventionella inlärningsordböcker, där några av de här diskuterade trenderna och frågorna säkerligen kommer att spela en central roll. Ett nytt millennium står för dörren: hur kommer 2000-talets upplagor av "The Big Four" att se ut? Startskottet har gått; jakten har börjat.

Litteratur

- Aijmer, Karin, Altenberg, Bengt & Johansson, Mats (eds) 1996: *Languages in Contrast*. Lund University Press.
- Atkins, B. T. S. 1996: Bilingual dictionaries: past, present and future. I: Gellerstam *et al.* 1996, 515–546.
- Calzolari, Nicoletta 1996: Lexicon and corpus: a multi-faceted interaction. I: Gellerstam *et al.* 1996, 3–16.
- COBUILD 1987 = *Collins COBUILD English Language Dictionary*. First ed. Editor in Chief: John Sinclair. London/Glasgow: Collins.
- Collins COBUILD Student's Dictionary* 1990. Editor in Chief: John Sinclair. London/Glasgow: Collins.
- Concise Oxford Dictionary* 1990. Eighth ed. Oxford: The Clarendon Press.
- Gellerstam, M., Järborg, J., Malmgren, S.-G., Norén, K., Rogström, L. & Røjder Papmehl, C. (eds) 1996: *Euralex '96. Proceedings I–II*. Göteborg University, Department of Swedish.
- Granger, Sylviane 1996: From CA to CIA and back: an integrated approach to computerized bilingual and learner corpora. I: Aijmer *et al.* 1996, 37–51.
- Hanks, Patrick 1987: Definitions and explanations. I: Sinclair, John (ed), *Looking Up. An Account of the COBUILD Project in lexical computing*. Collins 1987, 116–136.
- Hill, Robert J. 1982: *A Dictionary of False Friends*. London: Macmillan.
- Kernerman, Lionel 1996: English learners' dictionaries: How much do we know about their use? I: Gellerstam *et al.* 1996, 405–414.
- LDOCE 1978 = *Longman Dictionary of Contemporary English*. First ed. Editor-in-Chief: Paul Procter. Longman.

²⁹ För en ingående diskussion om framtidens elektroniska lexikon, se Atkins (1996).

- LDOCE* 1987 = *Longman Dictionary of Contemporary English*. Second ed. Editorial Director: Della Summers. Longman.
- Longman Language Activator* 1993. Editorial Director: Della Summers. Longman.
- OALD* 1989 = *Oxford Advanced Learner's Dictionary of Current English*. Fourth ed. Chief Editor: A.P. Cowie. Oxford University Press.
- Ohlander, Sölve 1991: Efterskrift till U. Ohlander 1991, 242–295.
- Ohlander, Sölve 1995: Le mot juste? Några tankar kring "världens första produktionslexikon". I: *LexicoNordica* 2, 237–259.
- Ohlander, Urban 1991: *Språk som bro och barriär*. Stockholm: Carlssons.
- Oxford Advanced Learner's Encyclopedic Dictionary* 1992. Oxford University Press.
- Nationalencyklopedins ordbok*. Andra bandet. 1996. Vetenskaplig ledare: Sture Allén. Bra Böcker.
- Pinker, Steven 1994: *The Language Instinct*. Allen Lane, The Penguin Press.
- Swan, Michael & Smith, Bernard (eds) 1987: *Learner English. A teacher's guide to interference and other problems*. Cambridge University Press.
- Veisbergs, Andrejs 1996: False friends dictionaries: a tool for translators or learners or both. I: Gellerstam *et al.* 1996, 627–634.
- West, Michael 1953: *A General Service List of English words*. Longmans.
- West, Michael 1965: *An International Reader's Dictionary*. Longmans.

Sölve Ohlander

Ingvar Stenström: *Interlingua-svensk ordbok. 25.000 internationella ord.* 1995. Studentlitteratur.

Det mänskliga språket har två allvarliga fel. Det ena är uppenbart, inte bara för lingvister: det mänskliga språket är inte *ett* språk, utan många. Att folk i alla tider reflekterat över detta hinder för sant allmänmänsklig kommunikation visar bl.a. Bibelns välkända berättelse om Babels torn. Det mänskliga språkets andra stora fel är inte lika iögonenfallande. Det hindrar i regel inte heller normal kommunikation. För filosofer och lingvister har det dock länge stått klart att varje mänskligt språk tycks lida av en besvärande brist på ordning och reda i förhållandet mellan form och betydelse. I sin berömda *Essay Towards a Real Character and a Philosophical Language* från 1668 berör den engelske biskopen och vetenskaplige mångsysslaren John Wilkins frånan av ett mer genomskinligt ett-till-ett-förhållande mellan uttryck och innehåll i mänskliga språk, där han i flertydighet och synonymi ser betänkliga brister ("Defects"):

1. In regard of *Equivocals*, which are of several significations, and therefore must needs render speech doubtful and obscure; [...]
2. In respect of *Synonymous* words, which make language tedious, and are generally *superfluities* [...] And yet there is no particular language but what is very obnoxious in this kind. (s. 17f.)

I själva verket kan även den babelska "språkföbistringen" betraktas som ett storskaligt symptom på den av Wilkins påtalade bristen på en ett-till-ett-relation mellan form och betydelse. I det ena språket uttrycks en viss betydelse genom en viss form, i det andra språket på ett helt annat sätt; omvänt kan samma form ha helt olika betydelse i olika språk. Om man som tankeexperiment, i en utomjordings globalt chomskyanska perspektiv, betraktar alla mänskliga språk som ett enda framträder således synonymi och flertydighet som gåtfulla – och onödiga.

Varje vän av språklig ordning och reda inser lätt att den uppkomna situationen är långt ifrån idealisk. Här finns med andra ord ett betydande utrymme för förbättringar via aktiv språkplanering. Historiskt sett har heller inte förslag till reformer saknats, vare sig i det inomspråkliga eller det mellanspråkliga perspektivet. I båda fallen har det

handlat om att på olika sätt, och med olika syften, söka skapa universella språk.

Detta är inte platsen att närmare gå in på de språkfilosofiskt inspirerade försök att konstruera konstgjorda, logiskt tillfredsställande universalspråk som hade sin glansperiod under 1600-talet, med rötter hos Francis Bacon.¹ Förutom den ovannämnde Bishop Wilkins – vars grandioöst upplagda projekt, kopplat till sin tids vetenskap och utförligt redovisat i hans *Essay*, brukar hänföras till avdelningen heroiska misslyckanden – fanns det många andra vid denna tid (Comenius, Descartes, Leibniz m.fl.) som såg skapandet av ett artificiellt universalspråk som ett angeläget, ej blott utopiskt projekt. Allmänt kan sägas att 1600-talets språkuniversalister inte enbart, eller kanske ens primärt, hade i åtanke att förbättra möjligheterna till internationell kommunikation. De syftade främst till att genom ett tydligare, mer logiskt språksystem tillhandahålla ett bättre instrument för tänkande, vetenskap och kunskapsspridning.²

Den bok som givit upphov till ovanstående reflektioner kring en speciell typ av språkplanering, *Interlingua-svensk ordbok (ISO)*, hör hemma i en annan, mer utåtriktad tradition av språkuniversalism. Dess mest välkända representant är esperanto. Med början för drygt hundra år sedan har den också givit upphov till bl.a. volapük, ido, occidental och novial, skapat av Otto Jespersen. De olika konstgjorda språken inom denna tradition har, tydligare än 1600-talets ansatser, ett direkt praktiskt syfte, dvs. att i stor skala fungera som medel för internationell kommunikation. Att framgången blivit begränsad – även för det i särklass mest framgångsrika språket, esperanto – har flera förklaringar. En mycket viktig sådan är naturligtvis engelskans under 1900-talet alltmer dominerande ställning som internationellt språk. När volapük – ordet betyder "världsspråk" – introducerades år 1880 kunde inget naturligt språk med samma självklarhet som dagens engelska göra anspråk på en liknande position. Engelskans segertåg har för övrigt också utgjort en effektiv spärr mot C.K. Ogdens på trettiotalet lanserade

¹ En bred diskussion kring 1600-talets intresse för olika "*a priori* philosophical languages", som en variant av "the search for the perfect language", ges i Eco (1995), kap. 10–14. I ett annat perspektiv resonerar Chomsky (1993:27f.) kring Freges idé om "a logically perfect language" som ett nödvändigt redskap för "a common science".

² Se Slaughter (1982:1f.); jfr Crystal (1987:352f.)

Basic English, ett internationellt hjälpspråk baserat på 850 riktiga engelska ord samt en uppsättning autentiska grammatikregler.³

ISO bygger på tidigare utgivna "interlinguistiska" ordböcker (interlingua - engelska, interlingua - holländska, interlingua - danska). Interlingua introducerades som internationellt hjälpspråk år 1951. Det var resultatet av ett projekt som sedan 1924 bedrivits inom International Auxiliary Language Association, med bl.a. Otto Jespersen och Edward Sapir som lingvistiska konsulter. I inledningen till boken (s. 6) kan man även inhämta att det nya hjälpspråket "redan efter ett par år togs [...] i praktiskt bruk som resumé-språk vid en hjärtläkarkongress, där man kunde konstatera att interlingua förstods av deltagarna utan att de ens hört talas om språket dessförinnan." Indirekt ger detta också en antydning om interlinguas karaktär, särskilt vad gäller dess ordförråd.

I likhet med andra hjälpspråk bygger interlingua sin vokabulär på i huvudsak västeuropeisk grund. Orden i interlingua och därmed i *ISO* återfinns således i "minst tre av språken engelska, franska, italienska, spanska/portugisiska, tyska och ryska" (s. 5); interlingua "förstås praktiskt taget utan förstudier av de 500 miljoner italiensk-, spansk- och portugisisktalande, och därtill av miljoner fransmän och engelsk- och latinkunniga" (s. 6). Detta innebär självfallet en kraftig övervikt åt det latinbaserade ordförrådet (i en del fall har interlingua-orden hämtats direkt från latinet), dvs. själva kärnan i det som med en eurocentrisk beteckning brukar kallas det internationella ordförrådet. Sedan länge har detta ordförråd en grundmurat stark ställning inom särskilt vetenskapens värld, så även vid hjärtläkarkongresser.

ISO har följdriktigt undertiteln "25 000 internationella ord". På pärmens baksida kan man bl.a. läsa följande: "Interlingua [...] är inget konstgjort språk, som diverse s.k. världsspråk, utan en språkvetenskapligt betryggande registrering av det internationella ordförrådet som det de facto existerar." Denna karakteristik är inte invändningsfri. Interlingua är lika lite som andra hjälpspråk ett i vedertagen mening naturligt språk. Det är precis som t.ex. esperanto ett på artificiell väg tillkommet kommunikationsmedel, låt vara med ett ordförråd som, i likhet med många av interlinguas konkurrenter, inte är skapat *ex nihilo* utan på en för många västerlänningar igenkännlig, flerspråkig grund. Graden av konstgjordhet kan man naturligtvis diskutera. Att interlingua för västerländska öron har en mer hemtam klang än de flesta andra hjälpspråk är däremot uppenbart. Detta framgår av t.ex. följande stycke

³ BASIC står för **British American Scientific International Commercial**. Se Ogden (1968), där även Churchills och Roosevelts intresse för Basic English finns dokumenterat (s. 111 f.).

om interlingua som ett slags "mönsterspråk", hämtat från Crystal (1986:355):

Tote le membros del communitate de linguas occidental son in un certe senso dialectos individual que devia plus o minus de un *patrono commun*. [...] Le termine *interlingua* es solmente un synonymo plus technic de *lingua de patrono*.

Den romanska prägeln hos vokabulären i ovanstående rader är omisskännlig, även om det inte alltid är fråga om helt autentiska ordformer. Grammatiskt har interlingua, liksom andra hjälpspråk, sitt eget förenklade form- och regelsystem, även här med inslag från olika grundspråk (ISO, s. 9–12). Exempelvis har såväl obestämd som bestämd artikel alltid samma form, oavsett substantivets genus (i ursprungsspråket) eller numerus: *un femina, le femina, le feminas* (jfr även *le membros* ovan). Adjektiven är oböjda, som i engelskan. Verben bryter däremot genom sina tre konjugationer den grundläggande enkelhetsprincipen: *parlar* ("a-verb"), *vider* ("e-verb"), *audir* ("i-verb"); här kan man i stället tala om ett slags "igenkännlighetsprincip" – det latinsk/romanska inflytandet är uppenbart. Detta gäller också i hög grad pronomen och pronominaladverb (*tu, vos; qui, que; quando, ubi*), liksom räkneorden (*un, duo, tres, quatro, ..., vinti, trenta, ..., cento, mille*).

Även om det latinbaserade ordförrådet har en starkt dominerande ställning i interlingua, så är det inte allenarådande. Många andra språk finns också representerade bland bidragsgivarna till dagens internationella ordförråd. Självklart spelar den klassiska grekiskan en viktig roll: *democratia, helicoptero, phenomen, pneumonia, psychopatic, syntaxe, technologia*. Vår egen tids stora långgivare, engelskan, står självfallet också för många ord: *baby, football, meeting, sexy, week-end, whisky, yacht*. Även tyskan är representerad, om än med betydligt färre ord än engelskan: *hinterland, kaiser, putsch, walzer*. Några svenska bidrag kan också noteras, t.ex. *ombudsman* och *knäckebröd* – däremot inte, något oväntat, *smorgasbord* (eller *gravad lax*). Bland slaviska språk förefaller ryskan inta en obestridlig tätposition, med exempelvis ord som *balalaika, kolkhoz, kopeka* och *ukaz* – däremot inte *gulag* eller *glasnost*. Från sanskrit kommer bl.a. *karma* och *yoga*, från persiska *caravana*.

Som redan betonats har ordförrådet i interlingua en kraftig slagsida åt det europeiska – särskilt västeuropeiska – hållet; indoeuropeiska ord är i förkrossande majoritet. Men där finns också en hel del ord av icke-indoeuropeiskt ursprung, dvs. ord som blivit inlånade i många europeiska språk och därmed vunnit status av "internationella ord".

Finskans *sauna* hör hit. Går vi utanför det europeiska språkområdet är det inte svårt att finna exempel på interlingua-ord med mer "exotiskt" ursprung. Några exempel bland många: turkiska (*kiosque*, *yogurt* – dock ej *kebab*), arabiska (*harem*, *moslem*, *muezzin*), hebreiska (*kasher*, *kibbutz*), japanska (*haiku*, *jujutsu*, *kimono*, *shogun*), javanesiska (*junca* 'djonk'), eskimåiska (*anorak*, *kayak* – men inte *igloo*), nord- och sydamerikanska indianspråk (*wigwam*, *ananas*, *jacaranda*), australiska språk (*kanguru*, *koala*), tibetanska (*yak*).

Etymologisk information ges inte generellt i *ISO*, så t.ex. inte för det överväldigande flertalet ord av grekiskt, latinskt/romanskt eller germanskt ursprung. Detta må vara förklarligt i en ordbok av begränsad omfattning. Lexikonet lider dock av en besvärande brist på konsekvens när det gäller angivelser av etymologisk hemvist för många andra ord. Bland de uppslagsord som nämnts i föregående stycken ges upplysningar om ursprungsspråk för t.ex. *kolkhoz*, *kiosque* och *kibbutz*, däremot inte för *caravana*, *yogurt* eller *ananas*; *jujutsu* anges som ursprungligen japanskt, dock ej det lika japanska *judo*. Denna principlöshet beträffande ord av ett för västerlänningar mer exotiskt ursprung är svår att förstå, särskilt i ett lexikon inriktat på att sprida kunskap om den internationella lånordsvokabulären i stort. Frågor kring ords ursprung är också något som många människor hyser ett spontant intresse för. Här finns således gott om luckor att täppa till.⁴

Ett annat, allvarigare problem med *ISO* berör dess – och därmed interlinguas – avsedda funktion. Å ena sidan kan man använda lexikonet som en inventering och presentation av det internationella ordförrådet, "de främmande orden", i svenskan liksom i många andra språk. Detta är också lexikonets främsta syfte (s. 5). Denna funktion gagnas emellertid knappast av att många av uppslagsorden förekommer i en icke-autentisk form, ej heller av de inslag av rent latin som förekommer. Å andra sidan är interlingua avsett som ett internationellt språk, med ett eget ordförråd och med sina egna grammatiska principer. I detta perspektiv inställer sig frågan om ordförrådets sammansättning, såsom det kan utläsas ur *ISO*.

Ett ordförråd på 25.000 ord kan självfallet inte göra anspråk på att vara heltäckande. Ändå är det rimligt att vänta sig något mer än ett allmänt krav på representation i minst tre språk (jfr ovan) för att ett ord skall ges status av uppslagsord. Utan en mer övergripande urvalsprincip som ledstjärna mot ett någorlunda "centralt" ordförråd riskerar annars

⁴ Jfr t.ex. *Prismas främmande ord* (1984), som ger etymologisk information, liksom *Nationalencyklopedins ordbok* (1995–96). I engelska "desk dictionaries" har etymologisk information sedan länge varit regel.

ett lexikon som *ISO* att i alltför hög grad te sig idiosynkratiskt eller rentav excentriskt i fråga om val av uppslagsord. Att utgivaren av *ISO* inte är omedveten om denna risk antyds i inledningen till ordboken (s. 7): "Jag ber den förste läsare som får nytta av det ursprungligen grekiska ordet rhabdomante 'slagruteman' att sända mig en hälsning..." Inställningen kan betraktas som en närmast hädisk kontrast till moderna lexikografers användning av stora, representativa, datoriserade korpusar som grund för sitt arbete, där bl.a. de inkluderade ordens frekvens är av central betydelse för valet av uppslagsord.

I brist på en konsekvent urvalsstrategi förefaller uppslagsorden i *ISO* ofta ha inkluderats mer av en slump (ordet X råkar finnas), snarare än som resultat av ett rationellt val (ordet X finns och *behövs*). Med andra ord: inte bara en slagruteman kan dyka upp, utan även en hustrumördare (*uxoricida*), måhända beväpnad med en *yatagan* ('avrättningssvärd i arabländerna'). Man skall således inte bland *ISO*s 25 000 internationella ord vänta sig ett "centralt ordförråd" i mer vedertagen bemärkelse. Däremot erbjuder lexikonet en grunduppsättning "lärd" ord inom såväl det vetenskapliga som det allmännare ordförrådet; som redan noterats ges också en hel del lågfrekventa, kulturspecifika ord. Men även vissa mer jordnära ord finns med, t.ex. en talaktspartikel som *ya* 'ju' eller det som "vulgärt" betecknade verbet *pissar*, ord som också klart indikerar *ISO*s och interlinguas ambition att kunna användas i vardagligt talspråk. Ungdoms- och populärkulturens ordförråd finns också representerat, ofta med lån från engelskan: *jeans*, *walkman*, *pop* – däremot inte något ord för rockmusik, förvånansvärt nog, trots att *jazz* är med. Även EU får sin del av uppmärksamheten, bl.a. i form av uppslagsordet/förkortningen *Union Europee/UE*. För FN, *le Naciones Unite*, ges ingen motsvarande förkortning; förkortningen *UNESCO* är däremot ett eget uppslagsord.

Det ibland godtyckliga urvalet av uppslagsord i *ISO* kan illustreras inte endast med frånvaron av väntade ord. Vissa ämnesområden, av relativt perifert intresse inom ramen för 25 000 ord, synes ha fått ett oproportionerligt stort utrymme. Detta gäller t.ex. konstgjorda språk, i *ISO* representerade inte bara av uppslagsordet *esperanto*, utan också av de betydligt mindre välkända *novial*, *occidental* och *volapük* (liksom *interlingua* självt). Följdriktigt, som ett ytterligare indicium på detta möjligen överrepresenterade särintresse, återfinns även egennamnet *Babel*, uppenbarligen främst motiverat genom det i interlinguasammanhang centrala *le turre de Babel*.

Som framgått av resonemanget ovan finns det betydande inkonsekvenser och obalanser i det internationella ordförråd som *ISO* redovisar. Som en väg, om än något vindlande, till det internationella

ordförrådet kan lexikonet ändå fylla sitt syfte, i konkurrens med andra sätt att bekanta sig med de internationella orden i svenskan. För lexikonbrukare med goda kunskaper i våra vanligaste skolspråk kan *ISO* också fungera som en nyttig och nöjsam påminnelse om hur mycket dessa språk trots allt har gemensamt på vokabulärsidan, ytterst en påminnelse om den kreativa potentialen i döda språk som latin och klassisk grekiska.

ISOs ambition att via interlingua ge ökad medvetenhet om de internationella orden är naturligtvis värd all respekt. Mer tveksam ställer man sig till det slags framtidsoptimism som kommer till uttryck i inledningen till ordboken (s. 6): "Det finns en enorm potential i interlingua som en faktor i framtida internationell kommunikation." På ett rent teoretiskt plan skulle detta förstås kunna vara sant. Verkligheten ser dock annorlunda ut, särskilt jämfört med för hundra år sedan. Som redan påpekats har inte minst engelskans förkrossande dominans som internationellt kommunikationsmedel i praktiken ryckt undan det mesta av grunden – om kanske inte de ursprungliga motiven – för ett mer allmänt bruk av artificiella hjälpspråk som interlingua.⁵

Med andra ord, hämtade från och sammanfogade (förhoppningsvis korrekt) med hjälp av *ISO*: *Le futuro de interlingua qua un medio de communication veridicamente international es probablemente multo incerte.*

Litteratur

- Chomsky, N. 1993: *Language and Thought*. Wakefield, Rhode Island: Moyer Bell.
- Crystal, D. 1987: *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press.
- Eco, U. 1995: *The Search for the Perfect Language*. Oxford: Blackwell.
- Nationalencyklopedins ordbok 1995–96*. Utarbetad vid Språkdata, Göteborgs universitet. Bra Böcker.
- Ogden, C.K. 1968: *Basic English: International Second Language*. New York: Harcourt, Brace & World
- Prismas främmande ord. 25.000 ord med förklaringar och härledning*. 1984. Stockholm: Prisma.

⁵ Till skillnad från de flesta andra (jfr Crystal 1987:355) vill Eco (1995:332f.) dock inte helt utesluta möjligheten, i ett europeiskt perspektiv, att ett internationellt hjälpspråk skulle kunna framstå som en lösning på språkpolitiska tvistefrågor, t.ex. inom ett framtida EU.

Slaughter, M.M. 1982: *Universal Languages and Scientific Taxonomy in the Seventeenth Century*. Cambridge: Cambridge University Press.

Jette Pedersen, i samarbejde med Tine Pedersen

Jens Axelsen: *Dansk-Engelsk Ordbog*. København: Gyldendalske Boghandel, Nordisk Forlag A/S 1995, DKK 220,-.

Jens Axelsen: *EN RØD KLASSIKER. Dansk-Engelsk Ordbog 1845–1995*. København: Gyldendalske Boghandel, Nordisk Forlag A/S 1995.

Dansk-Engelsk Ordbog i Gyldendals røde ordbogsserie er den 10. i rækken i ordbogens 150-årige historie.

Bortset fra omslaget, som er det velkendte røde, slidstærke, synes den foreliggende udgave ikke at have meget til fælles med sine ni forgængere. Dels er denne udgave væsentligt tykkere, dels er der tale om et nyt layout, idet ordbogen fx fremstår med en relativt større, behageligt letlæselig skrift og en opdeling af artiklerne i polysemiindeks, hvor overskueligheden yderligere understøttes af, at hvert nyt betydningsnummer indleder en ny linie. Sidst men ikke mindst er der sket en ændring af konceptet, som snarere end på tidligere udgaver leder tanken hen på Dansk-Engelsk Undervisning, som udkom i serien Gyldendals Røde Undervisningsordbøger i 1990. Dansk-Engelsk Ordbog fås også i en elektronisk udgave, som ikke er modtaget til anmeldelse.

Dansk-Engelsk Ordbog (i det følgende DEO) har fire byggedele i forteksten: Forord, Vejledende bemærkninger, Tegn og forkortelser samt Anvendt litteratur, men ingen indholdsfortegnelse. Listen over anvendt litteratur, der er opdelt i ordbøger, idiomatik og synonymik, skal fremhæves som et efterfølgelsesværdigt initiativ, ikke alene set fra et kollegialt synspunkt, men også af hensyn til brugerne, som hermed gives en mulighed for at bedømme ordbogens baggrund og oplysninger. Efter forteksten følger den dansk-engelske lemmaliste med et omfang på godt 1.000 sider. Bagteksten indeholder en fortegnelse over uregelmæssige engelske verber. På omslagets indersider findes et kort over UK henholdsvis USA og Canada.

I fortekstens vejledende bemærkninger beskrives DEO's intendede brugere og brugsfunktioner som følger: "Ordbogen er tænkt som en aktiv ordbog for danske, men af hensyn til udenlandske brugere er der angivet grammatiske oplysninger til de danske opslagsord." I *EN RØD KLASSIKER*, som er udgivet i anledning af 150-års jubilæet og forfattet af ordbogsredaktøren, beskrives og afgrænses en *aktiv* over for henholdsvis en *passiv* og en *biskopal* ordbog som følger:

En aktiv ordbog hjælper en til at udtrykke sig på det fremmede sprog.
For eksempel vil en aktiv dansk-engelsk ordbog have oplysninger om de

engelske oversættelser, om deres bøjning og brug, til oplysning for danskere, mens den ikke behøver oplyse om de danske opslagsords køn og bøjning, for det ved danskerne i forvejen. [...] Omvendt vil en passiv dansk-engelsk ordbog have oplysninger om de danske ords køn og bøjning, til brug for englændere eller andre udlændinge. [...] En biskopal ordbog [...] prøver at forene de to funktioner. (p. 144).

I henhold til denne forståelse må der dog snarere være tale om en biskopal ordbog, men en yderligere diskussion af spørgsmålet hæmmes af den manglende præcisering af DEO's målgruppe, ikke mindst de forudsatte brugerforudsætninger, samt oplysning om konkrete brugsfunktioner, herunder hvilke problemtyper ordbogen skal medvirke til at løse, jvf. gennemgangen nedenfor.

Bilingvale danske ordbogsudgivelser udmærker sig traditionelt ikke ved deres udførlige brugervejledninger. Således heller ikke DEO. Der skal ikke her plæderes for de meget lange og ofte temmelig uoverskuelige vejledninger, som fx kendes fra flere monolingvale engelske ordbøger, og som brugeren erfaringsmæssigt alligevel sjældent tager sig den fornødne tid til at sætte sig ind i. Hvad der savnes, er en redøgørelse, i form af konkrete skematiske fremstillinger, for den strukturelle opbygning af de enkelte ordbogsartikler, herunder hvilke oplysningstyper der kan findes i hvilke artikelkomponenter.

En væsentlig fordel ved sådanne skematiske fremstillinger er, at man umiddelbart får et præcist indtryk af, hvordan en konkret ordbogsartikel er opbygget. Det er derfor ikke nødvendigt at slå op i ordbogens ordliste(r) for at se de konkrete ordbogsartikler. En anden fordel er, at en sådan konkret skematisk fremstilling klart illustrerer de grafiske midler, der anvendes i ordbogsartiklerne. Denne fremgangsmåde ved præsentationen af artikelstrukturer er både pladsbesparende og en praktisk måde, hvorpå man i letforståelig form kan præsentere ofte indviklede strukturelle sammenhænge. Dette betyder dog ikke, at konkrete skematiske fremstillinger kan stå alene. De skal [...] ledsages af en forklarende tekst. (Nielsen 1994:181).

Kommenterede konkrete ordbogsartikler som en del af brugervejledningen findes fx i Cobuild: Guide to the Dictionary Entries, Longman Language Activator: Explanatory page og ALD: Key to Entries og kan anbefales som eksempler til efterfølgelse.

DEO er med sine 65.000 lemmata absolut at betegne som en maksimerende ordbog. Foruden at være baseret på tidligere udgivelser og de i litteraturlisten anførte opslagsværker er lemmabestanden i henhold til forordet et resultat af avis- og litteraturlæsning, radiolytning og TV-kigning. Herudover har redaktøren haft adgang til Den Danske Ordbogs

40 millioner ord store danske tekstkorpus. DEO lemmatiserer såvel leksemer som grammatiske ord, herunder uregelmæssige former af substantiver, verber og adjektiver, fx *køer*, *skreg*, *skreget* og *større*, *størst*, samt nogle få orddele, fx *langsomt-*, *stor-* og *store-* og *videre-*. DEO behandler først og fremmest ord, som traditionelt tilordnes almensproget, men herudover tages der i forkortelseslisten højde for en række fagområder som *arkitektur*, *arkæologi*, *astronomi*, *biologi*, *botanik*, *datalogi*, *elektricitet*, *forsikringsvæsen*, *fotografering*, *kemi*, *landbrug*, *lægevidenskab*, *meteorologi*, *søfart* m.v., i alt 46. Måske med undtagelse af *heraldik*, forekommer udvalget af fagområder hensigtsmæssigt i en almen sprogordbog, idet de alle typiske kan indgå i eller have en direkte relation til dagliglivets foreteelser, og dermed kan siges at tilhøre fællesmængden mellem almensprog og fagsprog. Den indbyrdes afgrænsning af de forskellige fagområder kan undertiden forekomme noget diffus. Hvad er fx relationen, eller vel rettere ikke-relationen, mellem *religion* og *bibelsk*? Mellem *parlamentsvæsen* og *politik*? Eller *anatomi* og *lægevidenskab*? Og hvorledes skal brugeren forholde sig til markeringen *fagl.* = fagligt sprog, jvf. fx

spiserør (*anat.*) gullet; (*fagl.*) oesophagus

hvor relationen mellem *fagl.* på den ene side og fagområdeangivelsen *anatomi* på den anden forekommer uklar. Måske er *fagl.* en stilmarkør? I givet fald burde markeringen være kontrasteret til en anden stilmarkør, fx T, som i henhold til forteksten betegner "uformelt sprog; uformelt skriftsprog og daglig tale".

I betragtning af det store antal check-opslag, jeg har foretaget, må det siges at være overraskende få gange, jeg er gået forgæves. Blandt de, som sagt meget få, ord, der ikke har kunnet slås op, er *apartheid*, *personnummerbevis*, *pædagogmedhjælper*, *svinehund* (*den indre svinehund*), *stilren* og *skriveblokering*. Sidstnævnte lykkedes det dog at finde under det mig bekendt ikke nær så hyppigt anvendte *skriveangst*. Stikprøveopslag viser, at DEO følger fint med udviklingen i sproget/samfundet, jvf. fx *betalingshospital*, *billigbillet*, *delebarn*, *deleforældre*, *dørspion*, *dørtelefon*, *elastikspring*, *kropsbevidst*, *kropsnær*, *overtræne* og *sampilsramt*, som alle er med.

Bortset fra flerordsforbindelser er alle danske lemmata ledsaget af en grammatisk angivelse, herunder oplysning om ordklassetilhørsforhold, uregelmæssig fleksion og, for substantivernes vedkommende, køn og pluralis ubestemt form, uanset om sidstnævnte er regelmæssig eller uregelmæssig. Ordbogen må således anses for et velegnet hjælpemiddel i forbindelse med dansk tekstproduktion.

Efter den grammatiske angivelse til det danske lemma følger den engelske ækvivalent hhv de engelske ækvivalenter. I henhold til EN RØD KLASSIKER (p. 121) har det været et mål ikke at anføre mere end to, allerhøjst tre, ækvivalenter til et lemma uden betydnings-differentierende oplysninger eller en stilistisk markering. Af sidst-nævnte er der forudset seks, fordelt med tre under umarkeret, neutral sprogbrug (uformelt sprog, slang, vulgært/tabuord) og tre over (formelt sprog, litterært sprog og poetisk sprog). Såvel absolut som relativt forekommer seks stilmarkører at være i overkanten, sammenlign fx ALD, som kan nøjes med tre: formel, uformel og tabu. Samtidig må det betegnes som uhensigtsmæssigt, at markeringen *litt.* har den dobbelte funktion af såvel stilmarkør som fagområdeangivelse. Ovennævnte princip med en øvre grænse for antallet af ukommenterede ækvivalentangivelser er medvirkende til, at ordbogen i mindre grad end det har været tilfældet med tidligere udgaver forudsætter den samtidige brug af en monolingval L2-ordbog.

Anmelderen er p.t. revisor på en monolingval dansk betydnings-ordbog, der bl.a. er konciperet med en opdeling af de enkelte ordbogs-artikler i delbetydninger, alle ledsaget af en definition. Sammen med en række andre kilder har jeg gennem de seneste to måneder løbende gjort brug af DEO som empirisk basis for denne opdeling. I første omgang udelukkende med henblik på nærværende anmeldelse, men efterhånden i stigende grad fordi DEO har vist sig at være en fortræffelig hjælp til at foretage den nævnte betydningsdifferentiering. Der er talrige eksempler på DEO's grundighed på dette område, her er valgt opdelingen under lemmaet *episode*. I Gyldendals Fremmedordbog listes følgende synonymmer til *episode*: *sidehandling*, *indskud*, *mindre optrin*, *enkeltstående hændelse*, mens den svenske Nationalencyklopedins Ordbok giver følgende forklaring under *episod*: (tillfällig och relativt obetydlig) händelse. I Vinterberg & Bodelsen Dansk-Engelsk Ordbog, som i parentes bemærket koster DKK 1.200,-, eller 5–6 gange så meget som DEO, behandles *episode* sådan her:

episode *en, -r* incident, //; (*sammenstød, polit. etc.*) incident (fx *frontier* i.s).

Sammenlign hertil DEO:

episode *sb. (en, -r)* **1.** (*optrin*) incident; (*større*) scene (fx *a dramatic, painful incident* (scene));

2. (*led i begivenhedsrække*) episode (*fx the love affair was only an episode in his life*);
3. (*del af tv-serie*) episode (*fx the fifth episode of the series*);
4. (*politisk, militær*) incident (*fx a frontier incident*).

I sammenligning med de øvrige kilder er der for DEO's vedkommende tale om en velkommenteret betydningsdifferentiering, som ikke blot begrænser sig til den opdeling og de ledsagende bemærkninger, der kan være behov for af hensyn til de forskellige engelske oversættelsesmuligheder, jvf. *incident*, som går igen i både betydningsnummer 1. og 4., hhv. *episode*, som optræder i såvel 2. som 3. Hertil kommer, at ingen af de øvrige tre kilder har taget højde for delbetydning 3. Samtidig er alle delbetydninger ledsaget af illustrative eksempler på engelsk, som, ud over at virke yderligere anskueliggørende, medvirker til at reducere behovet for supplerende opslag i fx learners' dictionaries.

Såvidt det kan bedømmes, er DEO særdeles pålidelig i sit valg af ækvivalenter. Det er således kun lykkedes at finde et par enkelte, lidt skæve forslag. Eksempelvis kan *computer control* for *datastyring* ikke verificeres, men derimod *data management*, jvf. fx Dictionary of Information Technology.

Grammatiske oplysninger til de engelske ækvivalenter må siges at være svagheden ved DEO. Til forskel fra hvad der gjorde sig gældende for uregelmæssig komparation af danske adjektiver som lemmata, gives der ingen oplysning om uregelmæssig komparation af adjektivale ækvivalenter. Til verbale ækvivalenter gives der hverken eksplicit oplysning om valens eller, endnu mere overraskende, om uregelmæssig fleksion (igen til forskel fra verbale lemmata) – ikke engang i form af et leksikografisk symbol som en asteriks el. lign., der kunne fungere som en implicit henvisning til bagtekstens fleksionsliste, jvf. følgende eksempel:

spise² *vb. (spiste, spist)* 1. eat (*fx bread, meat; there was nothing to eat*); (*om måltid, ret, især*) have (*fx breakfast, dinner, lunch, an egg, chicken*);

Denne relative prioritering af L1 og L2 når det drejer sig om grammatiske angivelser giver anledning til undren, ikke mindst på baggrund af at DEO i henhold til sin egen fortekst er "tænkt som en

aktiv ordbog for danske", jvf. ovenfor. En mulig forklaring kan være en generel overvurdering af de danske ordbogsbrugernes grammatiske forudsætningsniveau. Eksempelvis insisterer mere end en tredjedel af de studerende på en af landets højere læreanstalter efter godt 2 1/2 års videregående engelskstudier eksempelvis stadig på at bøje verberne *grind* og *cast* svagt. Om spørgsmålet vedrørende brugerforudsætninger hedder det i EN RØD KLASSIKER:

Kun i Gyldendals *Dansk-Engelsk Undervisning*, der er beregnet for mindre trænede ordbogsbrugere og mindre avancerede undervisningstrin, gives der direkte syntaktiske oplysninger i ordbogsteksten, i lighed med hvad der findes i de ensprogede engelske "learner dictionaries": uregelmæssige verber er mærket med *, utællelige substantiver som 'advice, information' har "utæl." efter sig, 'enjoy, avoid' har "+-ing", og i andre tilfælde gives der en forklaring i parentes, fx efter 'inform': (*NB inform skal have hensynslet med foran that-bisætning ...*). (p. 146).

Der er for de nævnte eksempler tale om leksemspecifikke fænomener, hvorfor de bør gøres til genstand for leksikografisk behandling i de relevante ordbogsartikler, jvf. nedenstående citat fra EN RØD KLASSIKER p. 146, hvor kun regelmæssigheder i L2 forudsættes indarbejdet i en ordbogsgrammatik. Man undrer sig over, hvor brugeren da skal finde oplysning om uregelmæssigheder? Det skal understreges, at jeg med leksikografisk behandling mener eksplicit leksikografisk behandling. Såfremt en ordbog skal medvirke til korrekt L2 tekstproduktion og oversættelse, kan det altså ikke anses for tilstrækkeligt at eksemplificere "syntaktiske ejendommeligheder ved oversættelserne ved udtryk i foden" (fortekstens Vejledende bemærkninger). Det er at stille for store krav til brugeren, herunder ikke mindst til hans gætteevne, at antage, at han er i stand til at udlede en grammatisk regel ud fra et eksempel på en syntaktisk ejendommelighed. I det hele taget er det behæftet med stor usikkerhed at slutte fra specifikt eksempel til generel regel, mens det omvendte ikke er tilfældet.

Om fordelingsstrukturen af grammatiske angivelser hedder det i EN RØD KLASSIKER (p. 146): "Det bedste ville nok være at indkorporere (= inkorporere) en grammatik som har de punkter der forudsættes bekendt ved brug af ordbogen, fx regelmæssig flertal af substantiver på -(e)s og regelmæssig bøjning af verber med -(e)d." Eksemplerne -(e)s og -(e)d viser, at der er tale om L2. Denne indfaldsvinkel er i overensstemmelse med de metaleksikografiske rekommendationer i henhold til hvilke det regelmæssige beskrives centralt i en ordbogsgrammatik, mens oplysning om uregelmæssigheder gives decentralt i artiklerne til de pågældende lemmata. Der har, stadig ifølge

EN RØD KLASSIKER, været flere gode grunde til ikke at indarbejde en selvstændig ordbogsgrammatik i DEO, herunder den helt udslagsgivende, at grammatikker, i modsætning til ordbøger, ikke må medbringes ved visse eksaminer. Hvor L2 er så godt dækket ind såvel undervisningsmæssigt som i form af tilgængelige grammatikker på alle niveauer, som det må siges at være tilfældet med engelsk, savnes der ikke en selvstændig ordbogsgrammatik. En sådan komponent er især indikeret i forbindelse med ordbøger til mere eksotiske sprog, således fx til Malagasy Alema, samt hvor særlige differensgrammatiske problemer gør sig gældende, jvf. Pedersen (1995).

Sidst i ordbogsartiklen, i det der som sagt benævnes artiklens fod, findes for ordbogen som helhed i alt ca. 27.000 "udtryk og vendinger" (EN RØD KLASSIKER, p. 156). I denne artikelkomponent gives oplysninger om ordforbindelser til lemmata, jvf. fx under verbet *gøre*, hvor der efter de forskellige delbetydninger gives en række kollokationer med *gøre* ledsaget af den engelske oversættelse, i alt ca. en spalte. Herefter følger to spalter med en række præpositioner og adverbier, som *gøre* kan optræde sammen med, igen ledsaget af de engelske oversættelsesækvivalenter. Andre lemmata, især verber, tager endnu mere plads op, fx *gå*, der strækker sig over 9 1/2 spalte. Der er her tale om en mikrostrukturel komponent, der bidrager til DEO's værdi som et særdeles værdifuldt hjælpemiddel i forbindelse med L1–L2 oversættelse og L2 tekstproduktion på alle niveauer.

Min datter, Tine Pedersen, spinder ende:

Jeg går i 2G, nysproglig linie, og har brugt DEO i to måneder. Der er to områder, hvor jeg synes, ordbogen kan blive bedre, nemlig dels omkring udtale og dels med hensyn til grammatiske oplysninger. Jeg vil prøve at forklare hvorfor.

DEO giver ingen oplysninger om udtale og man er derfor nødt til at slå op i andre ordbøger, når man skal forberede noget til mundtligt gennemgang på klassen. Jeg har aldrig lært lydskrift, men man udvikler efterhånden en ret god fornemmelse for, hvordan et ord skal udtales, selv om man ikke har set det før. Det sværeste er at vide, hvor trykket skal lægges. Det ville derfor være godt, hvis ordbogen indeholdt trykangivelser, også fordi man husker et nyt ord bedre, når man ved, hvordan det udtales.

Jeg har især brugt ordbogen ved oversættelse fra dansk til engelsk og i forbindelse med de essays og personbeskrivelser, vi selv skal formulere ud fra et forlæg, som regel en novelle. Her støder man tit på grammatiske problemer, fx om et substantiv er tælleligt eller utælleligt, eller man kan blive usikker på, hvordan et bestemt verbum skal

konstrueres, fx *succeed*. Slår man op i DEO under *lykkes*, får man følgende oplysninger:

lykkes *vb.* succeed (*fx* the plan succeeded); be a success (*fx* the experiment was a success); be successful (*fx* his attempt was successful); T come off (*fx* the plan (the experiment) did not come off);

□ **det** *lykkedes ham at gøre det* he succeeded in doing it; (*med besvær*) he managed to do it;

I Cobuild er problemet beskrevet meget tydeligere:

succeed

1. If you **succeed** in doing something, you manage to do it.
We have already succeeded in working out ground rules with the Department of Defense

Ud over at man her får oplysning om, hvordan man på engelsk konstruerer *lykkes nogen at gøre noget* allerede i forklaringen plus i det efterfølgende eksempel, gives der følgende oplysning i marginen:

V *in* -ing/N

I DEO skal man ikke bare et godt stykke længere ned i artiklen, før man støder på *succeed in* + *-ing*, men det, at konstruktionen ikke er beskrevet eller forklaret, men bare optræder i et eksempel, betyder faktisk, at man skal kende til problemet på forhånd og være opmærksom på det for at få udbytte af eksemplet. Og hvis man gør det og er det, slår man vel slet ikke op i en ordbog.

Bortset fra spørgsmålet om udtale og grammatik har jeg været meget tilfreds med DEO, som er behagelig i brug og enkel og nem at slå op i. Man kan næsten altid finde det ord eller den vending, man slår op, på det sted, hvor man slår op.

Litteratur

- ALD = *Advanced Learner's Dictionary of Current English*. London: Oxford University Press 1990.
- Cobuild = *Collins Cobuild English Language Dictionary*. London: HarperCollins Publishers 1995.
- Collins English Dictionary*. Glasgow: Harper Collins Publishers 1991.
- Dansk-Engelsk Undervisning*. København: Gyldendalske Boghandel, Nordisk Forlag A/S 1990.
- Dictionary of Information Technology*. Middlesex: Peter Collin Publishing 1987.
- Gyldendals Fremmedordbog*. København: Gyldendalske Boghandel, Nordisk Forlag A/S 1993.
- Longman Language Activator*. Essex: Longman Group UK Limited 1993.
- Malagasy Alema*. Antananarivo: Leximal 1991.
- Nationalencyklopediens Ordbok*. Höganäs: Bokförlaget Bra Böcker AB 1995.
- Nielsen, Sandro 1994: Ordbogens Byggede. I: Henning Bergenholtz/Sven Tarp (red.): *Manual i fagleksikografi. Udarbejdelse af fagordbøger – problemer og løsningsforslag*. Herning: Systime, 173–202.
- Pedersen, Jette 1995: Grammatiske oplysninger i tekniske fagordbøger. I: *LexicoNordica* 2, 121–142.
- Vinterberg & Bodelsen Dansk-Engelsk Ordbog*. København: Gyldendalske Boghandel, Nordisk Forlag A/S 1991.

Klaas Ruppel

Det rör på sig inom finsk-tysk lexikografi

Aino Kärnä: *Saksa-suomi-opiskelusanakirja* ('Tysk-finsk studieordbok'). WSOY Porvoo-Helsinki-Juva 1995. ISBN 951-0-19088-8. Pris FIM 175.

During the last years several middle sized German-Finnish dictionaries have been published in Finland. The one by Aino Kärnä is presented here because the author of this review considers it to be the best of these dictionaries. Finns who want to understand German texts are the target group of this "passive" dictionary. Some critical remarks are brought up mainly concerning the layout and formal matters but they are only of little importance compared with the highly reliable contents, i.e. the part of the publication its author is actually responsible for. With this highly recommendable work the German-Finnish lexicography made the step into the present closing a gap that had been deepening for decades. The most urgent desideratum now is an totally renewed German-Finnish comprehensive dictionary. In the review the item of grammatical gender vs. natural sex and the lexicographical handling of this complex is also discussed.

Hösten 1995 utkom en ny tysk-finsk ordbok. Den har sammanställts av Aino Kärnä och är resultatet av cirka tio års mödor (s. 5). Arbetet har utförts vid sidan av författarens arbete som lärare. Det är en mellanstor ordbok som enligt bakpärmen innehåller ca 43.000 ord och uttryck. Tyskan har granskats av veteranen inom finsk-tysk lexikografi, professor Ingrid Schellbach-Kopra, och den finska språkdräkten har granskats av fil.kand. Kitti Grönholm. Det finns alltså alla skäl att anta att ordboken är "en tillförlitlig tysk-finsk ordbok" (s. 5).

Upphovsrätten innehålls av författaren Aino Kärnä själv, som sig bör. Risker är numera stora att allt fler av författarna till mellanstora ordböcker får problem att behålla upphovsrätten till sina verk. För mindre ordböcker anges det inte ens alltid vem som har upphovsrätten. När författarna lämnar in manuskripten i elektronisk form vill förlagen utnyttja materialet också för andra syften och författarens upphovsrätt till den egna produkten vållar då problem, det vill säga den kostar. Därför ersätts författaren i form av ett engångsbelopp, och förlaget tillägnar sig rätten att göra vad det vill med materialet. Vad gäller fickordböcker kan det t.ex. vara frestande att med några enkla grepp åstadkomma en ordbok i den andra riktningen. Elektroniken är en utmärkt tjänare, men tar den herraväldet kan den åstadkomma stora skador. Ordboksanvändaren och den som står i beråd att köpa en ordbok

gör alltså klokt i att läsa igenom förordet och studera titelbladet och dess baksida för att få reda på hur ordboken kommit till, såvida detta framgår.

Målgruppen klart angiven

Kärnä anger osedvanligt klart i användardirektiven att ordboken är "en s.k. passiv ordbok" (s. 7). Den riktar sig alltså till finnar som vill förstå tysk text. Följaktligen ges nödvändig grammatisk information om de tyska uppslagsorden medan ingen sådan information ges om de finska ekvivalenterna. I en ordbok för skolelever och andra språkinlärare är det här en logisk lösning som spar utrymme och som säkert också har gjort det möjligt att ta med mera tyskt material.

Uppslagsorden har hämtats "bl.a. ur massmedier, dagstidningar, tidskrifter, från ordböcker och uppslagsverk från olika ämnesområden och dessutom har en del talspråkliga ord tagits med" (s. 7). Tyngdpunkten ligger dock på skriftspråket.

Ordboken är kompakt till sin storlek och det allmänna intrycket är i och för sig att artiklarna är tydliga och klara. Tyvärr har ett tråkigt modedefenomen trängt in även i denna bok, ett fenomen som man nästan inte undgår i några uppslagsböcker nuförtiden. En flerhundraårig tradition att inte blanda antikva och grotesk iakttas inte längre. Uppslagsorden (och de siffror som delar in artiklarna i betydelsemoment) är tryckta i grotesk och ekvivalenter, exempel och metaspråk i antikva. Man får hoppas att det här är ett övergående fenomen – att gamla visdomar segrar bara den första förtjusningen över alla fonter har lagt sig bland dem som gör layout på dator. Blandningen av de två typerna gör inte ordboksartiklarna tydligare, tvärtom ser uppslagen röriga ut.

En märklig lösning är att utelämna punkterna efter en mängd förkortningar (ett krav från förlaget för att spara utrymme enligt uppgift från författaren). Bara de allra vanligaste finska och tyska förkortningarna har fått behålla sina punkter. Av de 212 förkortningarna i förkortningslistan står 168 eller hela 79% utan punkt, i strid med finska och tyska rättskrivningsregler. Jag inser inte vad det är för fördel med det. Den godtyckliga ortografin i det här avseendet står i bjärt kontrast till författarens omsorgsfulla redigeringsarbete i övrigt. Samma fenomen kan dock iakttas även i andra nya ordböcker. Att utesluta punkten sist i en ordboksartikel är däremot helt acceptabelt. Artiklarna är överlag korta och huvudelementen utgörs i regel bara av uppslagsordet (med dess grammatiska angivelser) och ekvivalenterna.

Förordet finns både på finska och tyska och utöver de finska användaranvisningarna finns det kortfattade anvisningar på tyska. Det här går stick i stäv med det faktum att ordboken är avsedd "både för sådana

som just börjat läsa tyska och för längre hunna och sådana som t.ex. läser tyska texter i sitt arbete", d.v.s. att den alltså först och främst är gjord med tanke på finnar (s.7). Också i de tyska användardirektiven sägs det att ordboken är passiv och i första hand riktar sig till finnar. Men medan det i de finska direktiven sägs att "ingen böjning ges för de finska ekvivalenterna" är motsvarande formulering i de tyska direktiven försiktigare: "Entsprechend wird auf Angaben zur Grammatik finnischer Wörter, etwa auf die Flexionsformen, weitgehend verzichtet." (s. 13) Vad jag kunnat konstatera saknas dock sådana angivelser helt och hållet. Eftersom ordboken uttryckligen riktar sig till finnar är det inte berättigat med en utförlig presentation på tyska. Det kan rent av anses som vilseledande. Att uppslagsorden är alfabetiskt ordnade enligt tysk princip är en nödvändig och förtjänstfull upplysning i den finska bruksanvisningen, men i den tyska förefaller den högst märklig.

De komprimerade och sakkunniga användardirektiven visar tydligt att ordboksförfattaren har hög ambitionsnivå. Eftersom det handlar om en *studieordbok*¹ saknar man egentligen bara litteraturhänvisningar för uttal och grammatik.

Innehållet övertygande

Ordboksartiklarna är klara till sin uppställning. Varje uppslagsord får en egen artikel och det gör det lätt att hitta det ord man söker. Enda undantaget är synonyma ord som kommer efter varandra i alfabetet. Tryckangivelsen för uppslagsorden ges i form av en punkt under betonad vokal. Punkten sätts ut om trycket inte ligger på första stavelsen (t.ex. **Bulëtte**, **nervo<55s**). Vokallängden måste användaren själv sluta sig till, men den framgår ju i regel av skriftbilden. Det här är en konsekvent metod att tillämpa i en passiv ordbok. Hade man velat markera lång och kort vokal olika borde också de ord som har trycket på första stavelsen i konsekvensens namn ha fått betoningmarkerings, vilket inte utan vidare hade gjort det klarare. Trycket anges inte heller i sammansättningar; i flerledade sammansättningar kunde det ha underlättat gestaltningen av ordet om huvudtrycket hade getts, t.ex. **Heißwa5sser-bereiter**.

För att ange till vilken del av ordet ändelserna i böjningsangivelserna skall fogas används lodstreck (t.ex. **Schlu|ß** *m* -*sses*). Omljud

¹ Det finska verbalsubstantivet *opiskelu* har ett vidare betydelseomfång än sin svenska motsvarighet. Ordboken är en mellanstor ordbok av det slag som traditionellt brukar användas av skolelever och av folk som läser språk på sin fritid snarare än av språkstudenter vid universitet.

anges med en asterisk (**Raum** *m* -[e]s, -e*). För att markera att ett verb är löst sammansatt används bakstreck (t.ex. **offen\lassen, an\richten**). En speciell markering, som måste vara till glädje för en finsk användare är den punkt som anger ordledsgräns i ord som kan vara svåra att gestalta. Tyvärr har tecknet, "en liten punkt" (s.12), (på grund av tryckfel?) blivit störande stort (t.ex. **Topf•lappen, Schau•fenster**). Markeringen kunde ha utsträckts till ord som *Dornröschen, Höschen* (**Dornrös•chen, Hös•chen**) o.s.v., där kombinationen *sch* inte står för ett sje-ljud. Hur som helst visar den här förträffliga markeringen att författaren har gedigen erfarenhet som universitetslärare och lexikograf.

Utmärkt är också att ta med sådana verbformer som uppslagsord som avviker så pass mycket från infinitivformen att en nybörjare inte utan vidare kan sluta sig till den. För de oregelbundna verbens del tillhandahåller Kärnä två olika tabeller. Det ena är en tabell över de olika vokalväxlingstyperna, dit det hänvisas från artiklarna med hjälp av indexsiffror, t.ex. **tragen**⁴, **wachsen**⁴ (*sein*). Under ifrågavarande punkt i tabellen återfinns vokalserien och temaformerna för ett exempelverb, t.ex.:

		Inf / Preesens	Imperfekti	Partis perfekt
4	a/ä-u-a	laden/lädt	lud	geladen

Verben ovan skall alltså böjas *tragen, trägt, trug, hat getragen* och *wachsen, wächst, wuchs, ist gewachsen*. I slutet av ordboken finns en bilaga med en tabell över de vanligaste oregelbundna verbens temaformer. Mycket praktiskt och användarvänligt!

Uppslagsordet följs av de grammatiska uppgifterna om ordklass, genus och böjning. Uttalsuppgifter ges vid behov (t.ex. **Karriere** [ka'r—e:rβ]), likaså uppgifter om stil och fackområdestillhörighet. Uppslagsordens betydelse klargörs med hjälp av ekvivalenter i enlighet med den strävan som anges i användardirektiven, "att de olika betydelserna framgår av de ekvivalenter som ges". Om ett uppslagsord har två eller flera olika huvudbetydelser delas artikeln in med hjälp av siffror. Om ett ord kan höra till olika ordklasser används romerska siffror.

Att författaren i första hand anger betydelserna med hjälp av ekvivalenter torde bero på en strävan att spara utrymme. När man bläddrar i ordboken upptäcker man dock att det också finns exempel, särskilt exempel på idiom, i väldigt många av artiklarna. Mitt intryck är att ekvivalenterna genomgående är väl valda och ändamålsenliga. Aino Kärnä har, biträdd av språkgranskaren Kitti Grönholm, bemödat sig om den finska idiomatiken. Hon har inte nöjt sig med direkta översättningar, där sådana inte gör sig, utan strävat efter att ge typiska finska uttryck som motsvarighet. Författaren uppger att det har varit till nytta i arbetet att också aktivt leta efter tyska motsvarigheter till finska

uttryck. Tack vare denna bidirektionella metod har hon både fått fram nya motsvarigheter och också kunnat få bekräftat att tilltänkta motsvarigheter är korrekta. Frukten av metoden syns bl.a. i ekvivalenterna till sammansättningar och i idiomerna, t.ex.:

- Kaufhaus** tavaratalo 'varuhus' (i Rosendahl-Hirvensalos skolordbok: kauppaahuone 'handelshus')
- Luftfahrt** ilmailu; lento[liikenne] 'luftfart; flyg[trafik]' (R-H: ilmailu, ilmaliikenne 'luftfart; lufttrafik')
- Meeresarm** [meren]lahti; vuono '[havs]vik; fjord' (R-H: merenhaara 'havsarm')
- Kaffee das ist kalter** ~ tuo on vanha vitsi, tuo nyt ei ole mistään kotoisin 'den där är inte mycket att hurra för'
- Katzensprung es ist nur ein ~ bis dahin** se on vain kivenheiton päässä 'den ligger bara ett stenkast ifrån' (R-H: kissanloikkaus; lyhyt matka, pieni harppaus "'kattskutt"; kort väg; litet kliv')

Ett resultat av den bidirektionella arbetsmetoden är också att ordboks-användaren kan hitta uttryck som avser tillstånd och som lexikaliseras olika på finska och tyska. För adjektiv ges t.ex. ofta konstruktionen adjektiv + *werden*, om den i finskan motsvaras av enkla verb, t.ex.:

- alt** ~ *werden* 'bli gammal' vanheta, vanhentua; **älter werden** 'bli äldre' (*myös*) ikääntyä, vanheta
- klein ganz** ~ *werden* 'bli generad' nolostua, mennä hämilleen
- stark** ~ *werden* 'bli stark' vahvistua, voimistua
- staubig** ~ *werden* 'bli dammig' pölyyntyä, pölyttyä

I enstaka fall kan man diskutera om en ekvivalent är riktig eller om inte en betydelsenynans fattas, men de fallen är så få, att det vore småpetigt att lyfta fram dem här. Det vore inte att göra ett synnerligen lyckat verk rättvisa.

Överlägset bäst i sin serie

För jämförelsens skull har jag gått igenom och jämfört orden på *h* i den förhandenvarande ordboken (s. 284–318), i en annan nyutkommen ordbok, Klemmt 1993, s. 283–313, och i en äldre ordbok, Rosendahl–Hirvensalo 1973, s. 221–249. De tre ordböckerna är jämförbara vad gäller omfång och målgrupp; Kärnäs ordbok innehåller ca 43.000 ord och uttryck (enligt texten på bakpärmen), Rosendahl–Hirvensalos ordbok innehåller ca 45.000 uppslagsord (enligt förordet) och Klemmts

ordbok ca 30.000 ord och uttryck (enligt uppskattning). Orden på *h* är 1.500 i Kärnä (3,3%), 2.000 i Rosendahl–Hirvensalo (4,5%) och 1.230 i Klemmt (4%).

När man jämför Kärnäs ordbok med dess föregångare, Rosendahl–Hirvensalos tysk-finska skolordbok, vars 12:e upplaga från 1973 jag använde, märker man att Aino Kärnä har varit mycket självständig när hon samlat material för sin ordbok. En helt föråldrad ordbok har ersatts med ett modernt och tidsenligt verk. Av uppslagsorden på *h* är det bara 872, eller drygt hälften som förekommer i bägge ordböckerna.

En bild av vilket slags ord det är som förekommer i bara endera ordboken får man om man jämför med Dudens stora ordbok (banden 3 och 4 som innehåller *h* utkom 1993–94). Av de 1.130 ord som återfinns bara i den äldre boken är över 100 (knappa 10%) markerade som föråldrade eller något ålderdomliga i Duden. Av de ord som finns bara i Kärnäs ordbok har ett enda ord den markeringen i Duden:

Heredität *f* -, -en **1** perintö 'arv' **2** periytyminen 'nedärvning'

Eftersom det motsvarande adjektivet finns med i ordboken kunde substantivet gott och väl ha utgått:

hereditär (*adj*) **1**perinnöllinen 'ärfilig' **2** perintö-,perimis- 'arvs-'

Så har skett i många andra fall.

Både Kärnä och Rosendahl–Hirvensalo innehåller uppslagsord som inte finns i Duden. I Kärnä är de 20 till antalet (1% av uppslagsorden på *h*); det rör sig om neologismer och aktuella sammansättningar (t.ex. *Handy* 'nalle, ficktelefon' och *Hauptspeicher* 'primärminne'), vilket vittnar om Kärnäs aktualitet. Av de 287 uppslagsorden på *h* som finns i Rosendahl–Hirvensalo men inte i Duden (14% av uppslagsorden på *h*) är största delen sammansättningar. Många är föråldrade eller något ålderdomliga.

Klemmts nya ordbok är något mindre omfattande än Kärnäs. Enligt förordet innehåller den en hel del talspråkligt material. Kärnäs ordbok innehåller 760 ord på *h* som inte finns i Klemmt. Klemmts ordbok innehåller, trots att den är mindre till omfånget, 475 ord som inte ingår i Kärnä. Av dessa är hela 115 stycken hänvisningslemman. Som hänvisningslemman har alla något avvikande pluralformer tagits med och alla, verkligen alla, perfekt particip-former bildade med prefixet *ge-*, också fast verben är helt regelbundna. Som ett exempel på detta en sekvens uppslagsord som följer efter varandra i alfabetet (ekvivalenterna har uteslutits):

heranbilden *bildet heran, bildete heran, hat herangebildet*
heranbringen *bringt heran, brachte heran, hat herangebracht*
herangebildet ks heranbilden

herangebracht ks heranbringen
herangekommen ks herankommen
herangemacht ks heranmachen
herangewachsen ks heranwachsen
herangezogen ks heranziehen
herankommen *kommt heran, kam heran, ist herangekommen*
heranmachen *macht sich heran, machte sich heran, hat sich herangemacht*
heranwachsen *wächst heran, wuchs heran, ist herangewachsen*
heranziehen *zieht heran, zog heran, hat herangezogen*

Det här gör att Klemmets ordbok har svällt ut så att den verkar vara dubbelt så stor som Kärnäs utan att information har tillförts. Att tyngdpunkten skulle ligga på talspråk verkar också överdrivet; av de ord som inte återfinns i Kärnäs ordbok är omkring tjugo talspråkliga medan ett tiotal är skriftspråkliga eller ålderdomliga. Dessutom förekommer det felaktigt bildade particip (*hänggeblieben* i.st.f. *hängengeblieben*, *hintergegangen* i.st.f. *hintergangen*) och tryckfel (*hevorgegangen* i.st.f. *hervorgegangen*, *hinweggeschen* i.st.f. *hinweggesehen*, *Hirsarten* i.st.f. *Hirsearten*).

Ett par tryckfel har jag faktiskt stött på också i Kärnäs ordbok; **Läuferei** i.st.f. **Lauferei**, **Schirmherr** f i.st.f. **Schirmherr** m.

I jämförelse med Klemmets ordbok, som förefaller ganska amatörmässig, framstår Kärnäs ordbok som ett gediget verk av en kunnig författare. När en ordbok är ett enmansverk brukar det finnas risk för att ämnesområden som ligger författaren varmt om hjärtat får för stort utrymme. På basis av uppslagsorden på *h* kan man konstatera att Klemmt tydligen intresserar sig för bilar och ekonomi. I Kärnäs ordbok har jag inte kunnat skönja någon liknande snedvridning vad gäller uppslagsorden; hon har tydligen bedrivit flitig källundersökning och har gedigen erfarenhet.

Generiskt genus ett problem

En av de hetaste potatisarna på tyskt språkområde är sedan flera år s.k. grammatiskt genus för ord som syftar på person och som utgörs av yrkesbeteckningar och liknande. Med hänvisning till jämställdheten mellan könen har det hävdats att det är kränkande för kvinnor om man använder ordet *Lehrer* om man talar om lärare i allmänhet. I stället borde man säga *Lehrer oder Lehrerin* 'lärare eller lärarinna' eller något liknande. Tyska platsannonser måste numera vara könsneutrala så att både den manliga och den kvinnliga yrkesbeteckningen används. Nya lagar sätts under lupp i det här avseendet, och gamla med för den delen.

Grammatiskt sett är det här ett underligt förfarande. Orden *Lehrer*, *Schüler*, *Student* o.s.v. syftar, beroende på kontext, antingen på en manlig individ eller på en person i största allmänhet som är lärare, elev, student. I det första fallet är det i regel fråga om en bestämd person och grammatiskt och naturligt kön stämmer överens. Talar man om en kvinnlig individ använder man i motsvarande fall *Lehrerin*, *Schülerin*, *Studentin*. I det senare fallet är det fråga om s.k. generiskt grammatiskt genus och i den användningen, som är mycket vanligare, tas det inte ställning till naturligt kön. Så långt grammatiken.

Samhällsutvecklingen har dock lett till att många språkbrukare inte längre godkänner kategorin generiskt genus, utan de likställer helt och hållet grammatiskt genus med naturligt genus. Denna paradigmförändring medför ett enormt tryck på det tyska språksystemet, ett tryck som knappast något språk med genus som en grammatisk kategori torde undgå.

Ordboksförfattaren måste beakta den här utvecklingen. I Rosendahl–Hirvensalos ordbok är uppslagsorden på *-in* (*Lehrerin*, *Schülerin*, *Richterin* o.s.v.) ytterst få. Kärnä däremot har haft som strävan att ange bägge formerna, vilket ofta går att göra i en och samma artikel:

Bewohner *m -s, -, -in f -, -nen* asukas 'invånare; boende'
Bezieher *m -s, -, -in f -, -nen* (*lehden*) tilaaja '[tidnings]prenumerant';
 (*tuen, eläkkeen*) saaja '[understöds-, pensions]tagare'
Verbraucher *m -s, -, -in f -, -nen* kuluttaja 'konsument'

Fallgröparna är dock många och en närmare granskning visar att *-in*-formerna trots allt i många fall fallit bort:

a) i en del fall speglar bortfallet snarast verkligheten: *Legionär* 'legionär', *Hafenarbeiter* 'hamnarbetare' (men *Mitarbeiter*, *-in* 'medarbetare, arbetskamrat'), *Seiler* 'repslagare' o.s.v.

b) ibland beror bortfallet tydligen på utförandetekniska skäl:

Fünfziger *m -s, - 1* viisikymmenpenninen 'femtiopfennigsslant' **2**
 viisikymppinen, viisikymmenvuotias 'femtioåring, -årig'
Hörer *m -s, - 1* (*radion*) kuuntelija 'lyssnare' **2** (*puhelimen*) kuuloke,
 luuri '[telefon]lur'
1 Läufer *m -s, - 1* (*urh*) [kilpa]juoksija '(idr.) löpare'; (*jalkapallossa*)
 tuki[mies] '(fotboll) halvback' **2** (*š akissa*) lähetti '(schack) löpare';
 (*hist*) airut 'härold'
2 Läufer *m -s, -* käytävämatto; kaitaliina ('korridormatta;
 [bords]löpare')

...

c) för en stor grupp personbetecknande ord ges endast den generiska varianten och ingen *-in*-form, t.ex. *Arbeitgeber* 'arbetsgivare', *Arbeitnehmer* 'arbetstagare', *Athlet* 'atlet; tävlingsidrottare', *Bettler* 'tiggare', *Funker* 'radiotelegrafist', *Funktionär* '[parti-, fackförenings-] funktionär', *Klassiker* 'klassiker', *Mittler* 'mäklare, medlare', *Stürmer* 'anfallsspelare', *Stylist* 'formgivare, designer', *Taucher* 'dykare' och många fler.

Vad den första gruppen beträffar är Duden konsekventare och ger femininformen för samtliga (med undantag för **Legionärin*).

För grupp två hade femininformen lett till komplikationer i framställningen:

- 1 **Fünfziger** *m -s, -* viisikymmenpenninen, viiden kyymmenen markan seteli [!] 'femtiopfennigsslant; femtiomarkssedel [!]
- 2 **Fünfziger** *m -s, -, -in f -, -nen* viisikymppinen, viisikymmenvuotias 'femtioåring, -årig'

- 1 **Hörer** *m -s, -* (*puhelimen*) kuuloke, luuri '[telefon]lur'
- 2 **Hörer** *m -s, -, -in f -, -nen* (*radion*) kuuntelija '[radio]lyssnare'

- 1 **Läufer** *m -s, -* (*urh*) [kilpa]juoksija 'löpare'; (*jalkapallossa*) tuki[pelaaja] 'halvback'
 - 2 **Läufer** *m -s, -* (*š akissa*) lähetti 'löpare'; (*hist*) airut 'härold'
 - 3 **Läufer** *m -s, -* käytävämatto; kaitaliina 'korridormatta; [bords]löpare'
- Läuferin** *f -, -nen* (*urh*) [kilpa]juoksija 'löpare'; (*jalkapallossa*) tuki[pelaaja] 'halvback'

Å andra sidan har komplikationer i framställningen inte alltid lagt hinder i vägen. Jämför t.ex. följande ordartiklar:

- 1 **Kunde** *m -n, -n* asiakas; ostaja 'kund; köpare'; *ein fester* ~ vakituinen asiakas 'stamkund'
 - 2 **Kunde** *f -, -n* (*ylät*) tieto, viesti, sanoma '(i högtidlig stil) budskap'
- Kundin** *f -, -nen* [nais]asiakas '[kvinnlig] kund'; *ks. 1 Kunde*

- 1 **Leiter** *m -s, -* 1 johtaja, esimies, vetäjä 'ledare, chef' 2 (*fys*) johdin, johde 'ledare'
 - 2 **Leiter** *f -, -n* tikkaat, tikapuut 'stege'
- Leiterin** *f -, -nen* johtaja, vetäjä 'ledare, chef'

- Leser** *m -, -* lukija 'läsare'; **optischer** ~ (*atk*) optinen lukija, skanneri 'optisk läsare, skanner'
- Leserin** *f -, -nen* lukija 'läsare'

Fallen i den tredje gruppen är obegripliga, eftersom man nog hittar uppslagsord som *Fahrschüler*, *-in* 'bilskolelev', *Greis*, *-in* 'äldring',

Lektor, -in 'lektor', *Lerner*, -in 'elev, student', *Kläger*, -in 'kärande' o.s.v. Det här är synd. Man hade önskat sig ett mera konsekvent grepp av författaren på den här punkten.

Avslutning

Aino Kärnä har gjort ett förnämligt arbete. Hon har åstadkommit en utmärkt och praktisk ordbok. Det jag har att anmärka på gäller mestadels layout och en del smaksaker, som man ju kan ha delade meningar om. Innehållsligt är boken mycket övertygande och jag kan uppriktigt rekommendera den. När det gäller medelstora ordböcker har tysk-finsk lexikografi äntligen tagit steget in i nutid. Tyvärr måste vi ännu vänta en tid på att en ny tysk-finsk storordbok skall utkomma och ersätta den ytterst föråldrade ordboken av Hirvensalo. Det är definitivt nästa viktiga projekt på området. Det är ingen lätt uppgift och det är knappast numera ett projekt som en person ensam har möjlighet att genomföra. Det kan också ifrågasättas om någon privat förläggare har tillräckliga resurser och tillräckligt långsiktiga visioner för att förlägga en storordbok. Hur som helst kan Kärnäs ordbok i ganska stor utsträckning ses som en ersättare för den tämligen föråldrade storordboken. Samtidigt är hennes verk också en utmaning för den eller dem som tar itu med att göra en storordbok.

Litteratur

Das große Wörterbuch der deutschen Sprache 1–8. Hrsg. und bearb. vom Wissenschaftlichen Rat und den Mitarbeitern der Dudenredaktion unter der Leitung von Günther Drosdowski. 2. völlig neu bearb. und erw. Aufl. Mannheim, Leipzig, Wien, Zürich; Dudenverlag 1993–95. ISBN 3-411-04732-1.

Hirvensalo, L. 1971: *Saksalais-suomalainen suursanakirja*. ('Tysk-finsk storordbok'). Porvoo: WSOY

Klemmt, R. 1993: *Suomi-saksa-suomi taskusankirja* ('Finsk-tysk-finsk fickordbok'). Jyväskylä: Gummerus (3 uppl.)

Rosendahl, A. – Hirvensalo, L. 1976: *Saksalais-suomalainen koulu-sanakirja* ('Tysk-finsk skolordbok'). Porvoo: WSOY.

Appendix

Mellanstora tysk-finska ordböcker

- Auramo, Arja – Winckelmann, Thyro: *Saksalais-suomalainen sanakirja*. Helsinki 1942. 462 s.
Deutsch-finnisches Taschenwörterbuch. Leipzig 1972.
- Hauvonen, N.: *Saksalais-suomalainen taskusanakirja*. Turku 1891. 462 s. | 3. p. Turku 1916.
- Klemmt, R.: *Saksa-suomi sanakirja*. Jyväskylä 1992. | 7. p. Jyväskylä 1994. 30.000 hakusanaa.
- Klemmt, R.: *Suomi-saksa-suomi taskusanakirja*. Jyväskylä 1993. 3.p.
- Kostera, P.: *Saksalais-suomalais-saksalainen yleiskielen käyttösanakirja ja kieliopas*. Helsinki 1991. | 3. p. Helsinki 1995.
- Aino Kärnä: *Saksa-suomi-opiskelusanakirja*. WSOY Porvoo-Helsinki-Juva 1995.
- Köykkä, L. – Luukkonen, M. (toim.): *Suomi-saksa-suomi-sanakirja*. Jyväskylä 1992. 176 s. | 3. p. 1994.
- Lares, J. – Tuominen, T.: *Suomi-saksa-suomi taskusanakirja*. Porvoo 1995. 883 s.
- Peters, L.: *Langenscheidts Universal-Wörterbuch*. Finnisch. 2: Deutsch-finnisch. Berlin-Schöneberg s. a. 173 s.
- Rankka, A. W. – Streng(-Renkonen), W. O.: *Saksalais-suomalainen sanakirja*. Porvoo 1918. 843 s. | 6. p. Porvoo 1944. 879 s.
- Renkonen, E. – Renkonen, W. O.: *Saksalais-suomalainen pienoissanakirja*. Porvoo 1931. 595 s. | 10. p. Porvoo 1943. 513 s.
- Rosendahl, A.: *Saksalais-suomalainen koulusanakirja*. Porvoo 1973. 12. p.
- Rosendahl, A.: *Saksalais-suomalainen sanakirja*. Helsinki 1916. 602 s. | 7. p. Helsinki 1946. 915 s.
- Rosendahl, A. – Hirvensalo, L.: *Saksalais-suomalainen opiskelusanakirja*. Porvoo 1989. 673 s.
- Vesalainen, R. – Greiner, M.: *Saksan perussanasto*. Helsinki 1987. 360 s. | 4. p.

Tysk-finska storordböcker

- Godenhjelm, B. F.: *Saksalais-suomalainen sanakirja*. Suomalaisen Kirjallisuuden Seuran toimituksia 49. Helsinki 1873. 1106 s. | 2. p. 1–2. SKST 112. Helsinki 1906–16. 1745 s.
- Hirvensalo, L.: *Saksalais-suomalainen sanakirja*. [Porvoo 1963.] | 2. p. *Saksalais-suomalainen suursanakirja*. Porvoo 1971. 1592 s. > 190.000 hakusanaa.

(Översättning av Nina Martola)

Ingrid Simonnæs

Gubba, Wilhem: *Juridisk ordbog tysk-dansk*. G.E.C. Gads Forlag, København 1995

Gubbas ordbok *Rechtswörterbuch Deutsch-Dänisch*, som tittelbladet lyder på, foreligger nå i 3. utgave og representerer etter forfatterens egne ord en "fuldstændig ny bearbejdelse og en [...] betydelig udvidelse" av 2.utgave som kom ut på W. Gubbas forlag i 1987.

Motstykket til denne ordboken (Dansk-tysk juridisk ordbog) kom ut i 1993 – også på Gads Forlag.

Selv om dansk og norsk (almenspråk) har mange fellestrekk, er dette så vidt det kunne konstateres, ikke alltid tilfellet for dansk og norsk juridisk språk. Jeg har derfor konsentrert meg om å se nærmere på *Rechtswörterbuch Deutsch-Dänisch*. Jeg er ikke i stand til å vurdere om det danske oversettelseforslaget "ækvivalenten" i alle tilfeller er korrekt.

1. Innledningsdelen

1.1 Generelt

Det første iøynefallende¹ er at ordboken innledes med en kontrastiv faglig innledning om begge lands rettssystem. En slik innledning har mye for seg dersom den påtenkte brukergruppen ikke antas å ha noen nevneverdig innsikt i de viktigste kjennetegn for de angjeldende rettssystem. Brukergruppen er ikke eksplisitt nevnt i forordet til 3. utgave, og en må derfor gå tilbake til forordet til 1. utgave der den defineres slik: "primært [...] oversættere, advokater, retsinstanser, offentlige kontorer og erhverslivet, samt [...] universiteter og handelshøjskoler". Brukergruppen er m.a.o. både legmannen og fagmannen. Dette gjør det spesielt vanskelig å ha en faglig innledning som tilfredsstillende alle grupper behov for informasjon og krav til detaljopplysninger.

Denne delen er ikke utarbeidet av Gubba selv, men av Engberg som vil være den som skal fortsette Gubbas ordboksarbeid i fremtidige utgaver. Med denne merknaden in mente kommenteres denne delen først, før jeg tar fatt på selve ordboken.

¹ Dette er en tilnærming som hittil har vært heller uvanlig for en ordbok. Men i serien Gads fagordbøker bygger flere, bl.a. Torben Henriksen *Juridisk ordbog spansk-dansk*, nettopp på det samme prinsippet.

1.2 Avgrensning av delområder i innledningsdelen

Den kontrastive faglige innledningen foretar først en avgrensning av hvilke delområder av rettssystemet som behandles i denne delen av ordboken. Her finnes også en begrunnelse for at det er kun prosessretten som behandles fra den offentlige rett, nemlig "Es werden keine weiteren Teile des öffentlichen Rechts hier dargestellt [als das Prozeßrecht], weil sie **weniger zentral** [min utheving] für den Lemmabestand im Wörterbuch sind" (s.17). Denne påstanden er objektivt sett neppe korrekt. Dersom brukeren er interessert i ytterligere informasjon om de respektive rettsområder, henvises han til litteraturlisten "am Ende der Arbeit". Dette må forstås dithen at det er på slutten av den delen Engberg har utarbeidet. En slik litteraturliste finnes da også (s. 49), men er etter anmelderens oppfatning med referanse til bare 16 verk i knappest laget. Således finnes det ingen henvisning til litteratur om bl.a. forfatningsretten og andre områder under prosessretten enn for sivile saker og straffesaker. Dette til tross for Engbergs egen vektlegging av prosessretten som sentral.

Et annet prinsipp som er lagt til grunn i denne kontrastive faglige innledningen, er at det under presentasjonen av det danske rettssystem så vidt mulig brukes tyske betegnelser for danske begrep etterfulgt av tilsvarende dansk betegnelse ved førstegangs bruk. Dersom forskjellene i rettssystemet er vesentlige, ses det bort fra denne fremgangsmåten og det brukes kun den danske betegnelsen. Grensen for å konstatere at "forskjellene er vesentlige" er ikke alltid lett å trekke og kan diskuteres.

Det forstyrrer også at en i innledningens § 4 ("Vertragsrecht") begynner med "Generelles" og tar for seg dansk rett først for å komme tilbake til forholdene i dansk rett etter å ha beskrevet de tyske rettsreglene på dette området. – Ved behandlingen av de andre rettsområder begynner forfatteren derimot alltid med forholdene i tysk rett.

1.3 Struktur og språkbruken i innledningsdelen

Den faglig kontrastive innledningen er skrevet på tysk. Språkbruken er bra, og det er som regel ikke noe å utsette på denne.

Noen trykkfeil forekommer dog (f.eks. s. 19 *Reichwe-ite*, S. 26 *Regel-ungen*, s. 35 *Nachlaßve-rwalter*, s. 37 *des Amtsgericht*, s. 45 *Entsche-idung*).

Enkelte steder er formuleringene ikke helt treffende eller heller mindre vellykkete [mine uthevinger]:

- s. 18 *einige Paragraphen hinzuzufügen*².
 s. 19 *für [...] Handlung ersatzrechtlich verantwortlich gemacht werden zu können*. Det heter "haftbar".
Handlungen ausführen (ibid.) er heller uvanlig i stedet for "vornehmen", i det minste i juridisk språk.
 Jeg er heller ikke enig i formuleringen på s. 20 *hat das Recht der Bevormundung [...] abgelöst*. I den gamle formuleringen het det "Vormundschaft"³.
 s. 23 heter det f.eks. at *Verträge schriftlich verfaßt*⁴ sind. Det menes sikkert "abgefaßt".
 s. 26 heter det *Handelsgewerbetreibende*. Loven (HGB) bruker "Kaufmann", mens en ellers også kan finne "Gewerbetreibende(r)".
Anteile an der Gesellschaft haben er heller uvanlig i stedet for "halten an ..." (s. 27).
 s. 35, 36 *als Ausgangspunkt* – det menes antageligvis "prinsipiell"
 s. 37 *in kleineren Strafsachen*. Det er nok en lite heldig formulering for "in weniger schweren Strafsachen".

En må anta at det er en glipp fra Engbergs side når det sporadisk dukker opp grammatiske feil i teksten, som ellers viser at han har et meget godt grep på det tyske språket.

Det vises her f.eks. til [mine uthevinger] :

- s. 22 *davon gebunden* "daran gebunden"
 s. 24 [*besteht aus*]: *dem kausalen [Rechtsgeschäft] [...] und dem abstrakten (die Übertragung [...]) "der Übertragung"*
 s. 28 *Sowohl kirchliche als bürgerliche Eheschließung ist möglich "sind"*
 s. 28 *keiner der zukünftigen Eheleute dürfen verheiratet sein "darf"*
 s. 35 *auf einzelnen Punkten "in [...] Punkten"*
 s. 44 *führt Handlungen aus, zu dem ihn [...] bevollmächtigt "zu denen"*

Når det gjelder innholdsmessige forklaringer, så finnes det her formuleringer som ikke er helt entydige.

² Som kjent er jo BGB blitt revidert fortløpende og å snakke om «einige Paragraphen» er for svakt.

³ Jfr. "Zweiter Titel. Vormundschaft über Volljährige, §§ 1896 BGB - Stand 15. Februar 1976" [min utheving].

⁴ *Verfassen* har ifølge Duden (1995:3649) innen det juridiske språk betydningen 'schriftlich niederlegen', og dermed er "schriftlich" overflødig dersom en fortsatt vil bruke «verfassen» her.

Det vises f.eks. til s. 23, der fremgangsmåten ved kjøp av fast eiendom beskrives slik den er vanlig i Danmark. Det er ikke lett å forstå hvorfor forfatteren fremhever *Der Vertrag ist aber* [min utheving] *auch ohne Eintragung im Grundbuch zwischen den Parteien wirksam*, for lenger oppe på samme side var det allerede blitt fremhevet at avtale om kjøp av fast eiendom i henhold til dansk rett kan inngås uten å være bundet til visse former. Tinglysingen som omtales like før den her siterte setningen berører derimot rettsvernet mot tredjemann!

Språklig unøyaktig er det også å tale om at ekteskap i Tyskland alltid blir inngått (*wird* [min utheving] [...] *geschlossen* (s. 28) for en offentlig myndighet, når kriteriet nettopp er et 'geschlossen werden muß' på grunn av prinsippet om obligatorisk ekteskapsinngåelse for en offentlig myndighet.

Under omtalen av prosessretten (s. 37) er det unøyaktig å tale om at det ved underretten Amtsgericht *können* [min utheving] *zwei Schöffen* [...] *teilnehmen* når forfatteren først har presisert lovens krav for den type straffesaker dette gjelder for.⁵

På s. 46 står det i forbindelse med "Beschleunigtes Verfahren" at denne type straffesaker gjennomføres ved *mündliche Anklageschrift* [min utheving]. Formuleringen er en språklig selvmotsigelse.⁶

Layout-messig er det uheldig at det brukes paragraftegn til å markere kapittelinnledning, eksempelvis "§ 1 Einleitung". Spesielt forstyrrende virker det når paragraftegnet dukker opp i den fortløpende teksten og en må tenke seg om hva henvisningen gjelder.⁷

2. Ordboksdelen tysk-dansk

2.1 Layout

Orboksdelen er oversiktlig og brukervennlig med oppslagsordene i fet trykk. Skriftstørrelsen er tilstrekkelig stor. Eksempelsetningene/kollokasjonene på tysk er skrevet kursivt. Hele ordboksartikkelen er rykket inn. Som ovenfor under kommentaren til innledningsdelen gjøres dog

⁵ Jfr. § 28 GVG "Für die Verhandlung und Entscheidung der zur Zuständigkeit der Amtsgerichte gehörenden Strafsachen werden, **soweit nicht der Strafrichter entscheidet** [min utheving], bei den Amtsgerichten Schöffengerichte gebildet."

⁶ Jfr. § 212 StPO "Im Verfahren vor dem Strafrichter [...] kann die Staatsanwaltschaft **schriftlich** oder mündliche den **Antrag** auf Aburteilung im beschleunigten Verfahren stellen [...]"[min utheving].

⁷ Eksempelvis anføres fra s. 39: "Das Gericht ist erste Instanz bei Verfahren die Anschläge gegen die Sicherheit des Staates betreffen, und Berufungs- oder Revisionsinstanz (§17) in Landgerichtsverfahren." '§ 17' er henvisning til kapittelinnledningen, ikke til angjeldende paragraf i domstolsloven.

samme innvendingen gjeldende med hensyn til bruken av paragraftegnet under ordboksartikkelen.

2.2 Generelt

Det er prisverdig at ordbokens foreliggende 3. utgave er sterkt utvidet i forhold til de tidligere to utgaver. Dette har tydeligvis også vært forfatterens intensjon for at brukeren "skulle være sikret [...] det optimale beslutningsgrundlag" (s. 12).

I lys av den meget bredt definerte brukergruppen er det bra at brukeren får oppgitt substantivenes kjønn for det tyske oppslagsordet⁸, mens han finner både kjønn og flertallsendelse for den danske ekvivalenten. Ser en på dette ut fra et brukerperspektiv, så er nok de grammatiske opplysninger for det tyske oppslagsord mer påkrevd enn for den danske ekvivalenten dersom en forutsetter at det er flest dansker som vil bruke denne ordboken.⁹ Den danske brukeren vil opplagt ha mer bruk for de grammatiske opplysninger på tysk. Ser en for seg at brukeren har tysk som morsmål eller at han som danske er meget sikker i tysk grammatikk, er det egentlig ikke behov for å ha med grammatiske opplysninger om det tyske oppslagsord. Her er det imidlertid ut fra markedssituasjonen for et språkpar som tysk-dansk-tysk nødvendig å ta hensyn til begge brukergrupper.

Forfatteren fremhever selv at eksempelmaterialet er utvidet så betydelig "at det i det store og hele belyser alle væsentlige sammenhænge inden for juridisk sprog" (ibid.) – Jeg vil senere komme tilbake til eksempelmaterialet.

Når det gjelder oppdelingen av oppslagsord som homonymer, er denne ifølge forfatteren utbygget ytterligere og henvisningenes antall er sterkt øket. Dette kriteriet¹⁰, er dog ikke alltid fulgt, f.eks. s. 51: **a.A.** s. 54: **ABG**. Disse oppslagsord har helt klart forskjellige betydninger, som ikke er markert som sådanne.

⁸ Dette gjelder ifølge forfatteren kun for "oppslagsordet". En finner derfor ikke oppgitt kjønn når en slår opp på en forkortelse (f.eks. **A.** som forkortelse for "Anmerkung"), men må gå til substantivet for å finne kjønn. Forkortelser av typen **AbIVO** eller **ABG** står oppført med hele ordlyden i parentes, men ikke med angivelse av kjønn.

⁹ Bruksområdet for denne type ordbok er da i første rekke ikke for å kunne produsere korrekte tyske tekster, men for å være i stand til å kunne lese tyske tekster f.eks. i oversettelsesøyemed.

¹⁰ Det er mulig at forfatteren ikke har tatt hensyn til at homografer kan være homonymer.

Den største innvendingen mot ordboken ut fra et prinsipielt standpunkt er at den inneholder for mange oppslagsord som ikke kan sies å høre til det juridiske språk dersom man da ikke tøyer definisjonen meget vidt fordi disse ord kan dukke opp i et eller annet juridisk dokument¹¹. Jeg anfører eksempelvis: *abändern, Abbau* 'nedrivning', *Abdruck, abrechen, Abbiegeabsicht, Abhandlung, Abhängigkeit, Abkürzung, Bau, dafür, demnach, Erholung, flüchten, Fußnote, Geschenk, Gewissen, Gewißheit, Industrie, jederzeit, kennen, kundig, Ladefähigkeit, leblos, loslösen, Quelle, Partie, Raum, regelmäßig, römisch, scharf, scheinbar, schneiden, schwerlich, Schulausbildung, Unfall, verkennen, wahrscheinlich, wegen*.

Det er riktig at forfatteren for mange av de ovenfor siterte oppslagsord har eksempler som belyser ordet ut fra en gitt juridisk kontekst, f.eks. *durch einen Unfall verursachte Schäden oder Kosten, wegen höherer Gewalt*, men dette er etter min mening ikke nok til å kunne si at *Unfall* eller *wegen* bør tas med som egne oppslagsord i en juridisk ordbok.

Heller ikke ord som *abfeiern* hører med i en juridisk ordbok. I det minste burde oppslagsordet ha blitt markert med "dagligdags" eller noe lignende.

Andre steder er forfatteren oppmerksom på at oppslagsordet også brukes i allmennspråket og har gjengitt det med henvisning til nettopp dette. Det er bra, men en del eksempler er dermed overflødige i en juridisk ordbok som denne: f.eks. *abgeben* (allg.) 'afgive, fremsætte, indsende', *abgrenzen* (allg.) 'avgrænse', *abberufen, ablegen, Prüfung* 'eksamen', *Vortrag (Rede)* o.l.

Også ordboksdelen er i det store og hele fri for trykkfeil, se dog eksempelvis: s. 75: *AKZ (Zeitschrift der Akademie für Defusches Recht)*, s. 315: *zeitliche* Freiheitsstrafe eller på omslagssiden bak: *Anwendung*.

2.3 Spesifikke kommentarer til mikrostrukturen

Forfatteren har valgt en mikrostruktur der han følger prinsippet: oppslagsord – ekvivalent /evt. delekvivalent – grammatiske opplysninger – setningseksempler /kollokasjoner og kryssreferanser (jfr. brukerveiledningen på s. 9 og 10).

¹¹ Selv om det f.eks. i et dokument om byggingstillatelse kan tenkes at *Abbau* kan dukke opp, kvalifiserer dette etter min mening ikke til å bli et juridisk faguttrykk. Slår en opp i en alminnelig tysk-norsk (Cappelens tysk-norsk og norsk-tysk ordbok, Oslo 1992) eller tysk-engelsk ordbok (Cassell's German-English, English-German dictionary, New York 1978) på dette ordet, vil en selvsagt finne det der.

Dersom oppslagsordet ser ut til å forekomme i en spesiell sammenheng, er dette tilkjennegitt med angivelse av et "anvendelsesområde". F.eks. følgende "anvendelsesområder" brukes: "Waren, Reisen, Bauwerk, Warenabsatz, Vertrag", men også "EU". Fra et brukerståsted hadde det vært mere ønskelig med å relatere oppslagsordene til de forskjellige rettsområder som det er vanlig å arbeide med. F.eks. kunne *Abbestellungsversicherung* (Reisen) (s. 52) ha vært markert med (Versicherungsrecht), *Abhebungsformular* (Bank) (s. 57) med (Bankrecht) og *Abnahmeprotokoll* (Bauwerk) (s. 60) med (Baurecht).

Det savnes dessuten en redegjørelse for prinsippet for bruken av "anvendelsesområder" og helst også en oversikt/liste over dem. "EU" er riktignok forklart på s.16 som inneholder en oversikt over brukte forkortelser og symboler.

Når det gjelder eksempelsetninger /kollokasjoner som brukes for å belyse det aktuelle oppslagsordet, er denne delen meget fyldig og allsidig. En kunne derimot så absolutt ønske seg en redegjørelse fra forfatteren i innledningen /brukerveiledningen fra hvilket kildematerialet han har ekserpert eksempelsetningene /kollokasjonene. Er alle autentiske?

Det kunne f.eks. være greitt å få vite hva som er kilden til *einheitliche Bedingungen* 'standardbetingelser, standardvilkår'.

Ut fra prinsippet om at oppslagsordene er ordnet strengt alfabetisk er det et stadig tilbakevendende problem hva som skal tjene som oppslagsord ved kollokasjoner. Vil brukeren lete under *abhelfen* når han skal finne betydningen av *einer Beschwerde abhelfen* eller ser han under *Beschwerde* når han leter etter betydningen av *Beschwerde einlegen*? Forfatteren har løst problemet med å anføre kollokasjonene både under substantivet som oppslagsord (f.eks. *Beschwerde*), men også under verbet (f.eks. *einlegen*). Her finner en gjengitt de forskjellige kollokasjonene der verbet (*einlegen*) kan dukke opp, heriblant i forbindelse med forskjellige rettsmidler, altså også sammen med *Beschwerde*.

Som nevnt ovenfor har forfatteren utvidet denne 3. utgaven "betydelig". Mine innvendinger mot en del oppslagsord som ikke juridiske er anført ovenfor. La oss til slutt se litt på de juridiske oppslagsord som eventuelt måtte mangle.

Det er klart at det ikke er mulig for noen forfatter å lage en ordbok hvor det ikke vil være ord en savner. Derfor taes det da også forbehold om at ordtilfanget ikke er ment å være uttømmende (se forord til 1. utgave). Men det kan allikevel være på sin plass å se etter noen eksempler som man kunne forvente å finne. Dette gjelder spesielt oppslagsord der forfatteren allerede er inne på et gitt delområde. Således finner en ikke *Disziplinarstrafe*, men riktignok *Disziplinar-*

maßnahme, ikke Freistellung vom Militärdienst aus Überzeugungsgründen, men Freistellung von einer Pflicht, ikke Trunkenheitsdelikt, men Trunkenheit am Steuer, ikke Überstellung [von einem Täter]¹², men überstellen 'henvise'. Hvorfor er formelles og materielles Strafrecht ikke tatt med som egne oppslagsord? Eller Wehrdienstverweigerung, Drogendelikt og Insiderhandel? Hvorfor finnes ikke Transformation som oppslagsord når forfatteren har en del EU-rettrelaterte oppslagsord med? Henger det sammen med kildematerialet for oppslagsordene? Eller anses noen av disse stort sett kun å forekomme i den juridiske teorien?

Konklusjon

Ovenfor anførte kritiske bemerkninger er gjort i håp om at forfatteren ved neste utgave vil ta hensyn til noen av dem. Dette vil så føre til en forbedring av ordboken i retning av en **fag**ordbok i mer snever forstand. Det forutsetter som påpekt at oppslagsord som ikke hører direkte til det juridiske fagspråk lukes vekk, og at det foretas en klarere/bedre tilordning av oppslagsordene etter et gitt klassifikasjonsskjema. I lys av ordbokens andre og positive sider kan dagens utgave likevel trygt anbefales, bare brukeren er oppmerksom på de svakheter som er påpekt ovenfor.

Litteratur

- BGB = *Bürgerliches Gesetzbuch*³¹, Stand 15.02.1976, München o.J.
 Duden 1995 = *Duden – Das große Wörterbuch der deutschen Sprache in acht Bänden, 2.*, völlig neubearbeitete und stark erweiterte Aufgabe. Mannheim/Leipzig/Wien/Zürich 1995.
 GVG = *Gerichtsverfassungsgesetz*, Stand 10.02.1993. In: Schönfelder: "Deutsche Gesetze"⁷⁹. München 1993.
 StPO = *Strafprozeßordnung*, Stand 10.02.1993. In: Schönfelder: "Deutsche Gesetze"⁷⁹. München 1993.

¹² Riktignok er *Auslieferung* med i betydningen 'utlevering'.

Marketta Sundman

Ilse Cantell, Nina Martola, Birgitta Romppanen, Mats-Peter Sundström: *Suomi-ruotsi opiskelusanakirja* ('Finsk-svensk studieordbok'). Helsingfors: WSOY 1995

Anja Sarantola, Tauno Sarantola: *Ruotsi-suomi opiskelusanakirja* ('Svensk-finsk studieordbok'). Helsingfors: WSOY 1995

The article evaluates the two dictionaries from the perspective of the potential user. Both dictionaries are intended for Finnish speaking users such as school children and students. The vocabulary is modern and covers general usage. A number of colloquial and even slang words are included. Instead, transparent and infrequent compound and derivative words are not listed. The dictionaries are easy to use. The typography is very clear, and every word is given a new line.

A particular strength of the dictionaries is extensive presentation of both typical contexts and idiomatic phrases. Particularly the Finnish-Swedish dictionary could include even more information about the grammatical behaviour of verbs and pronouns.

Allmänt

I serien studieordböcker som ges ut av förlaget WSOY i Finland utkom år 1995 både en finsk-svensk och en svensk-finsk ordbok. Böckerna har samma format och blågula färg, dock så att pärmen på den finsk-svenska ordboken domineras av blått och den svensk-finska av gult, som sig bör. Den förra har en omfattning på 930 sidor och sägs innehålla ca 46.000 uppslagsord och uttryck, den senare är på 939 sidor och ca 62.000 ord och uttryck. Den finsk-svenska ordboken har utarbetats av en arbetsgrupp bestående av Ilse Cantell, Nina Martola, Birgitta Romppanen och Mats-Peter Sundström (den kallas nedan "Cantell m.fl."). Den svensk-finska ordboken har utarbetats av Anja och Tauno Sarantola (boken kallas nedan "Sarantola"). Arbetsgrupperna har arbetat oberoende av varandra.

Båda ordböckerna är i första hand riktade till personer med finska som modersmål. Cantell m.fl. är således en produktionsordbok som är avsedd för dem som behöver information om hur svenska ord och uttryck används, böjs, stavas och uttalas. På motsvarande sätt är Sarantolas ordbok främst en receptionsordbok som ger information om vad svenska ord betyder. Den ger t.ex. inga upplysningar om hur de finska ekvivalenterna till de svenska uppslagsorden böjs eller

konstrueras. Däremot ges vid behov uppgifter om de svenska uppslagsordens böjning. Båda ordböckerna presenteras på baksidan enbart på finska, likaså står böckernas namn bara på finska på titelbladet. Däremot ingår ett förord och anvisningar i båda böckerna på både finska och svenska.

Det är naturligt att anta att den tänkta användargruppen är ungefär densamma för de båda böckerna; de bildar ju ett par och ser ut att vara avsedda att användas tillsammans. Cantell m.fl. sägs vara en "allmänspråklig ordbok" som vänder sig till personer i olika åldrar med varierande kunskaper i svenska. Boken antas uppenbarligen lämpa sig för bl.a. skolelever och studerande på högre nivåer. Sarantolas ordbok sägs vara avsedd för språkstudier och yrkesmänniskor men också för andra som är intresserade av svenska.

Det ges mycket lite information om böckernas tillkomst. Författarna till den finsk-svenska ordboken har arbetat på Forskningscentralen för de inhemska språken i Helsingfors, och boken har uppenbarligen delvis kommit till parallellt med den nya finsk-svenska storordboken som har utarbetats av samma redaktion och som beräknas utkomma hösten 1996. Vilka andra ordböcker Cantell m.fl. eventuellt bygger på sägs det ingenting om. Sarantolas ordbok sägs ha kommit till under åren 1991–1994 och bygga på Stora svensk-engelska ordboken (1988).

I denna artikel granskas och diskuteras båda ordböckerna främst med avseende på hur väl de motsvarar de förväntningar och behov som de avsedda målgrupperna kan antas ha. Ordböckerna har tidigare recenserats i *Språkbruk*, av Ingegerd Nyström (nr 2/1995) resp. Nina Martola (nr 4/1995).

Ordartiklarnas struktur

Innehållsligt och typografiskt är artiklarna uppställda på nästan identiskt sätt i de båda ordböckerna. I den finsk-svenska ordboken ges uppslagsordens svenska ekvivalenter grupperade enligt olika betydelser. För ekvivalenterna anges böjning (främst substantiv och verb), och vid behov ges anvisningar om uttal och avstavning. T.ex. i fråga om verb anges oftast konstruktionssätt. Ekvivalenternas betydelse och användning i olika kontexter belyses i många fall med förklaringar men också med exempel. Dessutom upptas vanliga uttryck och fraser. Uppgifterna om stavning, böjning och uttal sägs bygga på Svensk skolordlista och Svenska Akademiens ordlista. Uttalet följer den sverigesvenska normen (se nedan).

I den svensk-finska ordboken ges uppgifter om det svenska uppslagsordets böjning och vid behov även om betoning och uttal. De finska ekvivalenterna är grupperade enligt olika betydelse, och dessutom ges frasexempel på samma sätt som i systerordboken.

Typografiskt gör båda ordböckerna ett klart och behagligt intryck. Uppslagsorden är tryckta med fet stil och fraser och idiom på källspråket med fet kursiv vilket gör att de är lätta att hitta i artiklarna. Uppgifter om rektion samt betydelseförklaringar är tryckta med kursiv. Ord- och exempelöversättningarna är tryckta med normal stil. Ordartiklarnas struktur och de olika specialtecknens funktion förklaras på ett mycket åskådligt sätt i Cantell m.fl. (efter förordet), medan motsvarande bruksanvisning saknas i Sarantolas ordbok. Samma principer följs dock också där.

Något som betydligt underlättar användningen är att alla ord, också sammansättningar och avledningar, utgör egna uppslagsord som står på en egen rad. Också vanliga geografiska namn ingår som uppslagsord; om dessa anges i Cantell m.fl. uttal och genus. Naturligtvis har bara ett rätt litet urval av geografiska namn kunnat tas med; bland namn som saknas i båda böckerna är t.ex. *Peru*, som stavas lika men som uttalas olika i svenska och finska. I Sarantolas ordbok hittar man också förkortningar som uppslagsord, placerade bland andra ord i den alfabetiska ordningen (t.ex. *m.fl.* efter *mexikan*). I Cantell m.fl. ges förkortningarna inte på samma sätt, utan *j.n.e./o.s.v.* hittas i ordartikeln *ja* 'och', medan *y.m.* överhuvud inte ges (*ynnä muuta* hittas under *ynnä*, och som ekvivalent ges *med mera, etcetera* men inga förkortningar; *med flera* ges inte alls).

Uttal

I fråga om uttal har man gått in för en lösning vars lämplighet kan diskuteras. I båda ordböckerna anges att uttalet följer ett sverigesvenskt ("rikssvenskt"), (i Cantell m.fl.) "närmare bestämt mellansvenskt" mönster. Med tanke på de många finska användarna i Sverige är detta säkerligen lyckat, men de flesta finnar lär sig faktiskt svenska i Finland och tillägnar sig därmed oftast det finlandssvenska uttalet. (Detta uttal är också betydligt lättare för dem att lära sig än det sverigesvenska och har visat sig fungera bra också i Sverige och i nordiska sammanhang.)

Principen om rikssvenskt uttal följs dock bara delvis: den s.k. tonaccenten, d.v.s. distinktionen mellan grav och akut accent anges inte. Man kan också enligt ordboken uttala t.ex. sje- och tje-ljudet antingen på sverigesvenskt eller finlandssvenskt vis. Det sverigesvenska uttals-

mönstret gör sig klarast gällande i Cantell m.fl. då det gäller sammansatta ord och vissa avledningar, där tryckfördelningen är en annan än i finlandssvenska. Ordet *avfallsanläggning* får därmed huvudtryck på *av* och bitryck på *lägg-* enligt sverigesvenskt mönster. I finlandssvenskan uttalas däremot *anläggning* med (nästan) samma tryckfördelning både som efterled i en sammansättning och som ett enskilt ord; sammansättningen får alltså bitryck på *an-*. Med tanke på de ordboksanvändare som för övrigt talar svenska med finlandssvenskt uttal (vilket de flesta finner alltså torde göra) kan detta verka förvirrande och leda till ett uttal som varken är korrekt i Sverige eller i Finland.

I Sarantolas ordbok markeras inte tryckfördelningen i sammansättningar. Denna lösning kunde enligt min mening också ha valts i den finsk-svenska ordboken, varvid allmänna riktlinjer för betoningen av sammansatta ord kunde ha givits bland de allmänna anvisningarna i början av boken. Sammansättningar med avvikande uttal tas inte heller upp i Svensk uttalsordlista (utg. av Forskningscentralen för de inhemska språken); skillnaden mellan uttalet i Sverige och Finland noteras dock där i inledningen (s. 9).

Uttalsangivelserna ges inom klamrar och i dem används både fonetiska tecken och normala grafem. Båda ordböckerna har i stort sett samma notationssystem, även om Cantell m.fl. använder fler fonetiska tecken. Till skillnad från bruket i svenska ordböcker anges det svenska o-grafemets uttal med bokstäverna *u* (slutet uttal, t.ex. i *fot*) resp. *å* (öppet uttal, t.ex. i *sova*). Relationen mellan grafem och ljud följer här finsk ortografi, naturligtvis med den reservationen att grafemet *å* inte förekommer i finska ord även om det ingår i alfabetet och därmed är känt för finnar. Det svenska u-ljudet (som i *ful*), som saknar motsvarighet i finskan, anges med ett fonetiskt tecken [Á]. Systemet kan vara något förvirrande för finnar som använder både finska och svenska ordböcker. I t.ex. SAOL markeras ju slutet o-ljud med grafemet *o* (t.ex. [fo:t]), medan grafemet *u* betecknar svenskt u-ljud (t.ex. [fu:l]). Den betonade stavelsen i ordet *velour* får därmed olika uttalsbeteckning i olika böcker: enl. SAOL uttalas den [-or], enligt Sarantolas ordbok [-ur]; dock avses samma uttal. I Svensk uttalsordlista betecknar bostaven *u* likaså u-ljudet (som i *ful*), medan det slutna o-ljudet anges med ett fonetiskt tecken. Ordet *mor* får därmed samma uttalsbeteckning i Sarantolas ordbok som ordet *mur* skulle få i Svensk uttalsordlista, om det skulle tas upp där, d.v.s. [mur]. Det öppna å-ljudet tycks generellt betecknas med grafemet *å* i olika ordböcker; så i SAOL, i Svensk uttalsordlista och i de båda finländska ordböckerna som behandlas här. Ordet *kod* får således uttalsbeteckningen [kåd] eller [kå:d] i alla fyra ordböckerna.

Något som jag saknar i båda ordböckerna är uppgifter om uttal av bokstäverna *e* och *ä*. I finskan betecknar dessa grafem olika fonem och de uttalas alltid olika, *e* ungefär som i svenskans *väst* och *ä* ungefär som i *verk/värk*. Det hade varit skäl att ge principerna för det svenska uttalet i samband med de allmänna anvisningarna i början av boken. För finnar är det inte självklart om t.ex. *peruk*, *person* och *peruan* uttalas med [e] eller [ä]. I Cantell m.fl. ges uttalsangivelser för enskilda ord oftare än i Sarantola; t.ex. för orden *persika*, och *persiska* ges uttalet [pä-], för *vers* [vä-] och för *ters* [tä-]. Det finska ortnamnet *Pernå* får i Cantell m.fl. uttalsangivelsen [pä-], medan man i Sarantolas ordbok valt att ange vokallängden i sista stavelsen: [-å], vilket är fel. Namnet uttalas med betoningen på första stavelsen. Uttalet av grafemet *e* i verbet *erbjuda* med antingen [e] eller [ä] ges i båda ordböckerna. Likaså ges uttalet för adjektivet *modern*, kanske främst för betoningens skull, men det är också viktigt att det visas att ordet i svenskan uttalas med [ä] och inte med [e] som i finskan. Andra ord som har finska motsvarigheter med e-uttal är t.ex. *verb*, *version*, *periferi* för vilka uttal inte ges; här är uttalsfel vanliga hos finnar. I Sarantolas ordbok används bara grafemen *e* och *ä* i uttalsangivelserna, medan Cantell m.fl. också använder två fonetiska tecken, nämligen [E] och [β]. Jag har svårt att se någon princip för i vilka fall e-ljud betecknas med [E]. Någon skillnad mellan två e-ljud i t.ex. *mäta* och *meta* görs däremot inte. Det kan ännu noteras att pluralformen *kanslerer* får olika uttal i de två ordböckerna: Cantell m.fl. ger [kansler_β], medan Sarantola ger [-är-].

Uttalsangivelser ges endast vid behov, och de är – som t.ex. i SAOL – fragmentariska, d.v.s. gäller enbart för de ljud eller ljudsekvenser som antas vara problematiska. Huvudprincipen är att uttal som framgår av stavningen inte ges. T.ex. anges i princip inte "mjukt uttal" av *g* och *k* framför betonad främre vokal, medan undantagen från denna princip anges (t.ex. *kö*). Systemet följs dock inte helt konsekvent. I Sarantolas ordbok anges tje-uttal i flera fall, uppenbarligen främst i fråga om främmande ord där finskan har ett motsvarande ord som uttalas med [k-]. T.ex. anges tje-uttal i fråga om orden *kikare*, *kinin*, *kinuski*, *kemi*, *keramik*, däremot inte i fråga om t.ex. *kika*, *kela*, inte heller *kejsare*. Varför tje-uttal också anges vid t.ex. *kök*, *kön* är dock svårt att veta. För användaren kan en sådan inkonsekvens verka förvirrande (i synnerhet om man inte läst inledningen där uttalsprinciperna förklaras). I Cantell m.fl. anges tje-uttal i obetonad stavelse, t.ex. i orden *keramik*, *kemi* och *kinuski*, men inte t.ex. i orden *kikare*, *kejsare*, *kök* och *kön*, som följer principen.

På samma sätt följer man huvudregeln att grafemet *o* uttalas som i *fot*, varför normalt bara å-uttal ges (därmed undviker man problemet

med u-grafemets ljudvärde som diskuterades ovan). Sålunda anges i Cantell m.fl. å-uttal för *blomsteraffär*, medan *blombukett* inte får någon uttalsangivelse. I synnerhet Cantell m.fl. är mycket restriktiv på denna punkt. Sålunda anges uttalet för *lov*, *loft*, *lock* [å] men inte för *lo*, *lod*, *lom*, *lots* [u]. Med tanke på att grafemet o står för ett öppet å-ljud i finskan, kunde en uttalsangivelse vara minst lika berättigad i de senare fallen som i de förra. Sarantolas ordbok anger vokalens uttal vid alla dessa ord, också för *blomma* och *blomster*, och t.ex. för båda uttalen av *lov* (i olika betydelser), d.v.s. både [låv] och [luv]. Jag föredrar denna lösning som inte heller upptar alltför mycket utrymme.

Betoning och vokallängd anges i de båda ordböckerna på delvis men inte helt identiskt sätt. I den finsk-svenska ordboken anges betonad stavelse med antingen punkt eller streck under vokalen, varvid punkt anger att vokalen uttalas kort, streck att den uttalas lång. Betoningen anges i själva ordöversättningen. I *markant* är den senare vokalen således markerad med punkt och i *marsvin* båda med streck; där får den senare stavelsen bitryck. I den svensk-finska ordboken anges betoning med punkt under vokalen. Vokallängd markeras inte utan den skall härledas av stavningen: en betonad vokal är alltså lång då den följs av en kort konsonant. Ifall en konsonant uttalas som lång trots enkel stavning anges detta vid behov (t.ex. *dum*) med dubbelteckning av konsonanten inom klamrar [-mm]. Denna markering saknas dock t.ex. i fråga om *kläm* och *lom*, där böjningsformerna stavas med dubbel konsonant, t.ex. *klämnen*. (Dessutom saknas den vid pronomina *han* och *hon*.) I uttalsangivelserna som ges inom klamrar markeras lång vokal med ett streck under vokalen, som t.ex. i fråga om *klot*, *glob*, *knop*, *lov*, [u]. Systemet är som helhet betraktat rätt komplicerat och kan därför verka något förvirrande. För övrigt är markering av betoning med punkt eller streck **under** vokalen en bra lösning, eftersom den lämnar ordets skriftbild (nästan) intakt.

Betydelseangivelser

Det normala är ju att ett ord inte kan översättas till ett annat språk med ett och samma ord i ordets alla kontexter. De flesta uppslagsord får därför flera olika ordöversättningar. I en receptionsordbok kan man utgå från att användaren oftast kan välja rätt bland ekvivalenterna på sitt modersmål med hjälp av den aktuella kontexten. För den som använder en produktionsordbok är valet av rätt ekvivalent däremot ett problem. Redaktörerna för den finsk-svenska ordboken har satsat mycket på att hjälpa användarna på denna punkt. Ekvivalenterna place-

ras ofta in i frasexempel som visar ordens vanliga kontexter i svenskan. Betydelseskilnader mellan orden förklaras där det är möjligt med hjälp av finska synonymer. Ekvivalenterna till verbet *sopia* hänförs t.ex. till fem olika betydelsekategorier som anges med synonymerna *mahtua* (> *rymmas*), *olla sopiva* (> *passa*), *olla soveliasta* (> *passa*), *tehdä sopimus* (> *komma överens*), *tehdä sovinto* (> *förlikas*). Också beskrivningar av andra typer används. De olika funktionerna hos verbet *korjata* belyses t.ex. med hjälp av typiska objekt, som *vaatteita* ('kläder') (> *lappa*), *koetta* ('ett prov') (> *rätta*), *tekstiä myös* ('en text även') (> *korrigera*), *uudistaa esim. rakennusta* ('t.ex. en byggnad') (> *renovera*), *viljaa* ('säd') (> *skörda*). Ibland anges också det fackområde som begreppet tillhör; här används förkortningar som listas i en förteckning i början av boken. För substantivet *korko* ges sålunda tre olika betydelser, av vilka den första förklaras med *jalkineessa* ('på sko') (> *klack*), den andra med förkortningen *tal* ('ekonomi') (> *ränta*) och den tredje med förkortningen *kiel* ('lingvistik') (> *intonation* etc.).

I allmänhet är de olika betydelserna hos ekvivalenterna mycket väl beskrivna i den finsk-svenska ordboken och belyses på ett utmärkt sätt med exempel som representerar ordens typiska kontexter. I de fall då flera ekvivalenter ges hade jag önskat att de belysts med minst ett exempel. T.ex. ordet *kurkku* får ekvivalenterna *strupe* och *hals*, i denna ordning, men frasexemplen innehåller antingen *hals* eller något annat uttryck; när skall man alltså välja alternativet *strupe*?

I den svensk-finska ordboken är ekvivalenterna grupperade enligt betydelse på samma sätt som i Cantell m.fl., d.v.s. antingen åtskilda med semikolon (mindre betydelseskilnad) eller som numrerade moment. I en receptionsordbok behövs färre betydelseförklaringar; vid behov ges dock sådana inom parentes. Adjektivet *likbent* får t.ex. översättningen *tasakylkinen*, med tillägget *kolmio* ('triangel'). Det vanliga är dock att orden belyses med frasexempel som visar i vilka kontexter de olika betydelserna aktualiseras. Verbet *etablera*, som saknar direkt finsk motsvarighet, exemplifieras sålunda med bl.a. fraserna *etablera ett samarbete* och *företaget har etablerat sig i landet*. Jag ser det som mycket lovvärt att frasexempel används på detta sätt för att belysa ordens typiska användning. Men dessutom behövs de för att illustrera mera ovanliga betydelser som endast aktualiseras i begränsade kontexter. T.ex. ordet *melodi* sägs kunna användas bildligt och då motsvaras av *tyyli* ('stil'), och som exempel ges *det är min melodi*. (Men i vilken kontext ordet kan motsvaras av *punainen lanka*, d.v.s. 'röd tråd', framgår inte.)

För fackspråkliga ord och synonymer anges ämnesområdet med hjälp av en förkortning. För ordet *lagring* ges t.ex. ekvivalenterna

varastointi och *kerrostuminen*, *kerrostuma*, av vilka de senare markeras som geologiska termer. Också stilangivelser används, som i fråga om ordet *ministerium*, som förutom den normalspråkliga ekvivalenten *ministeriö* också får en ekvivalent som anges tillhöra äldre stil, nämligen *pappisvirka* ('prästämbete'). Angivelser om fackområde och stil ges i allmänhet mycket rikligt, också då ordet bara har en ekvivalent i finskan.

Också i Sarantolas ordbok hade jag gärna sett att varje ekvivalent hade belysts med minst ett exempel, naturligtvis med undantag av ord som har identisk distribution i svenskan och finskan (typ *bil/auto*). Ovanliga betydelser som inte åskådliggörs genom något frasexempel förefaller snarast förvirrande. Bland ekvivalenterna till verbet *lura* ges t.ex. *kyteä* (som i Cantell m.fl. översätts med 'pyra, glöda'), men inget exempel där ordet kan ha denna något osannolika betydelse. Det kunde också vara bra om exemplen gavs i samma ordning som ekvivalenterna, helst börjande med den vanligaste ekvivalenten. Det betyder t.ex. att det första frasexemplet i artikeln *lufta* borde ha varit *lufta kläder* och inte *lufta hunden*, som det nu är.

Val av uppslagsord

Antalet uppslagsord är något mindre i den finsk-svenska än i den svensk-finska ordboken. (Enligt min uppskattning som bygger på ett stickprov på 10 sidor innehåller Cantell m.fl. ca 39.000 uppslagsord och Sarantola drygt 45.000 uppslagsord.) En orsak till denna kvantitativa skillnad ligger i att den svensk-finska ordboken har en något tätare text med 48 rader per spalt, medan den finsk-svenska har fått plats för 45 rader per spalt. Men framför allt handlar det om böckernas olika funktion. En receptionsordbok bör ju gärna innehålla de ord som bokens användare någon gång kan möta och behöva förstå, också litterära och ålderdomliga ord. Ordförrådet blir därmed större än i en produktionsordbok där man bara behöver ge de ord som användaren kan behöva producera själv, i tal eller skrift. Många av de ord som förekommer som uppslagsord i Sarantolas ordbok saknas sålunda bland ekvivalenterna i Cantell m.fl. I den senare ges t.ex. för finskans *reppu* bara ordet *ryggsäck*, som är det normala uttrycket i svenskan i dag, medan Sarantolas ordbok också upptar *ränsel* som uppslagsord. I stället är ordartiklarna något längre i den finsk-svenska boken p.g.a. att man där velat ge rikligt med frasexempel för att belysa hur orden används. T.ex. upptar uppslagsordet *tand* bara 7 rader i den svensk-finska ordboken (med 4 idiom), medan det motsvarande finska ordet *hammas* i

den andra boken givit upphov till en artikel på 14 rader, med 10 idiom. De längsta ordartiklarna i Sarantolas ordbok handlar om verb där kombinationerna med olika partiklar tar mycket utrymme. Verbet *ställa* behandlas t.ex. på 67 rader där bl.a. kombinationer med 9 olika partiklar (från *av* till *ut*) tas upp; det motsvarande finska verbet *panna* har fått bara 20 rader i den andra boken.

I båda ordböckerna förekommer också geografiska namn samt motsvarande adjektiv och invånarbetesbeteckningar som uppslagsord, vilket är mycket positivt. Man får t.ex. veta att *EU* är neutrum (ett-genus). I Sarantolas ordbok saknas underligt nog *Svenskfinland* som visserligen inte betecknar ett geografiskt område men som är ett etablerat begrepp; ordet saknar en självklar finsk motsvarighet och borde därför förklaras. I den finsk-svenska ordboken ges *finsk* och *finländsk* som likvärdiga motsvarigheter till *suomalainen* (likaså *finne* och *finländare*). Ändå anser man ju i Finland att de två orden har olika innebörd och bör användas i olika kontexter. En utmärkt lösning i den finsk-svenska ordboken är att i varje ordartikel om ett geografiskt adjektiv ge ett exempel av typen *vi övernattade på ett hotell i Uppsala* för att visa att svenskan inte använder adjektiv (som finskan) utan prepositionsattribut i dessa fall.

Enligt texten på baksidan har målet för den finsk-svenska ordboken varit att inkludera ett modernt ordförråd där det även ingår talspråkliga uttryck och ord som tillhör olika specialområden (vilka dessa är meddelas inte). I jämförelse med den finsk-svenska storordboken av Cannelin, Hirvensalo och Hedlund framstår den nya finsk-svenska ordboken faktiskt som mycket modern. Något som saknas i jämförelse med den betydligt äldre storordboken är framför allt mer tillfälliga sammansättningar som vanligtvis omskrivs vid översättning till svenska. (T.ex. av sammansättningarna på adj. *laaja* 'vid' ingår i den nya ordboken bl.a. de mer lexikaliserade *laajakantoinen* 'långtgående', *laajakatseinen* 'vidsynt', och *laajapohjainen* 'bredbasig', medan t.ex. de mer sällsynta och genomskinliga *laajalatvuksinen* 'bredkronig' och *laajatietoinen* 'med vidsträckta kunskaper' inte tagits med.) Bland orden på bokstaven l saknas i den nya ordboken bl.a. *laamanni* 'lagman', *laatikoittain* 'lådvis', *laatoittaa* 'belägga med plattor', *laavu* 'skärmskydd', *lahjavero* 'gåvoskatt', alla ord som en finskspråkig någon gång kan behöva använda på svenska. Ett aktuellt ord som gärna kunde ha tagits med är *luovuttaja* i betydelsen 'donator (av ett organ)'.

Enligt min uppfattning har författarna till den finsk-svenska ordboken dock lyckats utmärkt väl med att åstadkomma ett urval av ord som är normalt förekommande i olika stilarter i svensk sakprosa. Såväl uppslagsorden som ekvivalenterna representerar ett modernt och fräscht

språk och en stilistisk variation som sträcker sig från en formell och något ålderdomlig stil till talspråk och t.o.m. slang. I överensstämmelse med frekvensförhållandena i t.ex. dagens tidningsprosa har de finska orden *isä* och *äiti* nu fått *pappa* resp. *mamma* som första ekvivalenter, medan orden *far* och *mor* kommer först på andra plats. Moderna ord av typ *faxa*, *bläckstråleskrivare*, *tåggluffa* ges. Exempel på vanliga talspråkliga ord och slangord som har inkluderats är t.ex. *euroviisu* ('eurovisionslåt'), *fyrkka* ('stålar'), *funtsata* ('fundera'), *jätkä* ('kille', 'snubbe'), *naida* ('knulla'). Urvalet talspråksord och i synnerhet slang är dock litet; ett ord som jag förgäves slog upp är *saundi/soundi*, som är vanligt också i rätt seriösa kontexter i fall av typen 'sound hos en orkester'. Av någon anledning ges *fatsi* ('farsa') men inte *mutsi* ('morsa').

Även om en stor mängd nya ord och begrepp finns med innebär detta inte att alla de ord och betydelser som tillhör det normala ordförrådet t.ex. i finska dagstidningar går att hitta i ordboken. Ett litet experiment med några artiklar i Helsingin Sanomat 22.3.1996 visar att texterna inte skulle gå att översätta bara med hjälp av denna ordbok. Ord som saknas är t.ex. *heimokansa* 'besläktat folk', *monipuolistaa* 'göra mångsidigare', *tähänastinen* 'hittillsvarande', *digitalisointi* 'digitalisering', *valtaenemmistö* 'överväldigande majoritet', *rintapieli* 'uppslag (på klädesplagg)', *osin* 'till en del', *pahimmillaan* 'som värst'. Inte heller kan man alltid hitta korrekta ekvivalenter för just de betydelser som orden har i de undersökta texterna. Exempel på moderna ord som man däremot hittar är *televiestintä* 'telekommunikation', *eturistiriita* 'intressekonflikt', *leikkaus* 'nedskärning', *tekoäly* 'artificiell intelligens', *uupumus* 'utbrändhet', *tyhjiöpakkaus* 'vakuumförpackning' *turvapaikan hakija* 'asylsökande' och *sädetys* 'bestrålning'.

Några sammansättningar som jag saknade är *runkopatja* ('resår-säng') och *pirttikalusto* som inte kan översättas ordagrant. Det sistnämnda betecknar en i Finland vanlig kombination av matbord och långbänkar som påminner om allmogestil. Någon svensk motsvarighet finns sannolikt inte. Orden saknas också i den finsk-svenska storordboken.

I den svensk-finska ordboken har målet på liknande sätt varit att inkludera ett modernt och centralt ordförråd. När man jämför uppslagsorden i Sarantolas ordbok med Stora svensk-finska ordboken (1984) noterar man att det främst är genomskinliga och ovanliga sammansättningar och en mängd specialterminologi som utelämnats. Av de ca 50 sammansättningarna på *levnad-* som ingår i Stora svensk-finska ordboken har i Sarantola medtagits 9 frekventa och tämligen lexikaliserade sammansättningar, bland dessa t.ex. *levnadskostnader* och *levnadsteckning*; bland orden som saknas finns t.ex. *levnadskonst-*

när och *levnadsår*. Dessutom har ord som tillhör mer speciella områden eller är sällsynta utelämnats, bl.a. *libation*, *libero*, *libhaber*, *librettist* och *librist*, men också *lexikon* och *lexikografi* (!).

Ett litet test på några artiklar i Svenska Dagbladet (1.4.1996) gav ett mycket positivt resultat: så gott som alla ord gick att hitta i Sarantolas ordbok. Till ord som saknades hörde *grundbult* och *vegan* och några genomskinliga och kanske mer tillfälliga sammansättningar som *byggkostnad*, *psykförsvar* och *själv tillit* som inte heller ingår i t.ex. SOB. Nya ord och begrepp finns i allmänhet mycket väl representerade. Bl.a. hittar man *okynneskörning*, *platsa*, *prao*, *rea*, *moms*, *rymdsond*, *slit-och-slängvara*, *sniffare* och *terminalvård*. Ordförrådet gör genomgående ett tidsenligt och rent av modernt intryck.

Sarantolas ordbok innehåller något överraskande ett antal ord som saknas i SOB. Sådana är t.ex. *luftbolag*, *lungpaj*, *luka* och *lusing*. Dessutom ges för vissa uppslagsord delbetydelser som inte nämns i SOB. Man kan fråga sig om så pass ovanliga ord och betydelser försvarar sin plats i en tvåspråkig studieordbok och om man inte i stället borde ha tagit med flera aktuella sammansättningar och specialord.

I en ordbok avsedd för finländska förhållanden aktualiseras frågan om i vilken utsträckning också finlandismer, d.v.s. svenska ord och uttryck som enbart används i Finland, skall tas med. Olika finlandismer har varierande spridning och stilnivå; en stor del av dem som räknas upp i t.ex. Stenmarks ordlista (Stenmark 1983) upplevs som provinsiella eller talspråkliga också i Finland. I båda ordböckerna som diskuteras här är ordförrådet naturligtvis allmänsvenskt, d.v.s. först och främst rikssvenskt. I fråga om finlandismer har man följt den enda riktiga principen att finlandismer som allmänt förekommer och godtas i skrift i Finland är medtagna, medan sådana som betraktas som provinsialismer eller rent av språkfel inte ingår vare sig i Cantell m.fl. eller i Sarantola. I en receptionsordbok kunde något fler finlandismer dock ha varit på sin plats, naturligtvis försedda med anmärkningen *vardagligt*, *finlandssvenskt* eller motsvarande. T.ex. uttrycket *på basen av något*, som är så vanligt i finlandssvensk sakprosa och bara delvis kan ersättas med det allmänsvenska *på basis av något*, kunde kanske ha medtagits i Sarantolas ordbok, liksom t.ex. uttrycket *det skulle inte slå hål i sidan*, vars innebörd inte kan härledas ur konstituenternas betydelse (ungef. 'det skulle inte skada'). Sarantola har visserligen som uppslagsord flera finlandismer som inte ges som ekvivalenter i den finsk-svenska ordboken, t.ex. *muntur* (fi. *suunvuoro*) (ung. 'tillfälle att yttra sig') och *kännspak* ('karaktéristisk').

Principen är i båda ordböckerna den att ord som är officiella begrepp i Finland är försedda med förkortningen *Suom* och deras sve-

rigesvenska motsvarigheter med *Ruots*. Andra finlandssvenska ord och uttryck har försetts med beteckningen *suomr* och deras sverigesvenska motsvarigheter med *ruotsr*. Ordböckerna stämmer inte helt överens i fråga om vilka ord som räknas till den ena resp. den andra kategorin. Begrepp av den första kategorin (*Suom*) i Cantell m.fl. är bl.a. *sinko* 'granatgevär', *kaplan* 'komminister', *skriftskola* 'konfirmationsundervisning'. I Sarantola har också *leverlåda* resp. *korvkaka* försetts med dessa beteckningar, medan de i Cantell m.fl. har beteckningen *suomr* resp. *ruotsr*. Ord av den senare typen är t.ex. *gårdskar/portvakt*, *sem/ småfranska*, *talko* (ung. 'arbetsgille'), *gulnäbb* ('recentior'). *Griskotlett* 'fläskkotlett' är markerat som finlandism bara i Cantell m.fl.

Enligt förordet till Cantell m.fl. innehåller ordboken sådan specialterminologi som har "ansetts ha allmänt intresse"; också Sarantola sägs innehålla specialterminologi. Det är självklart att egentliga facktermer inte kan ingå i så begränsade ordböcker som det här är fråga om. Ett stickprov på växt- och fågelarter visade att Sarantolas ordbok innehåller klart mer terminologi på dessa områden. Cantell m.fl. har inkluderat närmast sådana arter som är rätt välkända och t.ex. ofta förekommer i skönlitteraturen. Växtarter som upptas i båda böckerna är t.ex. liljekonvalj, vitsippa, prästkrage, klöver, svärds-lilja, gråbo, klibbal och säl; bland fåglarna nämns t.ex. ormvråk, skrake, nattskär, rödstjärt, rödhake och svartvit flugsnappare.

Ekvivalenterna

I likhet med uppslagsorden representerar ekvivalenterna i Cantell m.fl. ett modernt språk som står stilistiskt närmast en neutral sakprosa, även om också ekvivalenter som tillhör andra stilnivåer ges i lämpliga fall. Sålunda ges i artikeln *lähteä* förutom verben *gå* och *ge sig av* (i väg) även det talspråkliga *sticka* (markerat *puhek*, d.v.s. 'talspråk') och i artikeln *puhua* förutom verben *tala* och *prata* även *snacka* (*puhek*, vars *ruotsr*, d.v.s. 'talspråkligt, i synnerhet sverigesvenskt'). För *paljon* ges på samma sätt både *mycket* och *en massa*, för *toveri* förutom *kamrat*, *kompis* även *polare*, *kumpan* och i betydelsen 'bästa kamrater' även *bästis*. Det är mycket positivt att ordförrådet inte har begränsats till det allra neutralaste, samtidigt som det naturligtvis är viktigt att markera de ord som inte passar i vilken kontext som helst.

I en produktionsordbok är det inte väsentligt att ange alla de ord och uttryck som uppslagsordet i något sammanhang kan översättas med. I stället skall man ge de ord som oftast motsvarar uppslagsordet och som tillsammans täcker de flesta fall där ordet kan förekomma. Författarna

till Cantell m.fl. har enligt min uppfattning lyckats utmärkt väl med detta. Otidsenliga och sällsynta ekvivalenter har sålunda gallrats bort och ersatts med moderna och frekventa ord. Ordningföljden mellan ekvivalenterna motsvarar enligt min uppfattning i de flesta fall dessas sannolika frekvens som motsvarigheter till de svenska orden. T.ex. ordet *laajuus* får som första ekvivalenter orden *omfattning* och *omfång*, vilka är de ord som används i de högst vanliga uttryck av typen *verksamhetens omfattning*. (Som jämförelse kan nämnas att Cannelin m.fl. i stället ger orden *vidd* och *vidlyftighet* som första ekvivalenter och att t.ex. orden *volym* och *amplitud* inte överhuvud ges som översättningar till *laajuus* där.) I artikeln *liikunta* skulle jag dock ha väntat mig *motion* före *rörelse*; min gissning (som jag inte kunnat kontrollera) är att ordet normalt används i den förra betydelsen. I synnerhet när en av ekvivalenterna är en ordagrann motsvarighet till uppslagsordet (vilket *rörelse* är i detta fall) kan det vara skäl att vara försiktig med att ange den först. I artikeln *olla* ('vara' etc.) kunde *finnas* gärna ha fått stå före *existera* och *förekomma* som är betydligt mindre frekventa som motsvarigheter till detta verb.

I artikeln *näkökulma* hade jag väntat mig att *perspektiv* hade angetts som ekvivalent utöver *synvinkel*, även om båda sannolikt kan användas i identiska kontexter. I artikeln *liittää* (bl.a. 'foga, förena') saknar jag verben *ansluta* och *anknyta*; det förra behövs i kontexter av typen *ansluta en telefon till nätet*. I artikeln *liittyä* (reflexiv motsvarighet till föregående) ges bara formen *anknyta sig* även om verbet normalt används utan reflexivt pronomen också i intransitiv funktion; detta framgår också av Sarantolas bok.

En viktig ekvivalent saknas enligt min mening i artikeln om adverbet *kuitenkin/kuitenkaan*, som närmast har en textuell funktion. Ett vanligt fel hos finskspråkiga är nämligen att översätta ordet med *ändå* i alla kontexter, vilket mycket ofta är oriktigt. (T.ex. i nästa mening kan *ändå* inte ersätta *dock*, som i detta fall skulle motsvaras av *kuitenkin* i finskan.) I ordboken ges dock (!) bara orden/uttrycken *ändå*, *likväl*, *trots* allt och *i alla fall* som ekvivalenter till *kuitenkin*. Ordet *likväl* kommer semantiskt närmast men är rätt ålderdomligt och därför enligt min mening stilistiskt mindre lämpligt än *dock*. Som ekvivalent kunde man också ge konjunktionen *men*, som i alla dessa kontexter kan ha samma textuella funktion som adverbet *dock*. Här hade det också varit viktigt att ge riktigt med exempel som belyser skillnaden mellan de olika kontexterna.

Också i Sarantolas ordbok är ordöversättningarna genomgående väl valda och idiomatiska och representerar ett rikt och levande språk. *Luns* översätts t.ex. med *köntys*, *lunsig* med *lötköttävä*, *lunka* med *löntystellä*.

Direkta fel bland ekvivalenterna är vad jag kan se mycket sällsynta. Däremot förekommer som ovan nämndes delbetydelser som i varje fall inte får stöd i SOB. Till adjektivet *lufsig* ges också ekvivalenten *rähjäinen* (närmast 'ovårdad'), som saknas i SOB. Liknande fall är *ilmanvaihto* ('ventilation') för *luftombyte* och *salakuljettaja* ('smugglare') för *lurendrejare* (alla dessa ges också i Stora svensk-finska ordboken). Även om betydelsena eventuellt kan förekomma eller har förekommit, borde de enligt min mening ha strukits i en ordbok som vill presentera ett modernt och centralt ordförråd. Ordet *huvipuisto* betyder numera främst 'nöjesfält' (så t.ex. Cantell m.fl.), varför översättningen är missvisande för ordet *lustgård*. Det skulle åtminstone ha krävt en kompletterande förklaring. Uttrycket *köra på skongång* (i fråga om tvättmaskin) har översatts med *pestä säästöohjelmalla* som enligt min mening är inadekvat eftersom tyngdpunkten då ligger på att spara (vatten/energi). Enligt Svensk ordbok är det fråga om 'maskintvätt som är skonsam mot plaggen'. Rätt finsk motsvarighet till ordet *skongång* kunde ha varit *arkojen tekstiilien pesuohjelma*.

Ibland saknar man en motsvarighet som hade varit minst lika berättigad att tas med som de som ges. T.ex. ordet *lucka* har fått många ekvivalenter, dock inte *aukko* som aktualiseras t.ex. i kontexten *en lucka i resonemanget* och i *lucktest* (som saknas). Verbet *lukta* kan också översättas med *haiskahtaa*, t.ex. i kontexter av typen *det luktar bedrägeri* (som ges i Cantell m.fl. som ett exempel på *haiskahtaa*). Ofta ges dock i Sarantolas ordbok snarare onödigt många än för få ekvivalenter. T.ex. verbet *lura* (se ovan) får ca 15 olika ordöversättningar. Grundbetydelsena kan ju oftast belysas med några få ord, och alla översättningar som någon gång kan aktualiseras kan ändå inte ges.

Överhuvudtaget är urvalet av ordöversättningar i båda ordböckerna enligt min mening adekvat och skickligt gjort. Det är skäl att tala om ett "urval" eftersom det är omöjligt att ta med alla synonymmer som orden kan få i någon kontext eller stilart. Den som behöver mer sofistikerad information om svenska resp. finska ord kan ju gå vidare till enspråkiga ordböcker eller rent av slå upp ord i olika konkordanser som numera också finns i datoriserad form.

Fraser och idiom

Till båda ordböckernas starka sidor hör det rikliga urvalet av exempel, uttryck och idiom. Dessa representerar olika stilnivåer och genomgående ett fräscht och levande språk. Den som lär sig svenska som främmande språk kan faktiskt genom att bläddra i ordböckerna lära sig

mycket om svenskans fraseologi. En principiell fråga som gäller urvalet av frasexempel är om dessa snarast skall ge en uppfattning om ordens typiska användningssätt eller tvärtom utgöras av speciella idiom och talesätt som avviker från ordens normala bruk. Den första typen av exempel är ägnade att komplettera de anvisningar om ordens användning som ofta ges i formen av synonymer eller andra förklaringar. De hjälper också till att ge en exaktare uppfattning om ordets betydelse i svenskan.

Jag ser det som mycket positivt att man i de båda böckerna velat ge riktigt med exempel på ordens typiska användningssätt. I detta avseende är Cantell m.fl. mycket mer användbar än den äldre finsk-svenska storordboken där frasexemplen är rätt få. En annan, lika viktig funktion hos frasexemplen är att presentera uttryckssätt där den normala ekvivalenten inte används. I artikeln *pää* 'huvud' ges sålunda ett antal uttryck med ordet *huvud*, t.ex. *hålla huvudet kallt*, *pojken har huvudet på skaft*, *framgången steg honom åt huvudet*, men också uttryck med andra ord, t.ex. *på dåligt humör* (fi. *huonolla päällä*), *på egen hand* (fi. *omin päin*), *raka vägen* (fi. *suoraa päätä*), *med full fart* (f. *täyttä päätä*) och *håll käften!* (fi. *pää kiinni!*). Det viktiga är naturligtvis att frekventa uttryck finns medtagna, som t.ex. praktiskt taget (fi. *käytännöllisesti katsoen*) och efter omständigheterna (*olosuhteisiin katsoen*) som båda hittas i artikeln *katsoa* ('se, titta'). De svenska fraserna ser ut att vara övertagna från svenska ordböcker; t.ex. Svensk ordbok ger i många fall (nästan) identiska uttryck.

I Cantell m.fl. ges också ett stort antal vanliga finska idiom och deras svenska motsvarigheter. Det är också viktigt att ge idiom där språken motsvarar varandra ordagrant, det är ju inte sagt att man annars vågar översätta direkt. Sålunda ges t.ex. i artikeln *laakeri* bl.a. *skörda lagrar* och *vila på sina lagrar* som båda har (nästan) exakta finska motsvarigheter. Några vanliga finska idiom som jag förgäves slog upp i den finsk-svenska ordboken är uttrycket *ei ole minun heiniäni* (ungefär 'det är inte min sak') som inte kommer upp under *heinä* ('hö') och *jutussa ei ole päätä eikä häntää* ('det är helt huvudlöst').

I Sarantolas ordbok presenteras på motsvarande sätt en rik svensk idiomatik genom att dels exempel på ordens typiska kontexter, dels idiomatiska uttryck av olika slag ges. För görlig ges sålunda *i görligaste mån*, för gyllene bl.a. *den gyllene medelvägen* och för grepp t.ex. *han fick inget grepp om det*, *tappa greppet* och *falla på eget grepp*. Översättningarna är genomgående träffande och idiomatiska, t.ex. översätts det sistnämnda uttrycket *sortua omaan näppäryhteensä*. Ofta framgår betydelsen bäst av exempel, som i fråga om ordet *gänga* som belyses med exemplen *allt går i de gamla gängorna* och *vara ur*

gångorna. En viktig funktion hos frasexemplen är att illustrera uttryck och idiom där uppslagsordet måste översättas med något annat ord än de vanliga ekvivalenterna. T.ex. i fråga om ordet *luft* ges en rad fras-exempel där den normala finska ekvivalenten *ilma* inte kommer i fråga: *ett slag i luften, leva på luft, ligga i luften, gripet ur luften*. Fraserna får utmärkta, idiomatiska översättningar.

I svenskan utgör verbfraser bestående av verb och partikel (t.ex. *bryta av*) en viktig kategori som saknar motsvarighet i finska. I Sarantolas ordbok tas de lösa förbindelserna upp under ett och samma moment i verbartikeln, där varje partikel är placerad på egen rad, i alfabetisk ordning. Motsvarande fasta sammansättningar (typen *av-bryta*) ingår ofta som självständiga uppslagsord, åtminstone om de semantiskt skiljer sig från den lösa förbindelsen. Sålunda ges både *av-bryta* och *bryta av* men på olika ställen; likaså finns både *fastslå* och *slå fast, framträda* och *träda fram*, o.s.v., där betydelseangivelsen är densamma. Å andra sidan ges inte *framlägga* utan bara *lägga fram*. Nina Martola har i sin recension påpekat att stilistiska och semantiska skillnader mellan fasta och motsvarande lösa partikelkombinationer inte framgår tillräckligt tydligt av Sarantolas ordbok.

En liten jämförelse med Erkki Karis svensk-finska idiomlista (Kari 1993) visade att urvalet av idiom i Sarantolas ordbok mycket väl täcker de centralaste uttrycken i svenskan, i både tal och skrift. T.ex. *skära alla över en kam, vara som klippt och skuren, bli varm i kläderna, när allt kommer omkring, börja på ny kula, vara i valet och kvalet* hittas alla där. Några idiom som också kunde ha tagits med är t.ex. *ta något med en klackspark* och *gå från klarhet till klarhet* som inte kan förstås utifrån de ingående orden.

Grammatisk information i den finsk-svenska ordboken

Den som skall tala eller skriva ett främmande språk klarar sig inte bara med ord utan behöver också veta hur dessa konstrueras, d.v.s. hur de kombineras med andra ord och former. Den information som ges i grammatikor är inte tillräcklig, eftersom dessa bara kan ta upp generella principer som gäller för stora grupper (kategorier) av ord. Många grammatiska fenomen är begränsade till mindre grupper, och i grammatikor ges sällan listor på ord som betar sig på ett bestämt sätt. Enskilda ords konstruktionssätt kan i stället tas upp i lexika. För svenskans del är Svensk handordbok en god källa, men den är främst avsedd för personer med svenska som modersmål och kan därmed inte alltid hjälpa dem som använder svenska som främmande språk. Därför är en

tvåspråkig produktionsordbok den rätta källan för information av detta slag.

Vilken grammatisk information är det som en finskspråkig kanske vill hitta i en finsk-svensk ordbok? Det man troligtvis oftast söker är upplysningar om verbens rektion samt om prepositionsval. I de flesta fall hittar man dessa uppgifter i Cantell m.fl., men ibland skulle jag önska ännu mer information. I artikeln om verbet *lukea* 'läsa' ges konstruktionssätten *läsa en bön* och *läsa ur en bok*, men typen *läsa i en tidning att...* (som finskans *lukea lehdestä että...* oftast motsvaras av) saknas. I artikeln *nähdä* 'se' får man inte heller veta hur finska uttryck av typen *nähdä televisiosta/lehdestä* skall översättas till svenska. En besvärlig kategori för dem som inte har svenska som modersmål är indirekt objekt och motsvarande prepositionsfraser. I artikeln om *näyttää* 'visa' framgår att verbet konstrueras antingen *visa någon något* eller *visa något för någon*. Däremot får man inte veta hur den mottagande personen uttrycks t.ex. vid verben *skriva* och *köpa*. Om verben *berätta* och *förklara* framgår det att de konstrueras med *för*, men om verben *säga* och *tala* konstrueras med *åt*, *för* eller *till* framgår inte.

Om frasexemplen i första hand tas över från enspråkiga svenska ordböcker finns det en risk att konstruktionssätt som är problematiska för finnar inte råkar ingå bland dem. Jag hade t.ex. gärna sett att prepositionsvalet i uttryck av typen *köpa i en affär*, som avviker från det finska konstruktionssättet där man har riktningskasus, hade framgått av något frasexempel i artikeln *ostaa* 'köpa'; ett vanligt fel är ju **köpa från en affär*. (Konstruktionen ges inte heller vare sig i SOB eller i Svensk handordbok.) I artikeln *jalka* 'fot' hade man gärna fått ge uttrycket *ha skor på fötterna* för att eliminera fel av typen **ha skor i fötterna* som är resultat av en direkt översättning av finskans inessiv. I artikeln *löytää* ('hitta') saknar jag exempel som visar att verbet inte konstrueras på samma sätt i svenskan som i finskan, d.v.s. med riktningsuttryck (**hitta ngt från ngt*). Vilka verb som kan konstrueras med objektiv predikativ och hur predikativet skall se ut kunde också belysas med exempel av typen *jag betraktar honom som min vän*.

Något man också gärna kunde visa systematiskt är att svenska verb inte lika allmänt konstrueras med objekt + infinitiv som finska verb. I artikeln *haluta* ges visserligen konstruktionen *jag ville att han skulle gå* som svensk motsvarighet till den finska konstruktionen *halusin hänen lähtevän*, men t.ex. i artikeln *huomata* får man inte veta att finskans *huomasin hänen lähtevän* inte kan översättas med **jag märkte honom gå*. Vilka verb som kan ta infinitiv som komplement (med eller utan *att*?) kunde likaså ha visats. Finskan har i många fall verbalsubstantiv där svenskan har infinitiv, t.ex. *hän suunnittelee/harkitsee matkusta-*

mista ('han planerar/överväger att resa'). Också konstruktionstypen *huset planeras att bli färdigt nästa år* kunde ha givits. Konstruktioner som motsvarar finskans passiv i fall där svenskan inte tillåter att verbet passiveras kunde ha getts i några centrala fall, t.ex. verbartiklarna *osata* ('kunna') *voida* ('kunna'), *haluta* ('önska'), *päätä* ('besluta'), *luvata* ('lova') t.ex. *talo voitiin rakentaa*, sv. *huset kunde byggas*, och *talo luvattiin rakentaa*, sv. *man lovade bygga huset*. Artikeln *tietää* ('veta') upptar nästan en hel spalt och innehåller ett stort antal idiomatiska uttryck men inga exempel på passiva satsar eller konstruktioner med objekt + infinitiv i finska.

En annan kategori som jag hade önskat mera grammatisk information om är pronomen. I några fall saknas uppgifter om pronomenets böjning, i andra fall om konstruktionssätt. I artikeln *kukin* ges *var och en* men inte *vart och ett*; också alla frasexempel innehåller utrumform. (Neutrumformen ges i stället i artikeln *jokainen*.) I artikeln *yksin* ges bara formen *ensam*; det ges flera frasexempel men alla har subjekt i utrum singularis. Eftersom ordet i finskan är oböjligt (adverb) är det naturligt för en finsk användare att tro att detta också gäller i svenskan, vilket leder till inkongruens av typen **de gick dit ensam*. Ett annat besvärligt pronomen är *själv*. I ordboken ges visserligen de tre formerna *själv*, *självt*, *själva* men inga frasexempel med de två senare formerna. Den finska ekvivalenten *itse* böjs inte i plural varför finner ofta låter bli att böja *själv* där svenskan kräver kongruens. I artikeln *itse* visas att ordet ofta motsvaras av ett reflexivt pronomen i svenskan (t.ex. *jag fick ingen att tro på mig*), men att också *själv* i många fall måste ingå visas inte tydligt (det heter ju t.ex. inte **han älskar bara sig*).

Finskans *eri* och *erilainen* ställer ofta till med problem för dem som skall översätta dem till svenska, och här kan inte heller den nya finsk-svenska ordboken erbjuda dem mycket hjälp. I artikeln *eri* ges *olik* som tredje ekvivalent, och *olika* ges enbart som adverb (t.ex. *olika stora*). Något exempel på *olika* som attribut i singularis ges inte; en möjlighet till detta hade funnits i frasexemplet *olla eri mieltä*, där man gärna kunde ha givit både *han var av annan åsikt* (i stället för *vara av annan åsikt* som nu ges) och *de var av olika åsikt* (med en förklaring av typen "sinsemellan"). Hur *olik* används framgår egentligen inte av denna artikel utan av artikeln *erilainen*, där man ger *helt olik den här*, d.v.s. konstruktion med objekt. Utgående från dessa exempel kan man lätt sluta sig till att **vi var av olik åsikt* är korrekt svenska. Valet mellan *olika* och *annorlunda* blir inte heller helt klart. Jag hade också önskat att adjektivets form i substantivfraser med t.ex. *varje*, *denna* och *samma* som attribut hade framgått av frasexemplen. Oklart blir också hur *båda*

konstrueras om substantivet också har adjektivattribut; kan det alltså t.ex. heta *båda svenska ordböckerna*?

Det kan ytterligare nämnas att finskans s.k. tilläggsartiklar, som visserligen är bundna morfem men har närmast pragmatisk funktion, inte kommer upp systematiskt i ordböckerna. I Sarantolas bok ges både *myös* och *-kin* som ekvivalenter till *också*, men i Cantell m.fl. ges inte *-kin* som uppslagsord. Adverbet *ju* får som ekvivalent i Sarantola partikeln *-han/-hän*, och dessa former ges också som uppslagsord i Cantell m.fl. Partikeln *-ko/-kö* som bildar frågesatser ges inte, inte heller *-pa/-pä*, som har pragmatisk funktion.

Slutord

Mitt slutomdöme om de båda ordböckerna är mycket positivt. Till formatet är de behändiga att använda – det är t.ex. inte omöjligt att bära dem med sig till och från arbetsplatsen, vilket undertecknad kan intyga med stöd av några veckors erfarenhet. De typografiska lösningarna gör att man lätt hittar den information man söker. Uppslagsorden täcker det centrala ordförrådet i modern svensk sakprosa, medan urvalet av specialtermer och talspråksord förefaller något mer slumpmässigt. Båda ordböckerna innehåller rikligt med frasexempel som belyser ordens normala användningssätt i svenskan men också deras förekomst i idiom av olika slag. Några allvarliga fel har jag inte råkat hitta, även om nyanserna ibland kan diskuteras. Däremot hade jag önskat mera upplysningar om ordens konstruktionssätt: i en svensk ordbok avsedd för finskspråkiga kunde frasexempel i större utsträckning användas för att också belysa ordens normala grammatiska konstruktion i svenskan.

Litteratur

Ordböcker

Cannelin, Knut, Aulis Cannelin, Lauri Hirvensalo, Nils Hedlund:
Finsk-svensk storordbok. 4:uppl. Helsingfors: WSOY 1976.

Kari, Erkki: *Svenska här och nu. Ruotsalais-suomalainen idiomisanakirja*. Helsingfors: Otava 1993.

SAOL = *Svenska Akademiens ordlista över svenska språket*. 11:e uppl. Stockholm: Norstedts 1986.

SOB = *Svensk ordbok*. Språkdata. Göteborgs universitet. Esselte Studium 1986.

- Stenmark, Harry: *Finlandssvenska ord och uttryck*. 3:e uppl. Stockholm: Proprius 1983.
- Stora svensk-finska ordboken 1–3*. Huvured. Göran Karlsson. Helsingfors: Suomalaisen kirjallisuuden seura 1982–1987.
- Svensk handordbok. Konstruktioner och fraseologi*. Nämnden för svensk språkvård. Red. Ture Johannisson & K.G. Ljunggren. Stockholm: Norstedts 1966.
- Svensk uttalsordlista*. Helsingfors: Forskningscentralen för de inhemska språken 1994.

Övrig litteratur

- Martola, Nina 1995: Ny mellanstor ordbok från svenska till finska. I: *Språkbruk* 4/1995, 22–25.
- Nyström, Ingegerd 1995: Suomi-ruotsi opiskelusanakirja – en pigg nykomling. I: *Språkbruk* 2/1995, 20–22.

Marketta Sundman

Turid Farbregd og Hannele Seppinen: *Finsk-norsk ordbok*. Oslo: Universitetsforlaget 1993.

Med *Finsk-norsk ordbok* (FNO) er talet på ordbøker mellom finsk og norsk fordobla (også den første stod Turid Farbregd bak, den gongen i samarbeid med Aili Kämäräinen). Sjølv om FNO naturleg nok er ei viktig hending for norske fennistar og finske nordistar, fortener ho merksemd også frå leksikografar, for dei (både heldige og mindre heldige) løysingar forfattarane har vald til ein del klassiske ordboksproblem. Men først ein kort presentasjon av FNO.

FNO inngår i Universitetsforlagets mellomvektarserie (ei rekke ordbøker frå framandspråk til norsk, større enn lommeordboka, men mindre enn storordboka). Ho skil seg frå forgjengaren sin (*Finsk-Norsk Norsk-finsk lommeordbok*, 1978 (NFL)) ved at ordforrådet er større (23.000 ord mot i underkant av 11.000 finske oppslagsord i NFL. Merk likevel at den norske delen i NFL er større, omtrent 22.000), ved at det i mange tilfelle er gitt eksempelsetningar for bruk, og ved at FNO gir grammatiske opplysningar om dei finske oppslagsorda der NFL gav opplysningar berre om dei norske.

Symmetri

Som allereide nemnt har FNO finske, men ikkje norske grammatiske opplysningar, ho er med andre ord *asymmetrisk*, i likheit med NFL, som er asymmetrisk den andre vegen. Dei fleste tospråklege ordbøker er asymmetriske. Dette gjeld særleg der mest alle brukarane brukar boka same veg, men og i mange andre tilfelle. Det nederlandske forlaget van Dale har som politikk å fordele dei grammatiske opplysningane, slik at med to storordbøker engelsk-nederlandsk / nederlandsk-engelsk vil opplysningane om engelsk stå i det eine bandet og om nederlandsk i det andre. Sjølv er eg motstandar av asymmetriske ordbøker, av følgjande grunnar: Som språkinnlærarar treng vi dei grammatiske opplysningane om framandspråket både når vi slår det opp i ei L2-L1-ordbok (for å lære det), og når vi leitar dei opp via ei L1-L2-ordbok (for å bruke det). Med asymmetriske ordbøker treng vi dermed ideelt sett fire ordbøker for kvart språkpar for å gjøra alle til lags. Mellom store språk som tysk og engelsk er ikkje dette noko problem, men det er utopisk å tenke seg fire ordbøker sjølv mellom såpass store språk som norsk og finsk (for

ein meir utdjupande argumentasjon sjå Bergenholtz 1995 og dei referansane han siterer). To vakre eksempel på symmetriske ordbøker har vi i Davíðsson og Holms *Svensk-isländsk ordbok / Sænsk-íslendsk orðabók* frå 1982, og i Valerij Berkovs *Russko-norvezjskij slovar' / Russisk-norsk ordbok* (Moskva 1987, sjå nedanfor), så reint teknisk er det ikkje noko problem å produsere dei. Det som eventuelt talar mot symmetriske ordbøker, er for det første at framstillinga kan bli tyngd av mange grammatiske opplysningar viss både oppslags- og særleg forklaringsord skal vera grammatisk merkte. Det andre momentet som talar mot symmetriske ordbøker, er tilfella der alle potensielle brukarar kjenner det eine språket betre enn det andre. Eit slikt tilfelle har vi for *Åarjelsaemien-daaroen baakoegærja / Sydsamisk-norsk ordbok* (Bergsland og Mattson Magga 1993), der samtlege sørsamiskspråklege beherskar norsk/svensk betre enn sørsamisk på dei aller fleste samfunnsområde. Ordboka er asymmetrisk, og gjev opplysningar berre om dei sørsamiske oppslagsorda, noko som er forståeleg i dette tilfellet. Men i svært mange andre tilfelle er det ikkje slik, og for FNO har begge brukargrupper like stort behov dei grammatiske opplysningane om det andre språket. Ei eventuell nyttgåve av ordboka bør absolutt innarbeide grammatiske opplysningar også om dei norske orda, ikkje minst med tanke på salstala (det er **langt fleire** finnar med norskkunnskapar enn omvendt, for å uttrykke meg forsiktig).

Grammatiske opplysningar for dei finske orda

FNO følgjer systemet i *Suomen kielen perussanakirja* (SKP), med ein tabell med 51 ulike nominale og 27 ulike verbale bøyingsmønster, fortløpande nummerert frå 1 til 78. Talet 39 bak oppslagsordet **aavistus**³⁹ anelse tyder at ordet blir bøygt etter 39: *vastaus* : *vastauksen* : *vastausta* : *vastaukseen* : {*vastausten*, *vastauksien*} : *vastauksia* : *vastauksiin* (i ordboka vist i tabellform). Stadieveksling har ein tilleggskode, først ei stjerne som markerer stadieveksling, og deretter ein bokstav frå A til M som markerer typen av stadieveksling, slik: **aavikoitua**^{52*F} forvandles til ørken, der F viser til vekslinga t : d (*aavikoidun*, osb.). Dette er ein svært presis måte å gje opplysningar på, sjølv om det krev at brukaren blar fram og tilbake i boka. Systemet til SKP er og i ferd med å få ein posisjon som standardssystem, slik at når eg bruker ein av dei andre ordbøkene mine, *Suomalais-marilainen sanakirja*, møter eg dei same kodane der med. Dessverre ber ikkje systemet preg av å vera laga spesielt for tospråklege ordbøker. Kodane kunne t.d. ha vore nummererte frå 1 til 51 for nomena og frå 1 til 27 for verba (ordklassetilhørigheita ville ha gjort det klart kva paradigme det

refererte til). Dei ulike bøyingsklassene er heller ikkje like sjølvstendige. Viss vi ser bort i frå variasjonen i fleirtal kan vi få talet på nominale vokalstammer ned til 14 (mot no 31) og variasjonen i fleirtal kjem enten av automatiske vokalvekslingar foran fleirtalsaffikset *-i-* eller av at skriftspråket for dei ulike klassene har tatt i bruk ulike kombinasjonar av dei 2–5 ulike sideformene for genitiv, illativ og partitiv pluralis. Viss vi deler koden i to delar, ein for eintals- og ein for fleirtalsformene (evt. ein del for eintal og bøyingsstamma i fleirtal og ein del for variasjonen i genitiv og partitiv fleirtal), får vi eit system som skil mellom vesentleg og uvesentleg informasjon på ein heilt annan måte enn SKP sitt system. Ei form som **laatikko**⁴ vil i staden få koden ¹B, der 1 tilsvarer SKPs grupper {1–2,4–6,7–14}, og B fortel oss at ordet skil seg frå t.d. **palvelu**² (=1A i det nye systemet) ved å ha ei kortarte sideform i illativ pluralis. Finsk har 26 ulike kasus/numerus-kombinasjonar, av dei er det berre tre som varierer, og desse tre bør ikkje få komplisere resten av systemet. Klassene bør også grupperas etter frekvens, slik at venlege klasser får låge nummer. Det er lett å forstå at når systemet no ein gong eksisterer, er det svært enkelt for ordboksforfattaren å berre kopiere koden direkte frå kjelda. Det eksisterer likevel alternativ til SKP sitt system. Den store *Svensk-finsk storordbok* (SFSO) opererer med eit system der orda i utgangspunktet er delt inn i klasser etter endingsvokal(ar) eller -konsonant. Innafor kvar klasse er så den mest frekvente bøyingstypen umarkert, og andre grupper får koder frå 2 og oppover til maksimalt x. Alt i alt får SFSO langt fleire klasser enn sjølv SKP i si noverande form, så det er ikkje umiddelbart klart kva system som er det beste. Det vil likevel vera synd om SKP sitt system skulle oppnå status som standardssystem utan at ein revisjon etter dei retningslinene eg her har skissert har vorte prøvd ut. Dette er ei oppgåve for det framveksande forskingsmiljøet for finsk som framandspråk!

Opplysningane om stadieveksling er overtydelege. I nesten alle tilfelle går det automatisk fram kva veksling det er snakk om (*tt* kan berre veksle med *t*, osv.) I eksempelet ovafor kunne *t* teoretisk ha veksla med *tt* og ikkje med *d*, men retninga på stadievekslinga vil alltid gå fram av bøyingsstamma, og nybyrjarar kan slå opp i bøyingsparadigmet (her: 52) viss dei er usikre. Unntaka gjeld veksling mellom *k* og *Ø/v/j*, men her vil omgjevningane avgjøre, og vekslingstabellen kan oppgje vilkåra for variasjon.. Det viktige er å få vita om konsonanten vekslar (som i **aavikoitua**^{52*F}) eller ikkje (som i **auto**¹), og den informasjonen kjem fram også med ein forenkla notasjon. Praksisen med symbolet * for stadieveksling (utan tilleggsbokstav) er allereide i bruk i fleire av dei tospråklege storordbøkene. Viss stadievekslings-

notasjonen blir forenkla, blir det og plass til å dele bøyingsklasseinformasjonen i to, slik eg har gjort framlegg om.

Det grammatiske forordet er kort (5 sider), og fell mellom to stolar: Det er ein (vel) lang smakebit på finsk grammatikk, men han er likevel for kort og usystematisk til å hjelpe brukaren til å avgjøre kva form ordet ho er interessert i, står i. Uttaleavsnittet fortel ikkje om dei to unntaka til den fonematiske rettskrivinga (<ng, nk> = [ññ, ñk] og konsonantforlenging etter visse finale -e og etter infinitivs- og imperativ singularis-suffikset), og heller ikkje at finske lukkelydar er uaspirerte. I kasustabellen manglar dei perifere kasusa abessiv, komitativ og instruktiv. I staden blir dei forklarte i teksten, noko som fører til at dei kasusa vi får best forklart, er dei perifere! Dei 6 lokalkasusa burde ha vore sette opp systematisk etter parameterane INDRE/YTRE og TIL/I~PÅ/FRÅ. At akkusativ ikkje har eigne former, som forordet hevdar, er feil. For det første har dei personlege pronomena i både eintal og fleirtal ei unik akkusativform på *-t*. For det andre, og det er det viktigaste, er akkusativ formelt skilt frå alle andre kasus ved å ha **suffikskombinasjonen** *-n* i eintal og *-t* i fleirtal, noko ingen andre kasus har (finske grammatikarar har ikkje alltid dette klart for seg, til stor skade for oss finskstudentar)¹. Nominativ- og partitivobjekt (*Avaa ovi!*, *En avannut ovea*), som teksten hevdar er eksempel på akkusativ (fordi dei er objekt kjem dei inn under samleomgrepet akkusativ), har ingenting med akkusativ å gjøra, men er nettopp det dei blir kalla: objekt i kasus nominativ og i kasus partitiv.

Possessivsuffiksa blir omtalt, men ikkje sett opp i tabellform. Suffikset for 3. person er ikkje nemnt i det heile, eg vil ha med både hovudforma *-nsA* og den viktige tilleggsopplysninga at det har forma *-Vn* etter kasussuffiks som endar på kort vokal.. Vokalharmoniavsnittet fortel ikkje at dei nøytrale vokalane krev fremre suffiks (regelen er: {a,o,u} i suffikset viss {a,o,u} i rota, elles, {ä,ö,y}). Vokalendringar framfor *-i* blir nemnt, men ikkje uttømmande framstilt. Det er ikkje fleire av dei (8) enn at dei kunne ha fått plass i ein tabell, med referanse til t.d. Fred Karlssons *Finskans grammatik* for ei forklaring. Litteraturlista bør skrivas med brukaren i tankane og ikkje berre fortelja kva oppslagsverk forfattaren har brukt. Fred Karlssons grammatikk er t.d. ført opp i litteraturlista, men berre i form av den finskspråkege utgåva, og ikkje som den svenskspråkege originalutgåva. Det er ført opp ei einspråkleg finsk lærebok i finsk, i staden for å kunngjøre det stolte faktum at det faktisk eksisterer ei norskspråkleg lærebok i finsk (Imerslund 1986).

¹ Også estisk har med andre ord akkusativ.

Oppsummeringsvis finn eg at denne innleiinga dessverre berre blir det ho kallar seg ("Litt om det finske språket") i staden for det ho burde ha vore fordi ordboka treng det, ei komprimert framstilling av all relevant informasjon brukaren treng for å bøye (evt. finne fram til grunnforma av) ordet ho er ute etter, med referanse til meir utfyllande grammatikkar for forklaring. Gode førebilete finn vi i den sovjetiske leksikografiske tradisjonen, t.d. i Berkov si russisk-norske ordbok, som har gode grammatikkskisser både av russisk og av norsk.

Grammatiske opplysningar for dei norske orda

Desse opplysningane manglar som allereie påpeikt dessverre heilt i FNO. Men kva form skal dei grammatiske opplysningane ha? Med ein i utgangspunktet så enkel morfologi som den norske ser dette ut som eit luksusproblem. Kompliserande faktor er likevel det faktum at bokmål har koda inn geografiske og sosiale skilnader i nettopp morfologien, slik at heile bøyingsklasser (fem.sg.best.form, nøytr.pl.best. form, pret. av I.konj.sv.verb) opererer med dobbelt sett affiks (danske for abstrakt / høgtideleg og norske for konkret / kvardagsleg / nedsettande bruk). Dette talar for eit kodesystem tilsvarande det finske, og eg ser ingen grunn til at systemet i *Bokmålsordboka* ikkje kan vera standard på same måten som Suomen kielen perussanakirja er det. Bokmålsordboka har dessutan eit betre system enn det finske: Kvar ordklasse er merkt med bokstav og nummer, slik at det regelrette bøyingskjemaet i kvart tilfelle er nummer 1. Tilfelle der variasjonen følgjer automatisk av forma på stammen, er slått saman (t.d. sterke og svake hokjønnsord). Jobben med å føre inn desse kodene er ikkje verre enn at Universitetsforlaget kan få nokon til å gjøra det relativt raskt, delar av ordforrådet kan og merkas automatisk. Også her bør det vera ei grammatisk innleiing, t.d. etter mal av forordet i *Bokmålsordboka*, men omsett til finsk, og med uttaleopplysningar. Finske brukarar vil sikkert sette pris på å få lista over bøyinga av sterke og uregelrette svake verb med seg frå NFL over i FNO.

WSOYs FNNFLO hadde opplysningar om både bokmål og nynorsk. Samanlikna med henne representerer FNO eit tilbakesteg, her er det berre bokmål. Bokmålsordboka og Nynorskordboka har dei same grammatiske kodene, nynorsk har ein litt meir komplisert morfologi, og kodesystemet i *Bokmålsordboka* er dermed eit undersett av kodesystemet i *Nynorskordboka*. Eit alternativ er å bruke det fulle (nynorske) systemet, slik at same kode viser både til nynorsk- og

bokmålsparadigma i forordet, og slik at to ulike koder kan vise til ulike paradigme for nynorsk men til same paradigme for bokmål. I dei tilfella der oppslagsordet kjem inn under ulik kategori (fordelinga av genus og av sterke/svake verb varierer ein del) kan vi få to kodar. Oppslagsorda kan vera doble **käännös** oversettelse °omsetjing, men eksempel-setningane henta berre frå eitt av skriftspråka. Om ei dobbelskriving vil ta for mykje plass, må i alle fall det norske grammatikkvedlegget vera tospråkleg, på same måte som i FNNFLO. I Noreg har vi lett for å gløyme skilnadene mellom nynorsk og bokmål, fordi vi kjenner dei. Når dei utanlandske studentane mine spør meg om dei skal lære seg nynorsk eller bokmål, svarer eg at mesteparten av den skriftkulturen dei møter vil vera på bokmål (med viktige unnatak, særleg viss det er litteraturstudentar), men at viss dei vil forstå språket dei same skribentane snakkar, må dei lære seg nynorsk. Viss ei ordbok for utanlandske norskstudentar skal kunne bli ein nøkkel til det norske språksamfunnet, må ho ta omsyn til dette.

Trykk og tonem på norsk

I norsk leksikografisk tradisjon har det av språkpolitiske grunnar vorte vanleg å la vera å seie noko om trykkplassering, av redsle for at det skulle virke talemålsnormerande. Sett frå ein utlendings ståstad er dette klart utilfredsstillande. Det politiske stridsspørsmålet går på om visse romanskbaserte lånord skal ha trykk etter opprinneleg eller norsk mønster. Reglane for trykk i norsk er som kjent (i stikkordsform) slik: Gamle germanske ord har trykk på første staving. For ord med innlånte (låg)tyske prefiks gjeld det at nokre prefiks får trykk, andre ikkje, og trykket fell på første stavinga av rota viss dei ikkje får trykk. Lånord (via diverse språk) frå gresk og latin: I dei austlege dialektane følgjer dei den germanske trykkregelen, mens dei i dei vestlege og nordlege dialektane følgjer trykktilordninga i långjevarspråket. Unntak frå dette er dei austlege høgstatusosiolektane, som har same trykkplassering som vest- og nordnorsk. Dette er kompliserte reglar, som utlendingar som vil lære norsk, absolutt bør få tilgang til. Særleg gjeld det fordelinga av trykksterke og trykksvake lågtyske prefiks. I og med at det austnorske trykkmønsteret kan avleias frå det vestnorske, bør ordboka følgje det vestnorske, men i forordet gjøra merksam på skilnaden mellom aust- og vestnorsk. Ein annan og vel så viktig grunn til å opplyse om trykk er at det dermed samtidig er mogleg å gje opplysningar om tonem. Bokmålsordboka markerer trykket viss det fell på ikkje-initiale stavingar, noko som er fornuftig nok viss det er berre trykk vi er ute etter, men dei har dessverre falle for freistinga til å

markere trykket som tonem 1 eller 2, slik at vi får den absurde situasjonen at vi får vita at *aksént* har tonem 1 (noko anna kan det heller ikkje ha), men ikkje om det heiter *kvínnfolk* eller *kvínnfolk*.

Det er og ei populær misoppfatning i Noreg at tonem berre er relevant for minimale par av substantiv/verb-typen (*lándet* – *lände*, *rénnet* – *rénne*), men slik er det sjølvsagt ikkje. Alle trykksterke stavingar er realisert som tonem 1 eller tonem 2. Viktige minimale par får vi på tvers av ordgrenser, t.d. for typen verb + partikkel: *tá på frakken* ('berøre') ~ *tà på frakken* ('kle på seg'), og også i dei tilfella der det ikkje er snakk om minimale par, vil bruk av feil tonem hos ein talar som elles skil mellom to tonem forstyrre kommunikasjonen. Det er sjølvsagt mogleg å snakke norsk utan å ha tonemdistinksjon, vi har fleire dialektar som gjør nettopp det, dialektar som ligg nettopp i den landsdelen der dei fleste finnane er. Intonasjonsmessig har dei eit system som fell saman med finsk (og med dei fleste andre av verdas språk): Trykksterke stavingar blir realiserte ved hjelp av høgtone. Eit grammatisk forord bør gjøre greie for desse alternativa, slik at dei som skal lære seg norsk med hjelp av ordboka, kan forstå systemet i norsk, og gjøra medvite val på grunnlag av denne innsikta.

Det bør med andre ord settas som ein standard for ordbøker over norsk at denne viktige sida ved norsk språk kjem med. I norske grammatikkar blir tonem tradisjonelt merkte med tal eller enkle og doble aksentar (¹*landet* – ²*lande*; *'landet* – *'lande*). Berkovs ordbok har innført eit system med akutt aksent for tonem ein og grav aksent for tonelag to, jfr. eksempla i forrige avsnitt. Dette gjennomfører han i **alle** dei norske orda, også eksempelsetningane (t.d. s. 87 *å skrive et skùespill*, men dessverre berre innafor ordgrensa, slik at han har (same side) *det å skru opp prìsene* i staden for det korrekte *det å skrù opp ...*). Fordelen med dette systemet er at det ikkje er plasskrevande, at det ikkje forstyrrar teksten så mykje som dei andre systema (lesarar som ikkje er interesserte i tonem, kan lettare ignorere retninga på aksenten og berre konsentrere seg om plasseringa), og at dei er ikonisk (rett nok for austnorsk, men likevel). Både FNO og andre norske ordbøker bør følgje Berkovs eksempel her.

Eksempelsetningar

Eit kort blick på ei tilfeldig side gir inntrykk av at det er svært få eksempelsetningar i boka. Grunnen til dette er snarare at kvart samansette ord har fått si eiga linje, slik at t.d. dei 32 oppslagsorda som er danna som samansetting med ordet *etu-* 'for(an)-', har fått kvar sin

separate artikkel, i staden for det meir økonomiske systemet med å plassere dei under same artikkel:

etu- for-

etuajassa (dagl) før tida *Juna tuli* ~ ~ Toget kom før oppsatt ankomsttid. Toget kom før oppsatt ankomsttid.

etuajo-oikeus forkjørsrett

etualla forgrunn *olla ~lla* være i forgrunnen

etuhalkio gylf

etuhammas fortann

etuilija køsniker

etuilla snike (i køen)

etuistuin forsete

etujärjestö interesseorganisasjon

...

etu- for, ~**ajassa** (dagl) før tida *Juna tuli*

~ ~ Toget kom før oppsatt ankomsttid.

~**ajo-oikeus** forkjørsrett, ~**ala** forgrunn *olla ~lla* være i forgrunnen, ~**halkio**

gylf, ~**hammas** fortann, ~**ilija** køsniker,

~**illa** snike (i køen), ~**istuin** forsete,

~**järjestö** interesseorganisasjon, ...

Ved nærare gransking finn eg likevel at eg som norsk brukar alltid finn ei eksempelsetning når eg har bruk for henne, trass i det litt ordlisteprega førsteintrykket. Finske brukarar vil sannsynlegvis ikkje vera like nøgde. Som regel gir FNO kontekst i dei tilfella det er fleire enn ei omsetting. Det heiter t.d. **kansio**³ perm, mappe, album; og **kannella**^{67*J} **1** sladre **2** påkjære, utan vidare forklaring, slik at den finske brukaren må kryssjekke i ei norsk(-finsk) ordbok for å finne skilnaden mellom *perm*, *mappe* og *album*, og mellom *sladre* og *påkjære*. I dei fleste tilfelle finn vi likevel kontekst, som i **kappalle**⁴² **1** del, stykke ~ *juusto* et stykke ost **2** avsnitt *jakaa teksti ~leisiin* dele teksten i avsnitt For å få ei symmetrisk ordbok må denne praksisen bli konsekvent, og nok ein gong kan Berkov si ordbok tene som forbilete.

Reksjon av verb blir markert med eksempelsetningar og ikkje med forkorta pronomen, som i fleire av dei finske storordbøkene. På den måten slår ein to fluger i ein smekk, i og med at tydinga av verbet kjem klarare fram med eit eksempel, samtidig med at den grammatiske informasjonen er der. Her har FNO lagt seg på ei heldig løysing.

Trykkfeil

Bokmeldingssjangeren krev at eg skal fortelja om boka inneheld trykkfeil. Det gjør ho. Den mest graverande er at siste side i bokstaven R manglar, noko forlaget også prøver å bøte på med å legge henne ved som lausblad (burde ha vore klistrelapp, til å lime inn bakarst, eg har allereie mista lausbladet mitt). Alle ordboksforfattarar og -produsentar bør føre opp denne ulukkeshendinga på lista si over Ting som ikkje skal skje i ordboksproduksjonen. Ein annan feil er at fleirtalsmerket *-t*

manglar i paradigmet som er sett opp i den grammatiske innleiinga (s. X). Det er trykkfeil i sjølve manuskriptet og, men ikkje såpass at det har plaga meg.

Omfang

Med 23.000 ord fordelt på 266 sider kjem FNO i ein mellomklasse mellom lommeordboka og storordboka, ein skjebne ho deler med dei fleste andre ordbøkene Universitetsforlaget har gjeve ut i same serie. Her ligg også FNOs desidert største svakheit: Ho er for stor til å gå ned i lomma, men for lita til bruk ut over basisnivå. FNO kjem t.d. til kort ved lesing av avisa *Helsingin Sanomat*, ordforrådet strekk ikkje til. Problemet er altså ikkje det som står i ordboka, men det som ikkje står der, og dette får meg til å sette eit spørreteikn ved denne ordboksstorleiken i det heile. Viss ordboka likevel er dømt til å stå i bokhylla og ikkje ligge i innerlomma, kan ho like godt vera minst eit par cm tjukkare, slik at eg kan vera sikker på at bladinga mi ikkje er bortkasta (blir det for mange bomturar, grip eg heller etter *Finsk-svensk storordbok* med det same). Med ei-linje-per-oppslagsord-praksisen sin er FNO ein av dei minste ordbøkene i serien, og står dermed lageleg til for ei drastisk utviding. Til neste utgåve bør denne praksisen endras til den vanlege komprimerte framstillinga av samansetningar, og ordforrådet bør aukast slik at boka likevel blir tjukkare enn førsteutgåva.

Samandrag

Sterke sider ved FNO er dei grammatiske opplysningane for dei finske orda. Også utvalet av eksempelsetningar er imponerende, ikkje så mykje for mengden av eksempel, men for evna til å skjera ned utvalet til eit minimum og likevel alltid ha eit eksempel klart når (den norske) ordboksbrukaren treng det. Den største svakheita ved sjølve ordboka (bortsett frå omfanget) er asymmetrien. Dette er særleg graverande når språket morfologisk sett er så ukomplisert som norsk: Å gjøra ordboka symmetrisk ville ikkje ha tatt mykje plass. Manglande forklaring av ulike norske omsettingsalternativ er heller ikkje så bra.

Det vi som brukarar aller mest ønsker oss, er likevel ei anna ordbok enn det som no er skrive, ei *større* ei. FNO fell (som dei fleste av søstrene sine i Universitetsforlaget sin ordbokserie) ned mellom to stolar, og bør allereide til neste opplag ha komme seg opp att og bli ei til alle høve brukbar *storordbok*.

Litteratur

Ordbøker

- Bergsland, Knut og Lajla Mattson Magga 1993: *Åarjelsaemien-daaroen baakoegærja / Sydsamisk-norsk ordbok*. Billefjord: I>ut.
- Berkov 1987= Berkov, Valerij: *Russko-norvezjskij slovar' / Russisk-norsk ordbok*. Moskva: Russkij Jazyk. 2. utgåve Oslo: Universitetsforlaget 1995.
- Bokmålsordboka* = Landrø, Marit Ingebjørg og Boye Wangenstein (red.) 1986: *Bokmålsordboka. Definisjons- og rettskrivingsordbok*. Oslo: Universitetsforlaget.
- Daví>sson og Holm 1982 = Daví>sson, A>alsteinn og Gösta Holm: *Svensk-isländsk ordbok / Sænsk-íslensk or>abók*. Lund/Reykjavík: Walter Ekström bokförlag/Almenna Bókafélagi>.
- Finsk-svensk storordbok* = Cannelin, Knut, Aulis Hirvensalo og Nils Hedlund 1976: *Suomi-ruotsi Suursanakirja / Finsk-svensk storordbok*. Borgå: WSOY
- FNNFLO = Farbregd, Turid og Aili Kämäräinen 1978: *Finsk-Norsk Norsk-Finsk lommeordbok*. Helsingfors: WSOY
- Moisio, Arto og Valentin Vasiljev 1995: *Suomalais-marilainen sanakirja*. Åbo: Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 52.
- Nynorskordboka* = Hovdenak, Marit, Lurits Killingbergtrø, Arne Luvhjell, Sigurd Nordlie, Magne Rommetveit og Dagfinn Worren (red) 1986: *Nynorskordboka. Definisjons- og rettskrivingsordbok*. Oslo: Det Norske Samlaget.
- SFSO = Göran Karlsson (red.) 1984: *Svensk-finsk storordbok / suomi-ruotsi suursanakirja I–III*. Suomen kirjallisuuden seruran toimituksia 358. Jyväskylä: Gummerius
- SKP = Harala, Risto (päätoimittaja) 1990: *Suomen kielen perussanakirja. Ensimmäinen osa. A–K*. Kotimaisten kielten tutkimuskeskuksen julkaisuja 55. Helsinki: Painutuskeskus Oy.

Annan litteratur

- Bergenholtz, Henning 1995: Grammatik i bilingvale ordbøger. *Lexico-Nordica* 2, 5–18.
- Imerslund, Bente 1986: *Ymmärrän! : lærebok i finsk*. Oslo: Universitetsforlaget.
- Karlsson, Fred 1978: *Finskans grammatik*. Helsingfors: SKS.
- Karlsson, Fred 1983: *Suomen peruskielioppi*. Helsingfors: SKS.

Franziskus Geeb

Magnar Brekke, Øivin Andersen, Trine Dahl, Johan Myking (udg.): *Applications and Implications of current LSP Research. Proceedings of the 9th European Symposium on LSP*. Vol I + II, Bergen, Aug. 2–6, 1993. Bergen: Fakkbokforlaget 1994.

Det er et imponerende tobindsværk på i alt 969 sider, som dokumenterer omfang og indhold af fagsprogskonferencen i Bergen 1993. Som udgivelsesårstal angives 1994, mig bekendt blev bindene først udsendt i slutningen af 1995. Der har måske været nogle tekniske problemer, som i sidste ende har udsat den endelige trykning. Rent teknisk kan udgiverne roses for at have bragt næsten alle artikler i et sådant format, at bindene fremstår som en helhed. Der er dog enkelte bidrag, som har en mindre linieafstand end hovedparten. Alt i alt drejer det sig om 107 artikler. Mange af dem er yderst korte på en 3–4 sider. Mange er også indholdsmæssigt forholdsvis tynde. Det har så tilsyneladende ikke været muligt for udgiverne at frasortere sådanne ret intetsigende artikler.

Som bekendt hersker der ikke bred enighed, hvad fagsprog er. Denne problematik tematiseres også i flere bidrag bl.a. af Christer Laurén "LSP-valid concept in 1993" og af Jennifer Kewley Draskau "LSP – A subordinate concept", men også i mange andre bidrag problematiseres fagsprogsopfattelsen. Den bredeste forståelse findes i de bidrag, som eksplicit eller implicit relaterer sig til Kalverkämpers artikler, dvs. at enhver tekst må forstås som en LSP-tekst, language for special purposes, dog sådan at der kan skelnes mellem forskellige grader af faglighed. Dette synspunkt gemmer sig også bag det første af de leksikografisk relevante artikler, som her skal omtales. E. Agirre et al., som beskriver i deres artikel "Intelligent Dictionary Help System" overførelsen af papirordbøger til en ordbogsdatabase. Interessant er det i denne sammenhæng, at de også opfatter sproget i ordbøger som fagsprog: "In this work dictionary definitions are considered as a special purpose language" (174). Selve omsætningen til en database minder principielt om den, der er kendt i andre sammenhænge, men et nærmere bekendtskab med dette system kunne muligvis være givtigt for både fagsproglige og almensproglige leksikografer.

Pirjo Järvi diskuterer i artiklen med titlen "Geldmarkterminologie-Beispiele von Begriffssystemen im Deutschen, Englischen und Finnischen" en række kulturafhængige økonomiske termer. F.eks. er der mellem "Geldmarkt", "moneymarket" og det finske "rahamarkkinat" kun

indtil en hvis grad ækvivalensrelationer. Dette ved forfatteren, som er ansat i Finlands Bank, men de undersøgte ordbøger har kun i utilstrækkelig grad været i stand til at klargøre forskellen. Kritik af denne art fra fagfolk er ganske vist mere destruktiv end konstruktiv, men dog meget vigtig, idet den påviser nødvendigheden af et samarbejde mellem fageksperter og leksikografer.

En lignende ækvivalensproblematik diskuteres i to bidrag om juridisk terminologi, som i endnu tydeligere grad end den økonomiske terminologi er kendetegnet af en stor grad af kulturbundethed. Åge Lind fremstiller i "Equivalence in legal terminology" grundproblematikken og giver nogle løsningsforslag. Ingrid Simonnæs diskuterer mere specielt ækvivalensrelationer mellem norsk og tysk juridisk sprog. Hendes fremstilling er yderst læseværdig, også for almensproglige leksikografer.

Endelig vil jeg nævne artiklen af Grethe Duvå, et al. "LSP-Dictionary translation – An integrated concept". Denne artikel er også kendt fra en tysk udgave i Bergenholtz/Schaeder (Udg.): *Fachlexikographie*, Tübingen 1994. – Grundtanken forudsætter en klar skillelinie mellem leksikografi og terminografi, en adskillelse, som terminologer, men ikke alle leksikografer, lægger vægt på. Der påvises en lang række informationshuller og fejl i foreliggende fagordbøger. Disse tilordnes leksikografien. Forfatterens tese er, at denne utilstrækkelige praksis vil kunne forbedres gennem en samtidig hensyntagen til terminologiske teorier og terminografiske metoder. Hertil hører efter deres mening specielt at fagordbogen skal have en systematisk makrostruktur. Herimod kan der i princippet kun komme den indvending, at den fagsystematiske arbejdsmetode og opbygning allerede er teoretisk kendt og også bliver anvendt i leksikografien. Desuden har den glatalfabetiske eller systematiske præsentation af leksikografiske, terminografiske eller terminologiske data ved den stigende brug af databaser og elektroniske medier til produktion og brug af ordbøger og termlister kun mindre betydning, da præsentationen af disse data næsten altid kan ændres efter brugerens behov.

De to bind udgør en samling af artikler, som både i positivt og negativt lys gengiver de tendenser, som har været kendetegnet for fagsprogsforskning i slutningen af 80'erne og begyndelsen af 90'erne. Som sådan må de to bind absolut have interesse for alle fagleksikografer, men også almensproglige leksikografer kunne måske have glæde af den ene eller den anden af bidragene, specielt de få, som nævnes i denne omtale.

Martin Gellerstam

Skandinavisk ordbok. Utarbetad av Birgitta Lindgren (huvudredaktör), Skirne Helg Bruland, Allan Karker och Ståle Løland. Utgiven i samarbete med Nordisk språksekretariat. Stockholm: Norstedts 1994.

Behovet av ett lexikaliskt hjälpmedel som kan underlätta det språkliga umgänget mellan de skandinaviska länderna är av gammalt datum. Många minns den klassiska berättelsen om "Den skinnsjuke greven" – en parodi på översättning från danska till svenska publicerad 1902 i Dagens Nyheter – som med jämna mellanrum har plockats fram och dammats av under det här århundradet. Tidigare har man tagit fasta på just de "lömska likheterna" mellan språken: att man exempelvis i normal svenska inte ledigt kan svänga en *stock* mellan fingrarna, att *mågar* som simmar på vattenytan inte hör till vanligheterna och att herrar inte brukar klä sig i *eleganta kjolar* (allt enligt den klassiska översättningen).

Från korta ordlistor med inriktning på sådana här lömska – och lätt anekdotiska – likheter har nu steget tagits till en hel Skandinavisk ordbok, frukten av ett samarbete mellan medarbetare i de tre inblandade länderna och med Nordiska språksekretariatet som tillskyndare. Detta är ett bra initiativ som har lett till en välkommen handbok i umgänget mellan de tre systerspråken. Man blir överraskad över hur många skillnader man trots umgänge med de nordiska språken inte kände till. Den nya ordboken tar inte bara upp de "lömska" ord som har samma form men olika betydelse (*rolig*), och de ord som har samma form och en gemensam betydelse men dessutom en skiljaktig betydelse (*stock*), utan också ord som är unika för ett (eller ett par) av språken (sv. *dyrgrip*). Rätt många ord hör till denna sista kategori men gränsen till de falska vännerna är svår att bestämma. Tittar man i början på bokstaven R i den danska listan så är drygt hälften av orden unika för danskan i förhållande till den svenske läsaren. Men en stor del är fortfarande alltså av den lömska sorten: *rabat* 'vägkant', *rask* 'frisk', *rede op* 'bädda', *rede sig* 'kamma sig' osv. Det är för övrigt inte bara enstaka ord som tas upp i ordlistan utan också längre uttryck: *det rabler for ham* 'det går runt för honom', *stikke i rend* 'sätta iväg', *rette bager for smed* 'straffa fel person', *rotte sig samman* 'gadda ihop sig' m.fl. nyttiga uttryck.

Ordboken har en ingång från varje språk och består alltså av tre ordlistor där varje uppslagsord förklaras genom en kort ekvivalent (mer sällan en längre beskrivning) på de två andra språken. I förordet sägs att det handlar om 10.000 ord "som kan ge problem i umgänget mellan danskar, norrmän och svenskar". Varje ordlista är paginerad för sig,

något som man upptäcker först efter en stunds användning. Det skulle inte ha skadat med en enkel (färg)markering som berättade var man befinner sig. Annars är boken behändig till formatet och har tydlig typografi.

Varje uppslagsord får en ekvivalent på de andra språken, även i de fall där ett av språken har samma ord som det som förekommer uppslagsordet. Detta är en medveten lösning. Om det danska ordet *sagtens* motsvaras av norska *saktens* och av svenska *förvisso* så kan en svensk "ha glädje av att se att samma ord också används i norska". Detta är möjligt men beror lite på hur ordboken används – läser man en dansk bok så vill man kanske i första hand veta vad den danska texten betyder. Ett annat sätt att signalera denna likhet vore annars att **inte** ha med sådana ekvivalenter som är lika med uppslagsordet (ett stickprov antyder att kanske hälften av uppslagsorden har en (nästan) identisk ekvivalent i ett av de andra språken). Men frågan om likhet är inte alltid enkel så man kan förstå beslutet att ha en och samma mall för alla ord.

Urvalet av ord och betydelser i en ordbok som omfattar tre språk är inte helt okomplicerat. Ett problem är hur långt man skall gå för att förklara ord som bara finns i ett eller två av språken. Här kommer man snabbt in på den tvåspråkiga ordbokens revir och det gäller att nöja sig med de vanligare förekommande orden av det här slaget. Ett annat problem är hur långt man skall hjälpa den läsare som huvudsakligen behärskar sitt eget språk men inte grannspråket. Exemplet *saft* som tas upp i inledningen får illustrera ett sådant problem. Enkelt uttryckt kan det danska och norska ordet *saft* dels betyda 'en typ av dryck' och dels 'sav' medan det svenska ordet bara har den första betydelsen. Hur gör man då för att ingen förvirring skall uppstå? Jo, i den danska och norska listan tar man upp ordet och ger den här informationen men i den svenska tar man inte upp ordet "eftersom det svenska ordet har samma betydelse som danska och norska *saft*"(?). Och – tillägger man – "ordet *sav* tas givetvis upp som uppslagsord i den svenska listan". Men hur vet danskar och norrmän som vuxit upp med att inte göra den svenska distinktionen mellan de två betydelserna vad det svenska ordet *saft* betyder? Om någonstans borde väl den svenska listan tala om att det svenska ordet *saft* enbart handlar om den där drycken. Men detta är alltså den redovisade principen. För att återgå till Den skinssjuke greven så saknas alltså svenska *stock* 'trädstam' (inte 'käpp') i den svenska listan, likaså *is* (som inte betyder 'glass'), *mangel* (som inte betyder "brist"), *sopp* (som inte betyder 'svamp') osv. Den här informationen kan man naturligtvis undertrycka om man vill tona ner en viss typ av de "lömska" likheterna ("false friends") men nog kan jag tänka mig att en sådan information skulle vara till nytta.

De ord som faller under denna kategori är inte fler än att man skulle kunna ta med dem i en ny upplaga av Skandinavisk ordbok. Tillskottet skulle göra en bra och välkommen Skandinavisk ordbok ännu bättre.

Guvarur Már Gunnlaugsson

Íslendskur-føroyskur orðalisti. Nordisk sprogsekretariat. 1995. 74 sider.

I de siste år har Nordisk sprogsekretariat i Oslo udgivet nogle små ordlister mellem nordiske sprog med den hovedhensigt at pege på ligheder og forskeller indenfor ordforrådet i to nærbeslægtede sprog. Lige nu er der udkommet en ny ordliste i denne serie, en islandsk-færøsk ordliste. Det fremgår af forordet at færingen Carl Jóhan Jensen har lavet listen, men Málstovan i Tórshavn har revideret den ud fra en sammenligning med en ordliste i en lærebog i islandsk (*Lærið íslenskt*) for færingers af Turi Sigurðardóttir.

Her som ellers i de andre internordiske ordlister skal lemmaselektionen for en stor del tjene det formål at forklare forskellig betydning af ord som har samme eller lignende udseende i begge sprogene. Men ordlisten indeholder også en god del almindelige islandske ord som ikke har modsvarighed af samme stamme i færøsk. Det er indlysende at i en ordliste af denne type lægges hovedvægten på at give færøske ækvivalenter til de islandske ord. Selvfølgelig er de grammatiske oplysninger minimale, faktisk begrænset til markering af ordklasse (og køn for substantiverne).

Som eksempel på ord taget op i ordlisten på grund af at de har samme grafiske form i islandsk og færøsk men forskellig betydning kan nævnes *epli* ('æble' på islandsk, 'kartoffel' på færøsk) og *synda* ('svømme' på islandsk, 'synde' på færøsk). Ordlisten indeholder ganske mange ord af denne type men forståelig nok er det nemt af føje til ord som ikke er taget med, som f.eks. *tíma* ('næppe' på islandsk, 'gide' på færøsk).

Ordlisten indeholder også ord som har bevaret sin gamle betydning men har fået forskellige former i de to sprog, både fonetisk og grafisk, på grund af forskellige lydændringer. Dette gælder for ord som isl. *gjá* – fær. *gjógv* 'kløft', isl. *amma* – fær. *omma* 'bedstemor' og isl. *oft* – fær. *ofta* 'ofte'.

En stor del af de islandske lemmaer er selvfølgelig ord som ikke har en morfologisk ækvivalent i færøsk, f.eks. isl. *jarðgöng* – fær. *berghol* eller *tunnill* 'tunnel'.

I en ordliste af denne størrelse er det naturligvis nemt at finde eksempler på ord en savner: ordet *belja* 'ko' er lige så almindeligt i islandsk talesprog som synonymet *kýr* men kun det sistnævnte er taget med i ordlisten. Dette lemmavalg stemmer ikke overens med at ordet *rolla*, som har lignende stilværdi overfor synonymet *ær* som *belja* overfor *kyr*, bliver lemmatiseret.

Som tidligere nævnt bygger lemmaselektionen delvis på en lærebog i islandsk for færinger. Det kan være grunden til at lemmalisten indeholder nogle temmelig sjældne ord af litterær art, som f.eks. *ari* 'ørn', *hnasl* 'mundfuld' og *munaraugu* 'længselsfuldt blik'.

I de fleste artikler er indholdet begrænset til én eller flere ækvivalenter til lemmaet. Men en del artikler indeholder desuden ordforbindelser og/eller eksempler. Den største del af disse forbindelser er oversat til færøsk men ibland findes der også konstruktioner, kollokationer og eksempler uden oversættelse for at illustrere brugen af lemmaet, som f.eks. *mig rámar í fla* (under *ráma*), *vera í uppáhaldi hjá e-m* (under *uppáhald*), *komast í kynni við e-n* (under *kynni*), *snúa úr íslensku á færeysku* (under *snúa*). En kan stille et spørgsmålstegn ved om en fremstilling af denne type, som må være beregnet på færinger med hensyn til aktiv brug, hører hjemme i en så knap tosproglig ordliste.

Når det gælder udseendet af ordbogsteksten, typografi og layout giver den islandsk-færøske ordliste et godt indtryk. Det er derfor en overraskelse at finde en række fejl i alfabetiseringen af lemmaerne: *afleitur* står efter *afli*, *beygja* foran *beyging*, *brúnn* foran *brum*, *búinn* efter *bústörf*, *bylur* foran *bylting*, *diskur* foran *disklingur*, *draugur* foran *drasl*, *fordyri* efter *forum*, *kviðmikill* efter *kvilli*, *flukla* efter *flulur*.

Det er all grund til at være taknemmelig for det initiativ Nordisk sprogsekretariat har vist med at udgive internordiske ordlister. Den islandsk-færøske ordliste er et værdifuldt bidrag til denne udgave. En logisk fortsættelse ville være at udgive en færøsk-islandsk ordliste.

Indgåede bøger maj 1995 til maj 1996

- Magnar Brekke/Øivin Andersen/Trine Dahl/Johan Myking:
Applications and Implications of Current LSP Research. Volume I+II. Bergen: Fagbokforlaget 1994. (969 sider)
- Jens Axelsen: *En rød Klassiker. Dansk-Engelsk Ordbog 1845–1995*. København: Gyldendal 1995. (169 sider)
- Jens Axelsen: *Dansk-Engelsk Ordbog*. 10. udgave. København: Gyldendal 1995. (1045 sider) (DKK 220)
- B. Kjærulff Nielsen: *Engelsk Dansk*. 5. udg. København: Gyldendal 1996.
- Hans-Peder Kromann/Anne Lise Kjær (Hrsg.): *Von der Allgegenwart der Lexikologie. Kontrastive Lexikologie als Vorstufe zur zweisprachigen Lexikographie. Akten des internationalen Werkstattgesprächs zur kontrastiven Lexikologie 29.–30.10.1994 in Kopenhagen*. Tübingen: Niemeyer 1996. (131 sider) (DM 98,-)
- Turid Farbredeg - Aili Kämäräinen, Suomi-norja-suomi taskusanakirja ('Finsk- norsk-finsk fickordbok') 1996, 110 FIM
- Suomi-englanti-suomi CD-ROM-sanakirja 1996, 580 FIM Författare: ???
- Suomi-englanti-suomi elektroninen sanakirja 1996, 580 FIM - diskettversion av föreg.
- Ilse Cantell, Nina Martola, Birgitta Romppanen & Mats-Peter Sundström/ Anja Sarantola & Tauno Sarantola
- Suomi-ruotsi-suomi elektroninen sanakirja 1996, 580 FIM - diskettversion av föreg.
- Suomi-ranska-suomi CD-ROM-sanakirja 1996, 580 FIM Författare: Jean-Michel Kalmbach/ Seppo Sundelin
- Suomi-ranska-suomi elektroninen sanakirja 1996, 580 FIM - diskettversion av föreg.
- Suomi-saksa-suomi CD-ROM-sanakirja 1996, 580 FIM Författare: ???
- Suomi-saksa-suomi elektroninen sanakirja 1996, 580 FIM - diskettversion av föreg.
- István Nyirkos, Suomi-unkari-suomi taskusanakirja ('Finsk-ungersk-finsk fickordbok')1996, 110 FIM
- Encyclopædia Iuridica Fennica III. Perhe-, työ- ja sosiaalioikeus. ('Familje-, arbets- och socialrätt'). Päätoimittaja ('huvudredaktör') Heikki E.S. Mattila. Suomalaisen lakimiesyhdistyksen julkaisuja. C-sarja n:o 26. Helsinki 1996.

Suomen sanojen alkuperä. Etymologinen sanakirja 2. L-P. Suomalaisen kirjallisuuden seura ja Kotimaisten kielten tutkimuskeskus. Helsinki 1995.

Vi skriver sidan 2

Meddelelser fra Nordisk forening for leksikografi

I det siste har foreningens virksomhet vært konsentrert om tre oppgaver, utarbeidelsen av Nordisk leksikografisk ordbok, oppbyggingen av en leksikografisk database og forberedelser til den fjerde nordiske leksikografikonferanse i Helsinki neste år.

Nordisk leksikografisk ordbok

Som kjent har foreningens hovedprosjekt de siste årene vært å utarbeide en ordbok over leksikografisk terminologi, Nordisk leksikografisk ordbok. Arbeidet med ordboka er nå i slutfasen, og den utkommer i mai 1997 på Scandinavian University Press, Universitetsforlaget i Oslo. Ordboka inneholder omtrent 850 definerte termer, og i tillegg kommer bortimot 150 henvisningslemmer (for en stor del synonymer), så i alt vil Nordisk leksikografisk ordbok inneholde omtrent 1000 termer. Redaksjonsspråket er norsk bokmål, som bl.a. betyr at bokmålstermer utgjør lemmabestanden i ordboka. Ved hvert lemma blir det angitt ekvivalenter på dansk, finsk, islandsk, nynorsk og svensk, samt engelsk, fransk og tysk. Gjennom indekser gis det også direkte adgang til de enkelte ekvivalentspråk. En utførlig systematisk innledning gir oversikt over leksikografien som disiplin, både teoretisk og praktisk, og plasserer sentrale begreper i teoretisk sammenheng. Fra de enkelte definisjonene vises det til en paragraf i den systematiske innledningen. Termene er også markert i forhold til et klassifiseringssystem som dekker samtlige termer.

Arbeidet med definisjoner og forklaringer er delt opp mellom medlemmene i redaksjonsgruppa, slik at enkelte medlemmer er ansvarlige for bestemte termklasser. Ordboka er redigert av Henning Bergenholtz, Ilse Cantell, Ruth Vatvedt Fjeld, Dag Gundersen, Jón Hilmar Jónsson og Bo Svensén, og har dessuten bidrag fra Helgi Haraldsson, Hans Kristian Mikkelsen og Jaakko Sivula.

Nordisk leksikografisk ordbok er et nybrottsarbeid i nordisk leksikografi, og også i internasjonal sammenheng har ordboka en verdi som det hittil største og mest systematiske oppslagsverk om leksikografisk terminologi. For Nordisk forening for leksikografi er utgivelsen av denne ordboka av stor betydning. Foruten å tjene som et praktisk oppslagsverk for nordiske leksikografer gir den en bred faglig oversikt over leksikografien og vil på den måte forhåpentlig inspirere til nye initiativ på det leksikografiske området.

Leksikografisk database

I sin tid ble det samlet omfattende materiale fra nordiske leksikografer til "Oversikt over nålevende utøvere av nordisk leksikografi". Styret for NFL har nå satt i gang arbeid med å bearbeide materialet, med den hensikt at oversikten i første omgang blir tilgjengelig som nettverksbase. Her går man ut fra at oversikten utgjør en del av en større leksikografisk database i NFLs regi. Styret anser det som en sentral oppgave å planlegge og bygge opp en slik database i nærmeste fremtid. Det første skrittet vil bli å lage en bibliografisk base med utgangspunkt i litteraturreferanser i konferanserapportene og LexicoNordica. Styret tar gjerne i mot forslag og råd når det gjelder den videre oppbyggingen av databasen.

Vi gjør oppmerksom på foreningens hjemmeadresse på Internett: <http://svenska.gu.se/nfl./nfl.html>.

Jón Hilmar Jónsson

Konferensen om leksikografi i Norden 1997 i Esbo i Finland 21–24.5. 1997

Nordiska föreningen för leksikografi anordnar den fjärde konferensen om leksikografi i Norden i samarbete med Hanaholmen – kulturcentrum för Sverige och Finland, Nordiska språksekretariatet (fram till årsskiftet 1995/96) och Forskningscentralen för de inhemska språken. Temat för konferensen är *Ordboken och användaren* och det är tänkt att täcka föredrag över ett brett spektrum. Frågor som kan tas upp är t.ex. olika *användargrupperns behov*, vad man vet om dem och hur väl de tillgodoses, *ordböckerna och språkundervisningen*, *ordbokskritiken* och hur den kan tänkas tjäna användarna, *informationskategorierna* i en ordbok och varför vissa typer av information finns med i vissa typer av ordböcker osv. En utförligare presentation av temat finns i det första inbjudningsbrevet och även på [www](http://www.nordiska.se) under nedanstående adress. Liksom under de föregående konferenserna är bidrag också om andra relevanta ämnen inom teoretisk och praktisk leksikografi välkomna. Deltagaravgiften är beräknad till FIM 600 för medlemmar och FIM 700 för icke-medlemmar.

Anmälningssblanketter och närmare upplysningar kan erhållas antingen från Nordiska språksekretariatet (se bakpärmen) eller från Forskningscentralen för de inhemska språken, Nina Martola, Sörnäs strandväg 25, FI-00550 Helsingfors, Finland, tfn +358-9-731 52 78, e-post Nina.Martola@domlang.fi. Bindande anmälan skall vara arrangörerna till handa senast den 13 december 1996. Sprid gärna information

om konferensen till alla som kan tänkas vara intresserade! Utförligare information finns tillgänglig på www under den direkta adressen <http://www.domlang.fi/nfl/index.html> eller via länk från föreningens hemsida.

Nina Martola