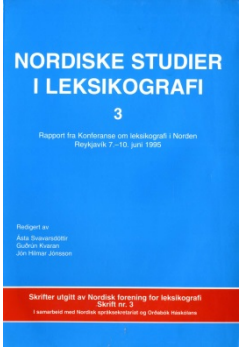


# NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Ordböcker som språkmodeller	
Forfatter:	Bo Ralph	
Kilde:	Nordiske Studier i Leksikografi 3, 1995, s. 339-350 Rapport fra Konferanse om leksikografi i Norden, Reykjavík 7.-10. juni 1995	
URL:	<a href="http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive">http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive</a>	

© Nordisk forening for leksikografi

## Betingelser for brug af denne artikel

Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavsmanden til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

## Søgbarhed

Artiklerne i de ældre Nordiske studier i leksikografi (1-5) er skannet og OCR-behandlet. OCR står for 'optical character recognition' og kan ved tegngenkendelse konvertere et billede til tekst. Dermed kan man søge i teksten. Imidlertid kan der opstå fejl i tegngenkendelsen, og når man søger på fx navne, skal man være forberedt på at søgningen ikke er 100 % pålidelig.

## Ordböcker som språkmodeller

It has sometimes been claimed that dictionaries which reflect the mental lexicon as faithfully as possible are likely to be more functional than dictionaries which do not. In this contribution, some fundamental properties of dictionaries which tend to create problems in this connexion are discussed. In particular, the way entry forms are traditionally selected in dictionaries and the lemmatization concept in itself give rise to asymmetries alien to the information network of the mental lexicon. However, lexicographers can contribute considerably when a realistic model of the lexicon is being established.

Det är normalt i vetenskapliga sammanhang att olika praktiska tillämpningar är sekundära i förhållande till (den primära) teoribildningen. Tillämpningarna är då inte bara logiskt sekundära, vilket är rimligt, utan också reellt, vilket inte är lika självklart. De kan alltså betraktas som senare uppkomna praktiska avknoppningar till redan existerande, vetenskapligt väletablerade forskningsdiscipliner. Detta är normalfallet inom exempelvis tekniska och medicinska vetenskapsgränar, och det kan nog sägas vara ett vetenskapligt ideal i den västerländska forskningstraditionen.

Ser man noga på saken framstår emellertid inte denna bild lika tydligt inom de humanistiska forskningsgränarna. Många discipliner har där så småningom kommit att fördjupa en från början ganska svagt utvecklad teoretisk profil. Vår egen disciplin, lexikografen, uppvisar denna utvecklingsgång. Ett mera målmedvetet teoriarbete har där först efterhand utvecklats ur en redan väl fungerande praktisk verksamhet.

### 1 Lexikaliskt teoriarbete

Den teoribildning som är relevant för det lexikografiska arbetet är av två grundläggande slag: specifikt lexikografisk teori kring själva det lexikografiska arbetet å ena sidan och allmän språkvetenskaplig teori med direkt tillämpning inom den lexikografiska verksamheten å den andra. Teori kring lexikografi har naturligtvis lång tradition på det viset att praktiskt inriktade lexikografer teoretiserat i någon mening kring sitt arbete. Det kan gälla urvalsprinciper vid informationsuppbyggnaden, organisationen av denna information, utformningen i förhållande till tänkta målgrupper etc. Emellertid har de teoretiska övervägandena vanligen redovisats endast implicit, genom de konkreta resultat som de teoretiskt baserade besluten manifesterats i. Mera sällan har den teoretiska grunden för de underliggande resonemangen redovisats explicit i inledningar och förord. En monografisk framställning som Svensén (1987) med principiella lexikografiska överväganden var också närmast unik när den kom.

Lexikografiska ställningstaganden som en reflex av den språkvetenskapliga debatten är på det hela taget ett ganska nytt fenomen. Det är väl korrekt att säga att det är först när ordböcker på senare tid tillkommit i anslutning till akademiska miljöer som språkvetenskapliga argument mera uttryckligen lagts till grund för lexikografiska beslut, framför allt sedan den teoretiskt mycket genomarbetade *Longman's Dictionary of Contemporary English* (1978) visat vägen.

Dessförinnan har vi haft en relativt lång tradition av i första hand kommersiell ordboksutgivning, dvs. rena förlagsprojekt. Förord och inledningar i sådana ordböcker kan visserligen vara ganska utförliga men innehåller oftast praktiska anvisningar för användningen snarare än en redovisning av det teoretiska grundarbetet. Det som här sagts betyder givetvis inte att rent kommersiella projekt, som bedrivits vid exempelvis förlagens ordboksredaktioner, inte skulle ha tagit ställning till språkteoretiska och lexikografiteoretiska problem. Tvärtom — så har det säkert normalt varit, men det har oftast stannat vid enbart en *implicit* redovisning, som sällan lett till mer utbredda principiella diskussioner, helt enkelt därför att traditionen inte har inbjudit till det. Med tradition menas då ett ganska kortvarigt bruk i ”modern” tid; i äldre ordböcker kunde det ofta förekomma långa utläggningar av både grammatisk och encyklopedisk art, breda forskningsöversikter osv. Ett välkänt sådant exempel är Adelung (1774–86) på den internationella arenan och från svenskt område i viss mån redan Spiegel (1712) och Ihre (1769).

Det kan också vara på sin plats att understryka än en gång, att den distinktion som här görs mellan praktisk lexikografi och olika slags teori inte innebär någon värdering till nackdel för den ”praktiska lexikografin”. Vi vet ju att så kallad praktisk lexikografi inte alltid haft något högt prestigevärde i teoretiskt orienterade språkvetares världsbild, men man kommer inte ifrån att det är ”praktiska lexikografer” och inte ”teoretiska lingvister” vi har att tacka för de ovärderliga språkliga hjälpmedel som vi kallar ordböcker. (Se även mitt bidrag till volymen från Oslokonferensen (Fjeld 1992), särskilt s. 28–32!)

Som en tredje aspekt på relationen mellan teori och lexikografi vill jag nämna den rörelse som uttryckligen proklamerades i USA för ett femtontal år sedan, som kallades ”computational lexicology” av utövarna själva. Det hela startade med en uppmärksam doktorsavhandling av Robert Amsler vid University of Texas, Austin (Amsler 1980). Man syftade till att genom noggrann datalingvistisk analys av maskinläsbara ordböcker dra generella slutsatser av betydelse för den allmänna språkteorin. Särskilt avsåg man att utveckla språkmodellens lexikaliska komponent, som dittills blivit i stort sett ignorerad. Därvid utmanade man den tidigare allmänt omfattade språkteoretiska uppfattningen att syntaxen skulle vara språkmodellens helt centrala komponent. I ett för computational lexicology-rörelsen typiskt arbete citerar Raoul Smith (1981:1) Reiger (1978) och Mel’čuk (1978) i två mycket entydiga uttalanden:

There has been an increased interest within cognitive science in the status of the lexicon. In particular, the position of the lexicon as the central element in models of natural language users is being expressed clearly and strongly in the literature. Reiger, in his development of a story comprehender, argues:

Syntax [...] is secondary, something which accrues from a more primary language phenomenon. This phenomenon, we will propose, is the organization and selection of *word senses* [...] [Most] of the complexity in natural language, in fact most of its information content, lies in the individual word sense, rather than in any central or uniform system of rules [...] [W]ord sense discrimination is the key to language comprehension (1978:97–98).

Mel’čuk, in his study of a lexicographical model for text processing, agrees:

Text analysis and synthesis are conducted on the basis of stored information about words [...] and their combinations [...] Any text is built up of words having a definite meaning and combined in accordance with definite properties. To analyze or synthesize a text one must know word meanings and their combinatorial properties. The respective data are stored in the dictionary; therefore a dictionary is the most important component of any meaningful A[utomatic] T[ext] P[rocessing] system [...] Grammar rules are no more than a means of abbreviating a dictionary, an attempt to formulate useful generalizations about the dictionary, something that completely depends on the dictionary [...] Without a dictionary no ATP system is possible (1978:3–4).

Jag är osäker på vad som senare kommit ut av de planer företrädarna för ”computational lexicology” hade. Det anordnades ett internationellt symposium med ett mycket begränsat antal deltagare vid Stanford Research Institute International 1983, varvid en permanent organisation och en särskild tidskrift diskuterades, men det förefaller som om de drivande krafterna sedan dess absorberats av annan datalingsvistisk forskning eller olika kommersiella åtaganden. Hur som helst kan man konstatera att lexikaliska frågor generellt kommit att spela en betydligt större roll i språkvetenskapen under den följande tidsperioden, samtidigt som den lexikografiska verksamheten intensifierats och förfinats ytterligare. Praktisk lexikografi som allmän ram för ett snarast språkteoretiskt översiktsverk som Aitchison (1987) hade knappast varit tänkbart tidigare.

Under alla omständigheter innebär den inriktning som nämndes strax ovan, ”computational lexicology”, ett totalt perspektivskifte från den i bästa fall lätt nådiga attityd som språkteoretiker tidigare ofta intagit, med innebörden att den enda relation som råder mellan praktisk lexikografi och språk teori är den som består i att lexikografer kan sköta sitt uppdrag bättre om de är språkteoretiskt medvetna, eftersom de då kan göra språkvetenskapligt bättre genomarbetade ordböcker. Även om det är rimligt att anta att lingvistiskt väl skolade lexikografer åstadkommer bättre ordböcker än språkteoretiska ignoranter, kvarstår det fortfarande som ett obevisat påstående att ordböcker som troget speglar människans sätt att spontant hantera den lexikaliska sidan av språket är generellt mer ändamålsenliga än ordböcker som saknar sådana egenskaper. Om det påståendet emellertid vore sant skulle den idealiska ordboken också kunna fungera som något slags modell över språkbrukarnas lexikaliska förmåga. Det är med utgångspunkt i sådana frågeställningar som jag i det följande skulle vilja göra några opretentiösa iakttagelser och överväganden. Ett av skälen är att jag själv tidigare gjort mig till tolk för den antydda ståndpunkten men efter att ha stått vid sidan av praktiskt lexikografiarbete under en dryg tioårsperiod nu har fått många av mina tidigare perspektiv förskjutna. De frågor som tas upp nedan är välkända lexikografiska problem, men de behöver ånyo övervägas i just det här perspektivet.

## 2 Ordböcker och databaser

Ett av skälen till att relationen mellan lexikografi och språk teori ändrades i början av 1980-talet var den tekniska utvecklingen. Dittills hade man visserligen på sina håll ägnat

sig åt att konvertera traditionellt framtagna ordböcker till maskinläsbar form i samband med uppdatering och publicering av nya upplagor, men nu blev nya ordböcker omedelbart maskinläsbara i samband med att de etablerades direkt i de nya elektroniska medierna. Man fick således tillgång till stora datorläsbara lexikaliska material. Ett mål som emfatiskt framhölls inom den nya datorbaserade lexikografin var att omfattande, väl genomarbetade databaser skulle upprättas, varur olika specialiserade produkter skulle kunna genereras. Detta var ett av skälen till att åtskilliga aktiva språkforskare var ivriga att engagera sig i verksamheten. Det idealtillstånd som eftersträvades bestod i att man skulle få en funktionsuppdelning mellan databasen och de olika lexikografiska specialprodukterna. Databasen skulle vara en reservoar att ösa ur efter behag, när specifikt utformade ordböcker skulle sammanställas för olika ändamål. Visionen var nog i allmänhet att den grundläggande databasen snarare än de konkreta ordböckerna skulle ha karaktären av teoretiskt fullgånge språkmodell.

Den senare utvecklingen har väl visat vissa tendenser åt detta håll. För lexikografiska projekt som för andra mänskliga aktiviteter gäller emellertid, med en lätt modifiering av en rad hos den svenske diktaren Stagnelius, att "verklighetens hårda band oss trycka". De ofta starkt begränsade resurserna har nödvändiggjort att även de mest storslaget planerade databaser fått anpassas till de krav som ställts på de konkreta produkter som skulle extraheras ur dem. Den hett åtrådda kombinationen av databasperspektivet och ordboksperspektivet i form av det *elektroniska lexikonet* — som alltså är något kvalitativt annorlunda än en datorläsbar ordbok och exempelvis skall kunna medge att informationen genomlysas från ett stort antal olika perspektiv — har veterligen inte förverkligats någonstans. Vad som tills vidare förekommer är *dels* elektroniskt tillgänglig lexikalisk information i mer eller mindre ordbokstrogen form, information som givetvis också kan ha ett allmänt språkvetenskapligt intresse av stor räckvidd, *dels* mera utpräglat modellliknande lexikaliska informationsmängder, antingen för teoretiska ändamål och då i allmänhet experimentellt utformade eller i starkt reducerad form, eller anpassade till mycket specialiserade ändamål av artificial intelligence-typ.

Vi kan bara konstatera att det ännu är långt kvar till en fungerande modell över människans totala lexikaliska förmåga. Lexikografer kan knappast ta något primärt ansvar för att en fullfjädrad sådan modell verkligen konstrueras. Frågan är om lexikografin över huvud taget har något att vinna på att ta del i dessa strävanden. En förutsättning är givetvis att man tror att det därigenom går att få bättre ordböcker.

Att en lexikonmodell syftar till att ge en trogen bild av människans lexikaliskt inriktade språkförmåga är självklart. Däremot kan ordböcker ha en hel rad olika funktioner. Ordböcker är i första hand praktiskt motiverade hjälpmedel, som följaktligen sammanställs i syfte att lösa en konkret uppgift. Syftet kan vara rätt generellt eller mer specifikt beroende på typen av ordbok. Frågan är då om det är relevant i de enskilda fallen att ordboken utformas i enlighet med en idé om det mentala lexikonets uppbyggnad. Det ligger kanske närmare till hands att ett sådant mål eftersträvas i enspråkiga definitionsordböcker än i tvåspråkiga allmänordböcker å ena sidan, olika specialordböcker å den andra. En enspråkig ordbok med en långtgående ambition att beskriva ett enskilt språks lexikaliska sida så utförligt som möjligt aktualiserar frågan om "psykologisk realitet" (i exempelvis Chomskys terminologi) på ett mer naturligt sätt än en specialordbok av något slag, låt säga en ordbok över flygterminologi. I det senare fallet kan det inte vara ett primärt önskemål att ordboken troget avspeglar en människas sätt

att mentalt lagra ifrågavarande ordförråd. Det viktiga är att den samlade termuppsättning som maximalt kan komma till användning inom det specificerade fackområdet på det aktuella språket finns väl beskriven i ordboken. På likartat sätt kan man givetvis resonera för de flesta ordböcker över specialområden. Ordbokens syfte i det enskilda fallet kan således vara avgörande för frågan huruvida en modell över det mentala lexikonet är ett eftersträvansvärt mönster eller ej.

Något svårare att besvara är möjligen frågan om önskvärdheten av att tvåspråkiga ordböcker troget avspeglar formen för mänsklig lexikalisk kompetens. Syftet med dessa ordböcker är ju att underlätta översättning mellan språk. Att en tvåspråkig ordbok är utformad som en lexikonmodell skulle exempelvis kunna innebära att man försöker efterlikna en genuint tvåspråkig persons lexikaliska förmåga, vad nu detta skulle innebära konkret. Vad lexikografiteoretiker observerat på det här området är emellertid snarare, att tvåspråkiga ordböcker för att optimalt kunna fylla sin uppgift bör vara olika utformade med hänsyn till om de vänder sig till språkbrukare som primärt har det ena aktuella språket som modersmål eller det andra, om man har ett källspråks- eller målspråksperspektiv. Det är alltså snarast de i varje enskilt fall inblandade språkparens kontrastiva egenskaper som behöver fokuseras, och för varje språkpar skulle det i det idealiska fallet finnas åtta ordböcker: en produktions- och en perceptionsordbok på vardera språket med särskild utformning för talare av respektive språk. Detta fulla utbud brukar av praktiska skäl vara en omöjlighet att realisera, men en integrerad modell av en tvåspråkig talares kompetens — hur den nu skulle se ut — fyller under inga omständigheter samma funktion. De lexikografiskt kritiska frågorna gäller snarare informationsurval och informationens organisation.

### 3 Ordurval och lexikalisk systematik

En primär fråga inom lexikografien är hur orden i varje given ordbok skall väljas ut. Alla vet att mängden av uppslagsord är det kanske viktigaste försäljningsargumentet när en ordbok marknadsförs, oavsett hur sofistikerade lösningar på diverse andra problem man kan uppvisa. I praktiskt lexikografiska sammanhang blir därför ordurvalet ett optimeringsproblem, en fråga om hur man på effektivast möjliga sätt presenterar så mycket information som möjligt om så stor ordmängd som möjligt till en tänkt användargrupp. Olika skarpsinniga lösningar på diverse problem kan naturligtvis vara till nytta, men det är ändå vanligen de strikt *praktiska* bedömningarna som blir avgörande för resultatet.

Om ordboken skall reflektera en *lexikonmodell* blir redan ordurvalet en *teoretiskt* viktig fråga. Urvalet blir inte i första hand en utrymmesfråga, utan det skall också spegla någon form av *lexikalisk kompetens*. I själva verket skall det ordurval som man stannat för i ordboken återspegla ett *naturligt ordförråd* i någon mening. En sådan inställning kan leda till olika slutsatser. Om man vill inta en extrem ståndpunkt kan man säga att ingen specialordbok, t.ex. en terminologisk ordbok över ett visst fackområde, kan återge språksamhällets kollektiva kompetens. Dels är den lexikaliska informationen i en sådan ordbok så specialiserad att den svarar mot endast mycket få individers kompetens, dels saknas det mesta som bygger upp den språkliga kompetensen i allmänhet, nämligen större delen av det allmänspråkliga materialet. Här blir också valet av teoretisk grund avgörande. Konflikten mellan vad som ingår i individers lexikaliska språkförmåga och den kollektiva kompetensen är ett generellt deskriptivt problem. Specialordboken kan i och för sig mycket väl, liksom Chomskys

grammatikmodell, tillåtas spegla en i sammanhanget "idealisk språkbrukares" kompetens, nämligen den specialiserade fackmannens.

Det är emellertid ofrånkomligt att urvalskriterierna för ordmängden i praktiskt användbara respektive modellteoretiskt trogna ordböcker lätt råkar i konflikt med varandra. En lexikonmodell måste beskriva det reguljära allmänspråket, dvs. koncentrera sig på generaliteter i språket, vilket även en ordbok kan göra, men ordböcker behöver väl så ofta betona det speciella, det anmärkningsvärda: fackspråk, dialektformer, ålderdomlighet, olika stilregister osv. Vad av allt detta som inkluderas i en ordbok kan vara beroende av exempelvis nationell lexikografisk tradition. Exempelvis lämnar Blöndals isländsk-danska ordbok (1920–24) och Árni Böðvarssons *Íslensk orðabók* (1983), liksom den isländska lexikografiska traditionen i allmänhet, betydligt större utrymme åt ålderdomliga uttryck än vad som är vanligt i de flesta jämförbara skandinaviska ordböcker. Om dessa ordböcker över huvud taget skall betraktas som synkroniska beskrivningar av det isländska ordförrådet, får man förutsätta ett ovanligt generöst tilltaget synkroniskt stadium.

Botolv Helleland tog för några år sedan upp frågan om vilken plats egennamnen skall ha i en ordbok (Helleland 1992). Problemet är mer komplicerat än man kanske brukar tänka på. I enspråkiga ordböcker utelämnas ofta alla ord som man skriver med stor inledningsbokstav, medan sammansättningar med och avledningar av sådana ord genom sitt ordslut betraktas som principiellt icke-propriella och tas med. Däremot kan tvåspråkiga lexikon innehålla vissa egennamn som uppslagsord. Vanligen gäller det då bara sådana som har speciellt anmärkningsvärd form i ettdera av de språk lexikonet involverar, eller båda. Så kan i en engelsk-svensk ordbok *The Mediterranean (Sea)* med den svenska motsvarigheten *Medelhavet* återfinnas, vidare namn som *Prague* med uttalsangivelse och svensk motsvarighet *Prag*, medan urvalet av ort- och personnamn för den stora majoriteten av logiskt möjliga fall blir en bedömningsfråga; lexikografens föreställning om vad som är betydelsefullt för ordboksanvändaren blir avgörande.

Som jag påpekat i ett par andra sammanhang (t.ex. Ralph 1992:33) spelar namn en statistiskt ganska betydande roll i uppbyggnaden av texter, men framför allt intar de en mycket framskjuten plats i ordförrådet. I en realistisk lexikonmodell måste därför egennamnen beredas den plats som är motiverat med hänsyn till kravet på psykologisk relevans. När enspråkiga ordböcker återger uppslagsord som innehåller propriella morfem som element, medan själva de egennamn som ligger till grund för dessa utesluts, uppstår teoretiskt viktiga effekter. För att ta namn på stater och så kallade nationalitetsord som exempel, så bildar dessa vanligen naturliga lexikaliska grupper bestående av statsnamn (t.ex. *Island, Ungern*), inbyggarebeteckning (*islänning, ungrare*), nationalitetsadjektiv (*isländsk, ungersk*), och det är i många sammanhang också relevant att hit räkna beteckning på officiellt språk (*isländska, ungerska*; se även Sigurd (1965) och *Statsnavne og nationalitetsord* (1994)). Dessa grupper kan i ett givet språk, exempelvis svenskan, visa en ganska förutsägbar morfologisk form (som i de just nämnda fallen) eller en mer oförutsägbar, idiosynkratisk form (såsom *Danmark* : *dansk* : *dansk* : *danska*; *Norge* : *norrman* : *norsk* : *norska*). I och med att avledningarna bildas med hjälp av ett antal olika derivationsmorfem, måste varje enskild manifestationsform, dvs. den verkliga, konventionaliserade formen, preciseras i lexikonet, såväl människans mentala lexikon och den teoretiska modellen av detta som den fysiska ordboken.

Det finns emellertid trots allt en viss parallellitet mellan de fyra potentiella orden i varje grupp, ett slags partiellt återkommande systematik, även om det inte är helt förutsägbart

exakt vilka former som är gängse i varje enskilt fall. Denna tendens till systematik framgår av exemplen i tabell 1, där några olika typer samlats. Nu finns det i detta namnmaterial en del specialfall. Vi behöver bara gå till den återstående nordiska nationen, Finland, för att hitta ett. Som inbyggarbeteckning till statsnamnet *Finland* fungerar både *finländare* och *finne*, som nationaladjektiv på samma sätt *finländsk* eller *finsk*. Varianterna kan i vissa sammanhang användas distinkt (finlandssvenskar inkluderade eller ej), men de är i många sammanhang funktionellt synonyma. Som språkbeteckning saknas av naturliga skäl den ena morfologiskt möjliga varianten, dvs. \**finländska*, medan den enda förekommande språkbeteckningen, *finska*, ansluter sig morfologiskt till *finne* : *finsk*.

Den lucka vi har i serien *Finland* : *finländare* : *finländsk* : — ser vi en motsvarighet till i en rad fall, där man saknar ett nationalspråk vars beteckning återgår på statsnamnet eller nationaladjektivet (utan att det finns en kompletterande serie som *Finland* : *finne* : *finsk* : *finska*): *Libanon* : *libanes* : *libanesisk* : —; *Liechtenstein* : *liechtensteinare* : *liechtensteinsk* : —; *Tanzania* : *tanzanier* : *tanzanisk* : —; osv. Det omvända förhållandet kan sägas råda i en del andra avledningsserier, såsom — : *same* : *samisk* : *samiska*, där något motsvarande statsnamn inte existerar av det enkla skälet att samerna inte bildat någon suverän stat. Beteckningen *Sameland* (*Sápmi*) har en något annorlunda status men kommer naturligtvis språkligt nära statsnamnen såsom beteckning på samernas traditionella bosättningsområde. Morfologiskt och semantiskt blir emellertid den ”icke-propriella” delen av serier som *same* : *samisk* : *samiska* och *finne* : *finsk* : *finska* fullständigt parallellt uppbyggd. Ännu lite annorlunda status har beteckningen *Kurdistan*, som i första hand har en etablerad geografisk innebörd. Området bebos huvudsakligen av den etniskt urskiljbara gruppen kurder, och även om Kurdistan inte är en suverän stat faller namnet således väl in i en serie *Kurdistan* : *kurd* : *kurdisk* : *kurdiska*.

Albanien	alban	albansk	albanska
Bulgarien	bulgar	bulgarisk	bulgariska
Danmark	dansk	dansk	danska
Frankrike	fransman	fransk	franska
Färöarna	färing	färöisk	färöiska
Grönland	grönländare	grönländsk	grönländska
Island	islänning	isländsk	isländska
Japan	japan	japansk	japanska
Norge	norrman	norsk	norska
Polen	polack	polsk	polska
Portugal	portugis	portugisisk	portugisiska
Sverige	svensk	svensk	svenska
Tjeckien	tjeck	tjeckisk	tjeckiska
Ungern	ungrare	ungersk	ungerska
Vietnam	vietnames	vietnamesisk	vietnamesiska

Tabell 1. Statsnamn och nationalitetsord. Symmetriska serier.

I en del andra fall med närmast etniska beteckningar saknas likaså allmänt accepterade beteckningar på landområdet. Så är det exempelvis för samojedernas del (om vi bortser från den skämtsamma beteckningen *Samojedien* för någon allmänt avlägsen plats). Vi får således



en ohjälpligt defekt serie, som i övrigt från språkliga utgångspunkter sett ansluter sig väl till de tidigare nämnda: — : *samojed* : *samojedisk* : *samojediska*. Ytterligare en variant, som kommer nära denna är — : *arab* : *arabisk* : *arabiska*, som likaså avser kulturförhållanden snarare än nationsförhållanden. Till denna defekta serie kan antingen territorialbeteckningen *Arabstaterna* läggas, som då är en sammanfattande beteckning på de stater som vi förknippar med ett i kulturellt avseende arabiskt livsmönster, eller *Arabien*, som får uppfattas som ett historiskt begrepp mer än ett politiskt. Den nutida motsvarighet i statshänseende som kommer närmast är väl *Saudiarabien* med avledningarna *saudiarab/saudier* respektive *saudi-arabisk/saudisk*, men man kan observera att beteckningen på det officiella nationalspråket är *arabiska*. Till denna svit av beteckningar som hänför sig till etniska grupper ansluter sig egentligen också på ett naturligt sätt *jude* : *judisk*, men här saknas såväl morfologiskt besläktad territorialbeteckning som språkbeteckning (om vi inte räknar *jiddisch*, vilket vore orimligt). De olika problemfall som behandlats här sammanfattas i tabell 2. De anförda exemplen är hämtade från svenskan men har uppenbarligen direkta motsvarigheter i andra språk.

Det är helt klart att man från språkliga utgångspunkter måste betrakta dessa etniska beteckningar som del av ett kontinuum, som mot den ena extremen rymmer de klara statsnamnen med deras respektive avledningar, mot den andra entydigt icke-propriella, rent appellativiska bildningar. Det är också rimligt att dessa nästan termliknande ord bildar en subkategori i språkbrukarnas mentala lexikon med väl utarbetade associationsbanor mellan de prototypiska fallen å ena sidan och till mer eller mindre avvikande fall å den andra. De mönster som delvis kan urskiljas i detta subsystem blir naturligtvis mindre tydliga om allt som liknar egennamn diskvalificeras, såsom faktiskt sker normalt i de flesta ordböcker. De asymmetriska förhållanden som råder mellan olika avledningsserier framstår också mindre tydligt. Detta är således ett exempel på hur ordböcker med nuvarande praxis omöjligt kan uppfattas som en trogen bild av människors lexikaliska kompetens.

Finland	finländare/ finne	finländsk/ finsk	— finska
Libanon	libanes	libanesisk	—
Liechtenstein	liechtensteinare	liechtensteinsk	—
Tanzania	tanzanier	tanzanisk	—
(Sameland)	same	samisk	samiska
Kurdistan	kurd	kurdisk	kurdiska
—	samojed	samojedisk	samojediska
(Arabstaterna/Arabien)	arab	arabisk	arabiska
Saudiarabien	saudiarab/ saudier	saudiarabisk/ saudisk	arabiska
—	jude	judisk	—

Tabell 2. *Asymmetriska serier.*

Man kan säga att egennamn som lexikaliska enheter vetter mot den encyklopediska kunskap som ordböcker sällan tar som sin uppgift att beskriva. Distinktionen mellan språklig och encyklopedisk information är emellertid ännu ett stående problem inom lexikografen. Kanske kan man för lexikonmodellens del påstå att en fullständigt realistisk sådan föreligger först när den är inordnad i en modell över människans totala kognitiva förmåga. I så fall

understryks diskrepansen mellan målet för lexikonmodellen och vad som kan vara ett rimligt mål för en ordbok ytterligare.

## 4 Den lexikaliska informationens lagringsform

Välkända är tidigare språkteoretiska beskrivningar av lexikonet som ”en lista över morfemen i ett språk”, ”över oregelbundenheter” eller ”oförutsägbara enheter som inte kan fångas med regelskrivning”. Den syn som kommer till uttryck i sådana formuleringar reducerar på ett avgörande och fatalt sätt lexikonets grundläggande mångdimensionella natur. I lexikonet kan ett stort antal subsystem av olika slag urskiljas. Egennamnen som nyss berördes kort bildar ett sådant system, och det går att urskilja ett stort antal liknande subsystem på olika grunder: grammatiska, semantiska, encyklopediska etc. Klasserna går sällan att skarpt avgränsa, utom i mycket speciella slutna system, som t.ex. beteckningar för militär rang. De måste också urskiljas i olika dimensioner, med följderna att en och samma enhet kan komma att ingå som medlem i olika klasser beroende på vilket perspektiv man anlägger.

Detta är ett synsätt som står i strid med grammatiktraditionens klassificering och därmed också ordbokstraditionens. Det skapar redovisningsproblem och aktualiserar en av de mest grundläggande frågorna i såväl arbetet på modellen över det mentala lexikonet som inom lexikografen: hur skall en lexikalisk enhet preciseras? I en ordbok som syftar till att efterlikna en lexikonmodell bör det rimligen vara god överensstämmelse mellan modellens lexikaliska enhet och ordbokens uppslagsord. Vad som skall kunna vara ett uppslagsord är naturligtvis också en lexikografiteoretiskt viktig fråga i sig.

Grundläggande för lexikonmodellens lexikaliska enhet är begreppet ”lexikalisering”, som vi ofta talar om, men som väl knappast är tillfredsställande definierat. Principiellt bör man av en lexikalisk enhet i det mentala lexikonet kräva att det är en konventionellt etablerad, fast bildning med specificerbar lexikalisk funktion. Liknande krav ställs ofta även på uppslagsord i ordböcker av teoretiskt intresserade lexikografer. Men frågan är hur enheternas grad av lexikalisering — i betydelsen ”konventionellt etablerade enheter” — skall uppfattas i de enskilda fallen. Eftersom ordböcker i första hand skall tjäna som praktisk hjälp, är det möjligen graden av förutsägbarhet som är det viktigaste urvalskriteriet, och det är inte riktigt detsamma som graden av lexikalisering. (För ett illustrativt fall i denna fråga se t.ex. Malmgren 1994.) Inte minst kan det just sagda gälla för tvåspråkiga ordböcker, där det inte bara rör sig om respektive enhets förutsägbarhet i respektive språk utan också förutsägbarheten av att två definierade motsvarigheter skall korrespondera till varandra i respektive språk, givet språkens regelsystem.

Uppslagsorden i ordböcker brukar representera en flexionsserie, ett morfologiskt sammanhållet paradigm. Den abstraktion som lemmatisering innebär och valet av kanonisk form som uppslagsord kan komma att dölja en skiftande verklighet. I en svensk flexionsserie *fort* : *fortare* : *fortast* kan möjligen formen *fort* uppfattas som logiskt primär framför övriga former, men hur förhåller sig komparativformerna *bättre*, *hellre*, *sämre* till superlativformerna *bäst*, *helst*, *sämst*, som saknar morfologiskt regelbundna motsvarigheter i positiv? Är någondera formen primär framför den andra? Normalt brukar dessa böjningsformer redovisas var för sig i ordböcker (alltså t.ex. *hellre* skilt från *helst*), med olika korsreferenser till varandra och till de positivformer som brukar suppleras (*bra*, *gärna*, *dålig*). Hur än redovisningsproblemet

blir löst kvarstår uppenbart en obalans för "suppletiva" böjningar såsom [*gärna*] : *hellre* : *helst* i förhållande till "regelbundna" flexionsserier som *fort* : *fortare* : *fortast*.

Total icke-förekomst av en böjningsform är bara ett extremfall av de frekvensförhållanden som kan råda inom paradigmerna. Vilken roll frekvens spelar för språkinläring och lagring av språklig information är en omstridd och besvärlig fråga, liksom det är svårt att avgöra hur frekvensuppgifter skall kunna utnyttjas för att göra ordböcker mera ändamålsenliga. För att ta ett enda exempel förekommer ordet *resurs* i ett stort svenskt språkmateriale från 1965 (redovisat i Allén et al. 1970–80) endast i pluralis, medan singularformen tio år senare börjat uppträda och sedan dess blivit allt vanligare. Det som i det allmänna medvetandet snarast var ett defekt paradigm för en generation sedan är nu kompletterat, men i ordböckerna syns ingen skillnad, trots att själva uppslagsformen på 1960-talet bör ha varit en närmast konstruerad form. För andra ord kan det vara andra böjningsformer som är dominerande inom paradigmerna. Det står helt klart att de olika konkreta formerna inom ett paradigm har helt olika egenskaper med hänsyn till en rad faktorer, förutom frekvens även syntaktisk ram, fraseologi, semantiska restriktioner osv., med andra ord: skillnader som vi är noga med att notera för olika lemman! En modell över språkförmågan bör anvisa mekanismer för generering av icke ännu konventionaliserade men potentiella regelbundna former, samtidigt som den bör spegla det rådande tillståndet. I ännu högre grad är det viktigt att ordböcker ger besked om det faktiskt utnyttjade systemet.

Är det över huvud taget möjligt att fånga hela detta spektrum av lexikaliskt viktiga egenskaper hos böjda former under en och samma kanoniska form och hävda att denna har psykologisk relevans? I vissa avseenden verkar onekligen de enskilda formerna i ett paradigm vara de enda psykologiskt rimliga lexikaliska enheterna. Detta är knappast en lexikografiskt acceptabel tanke, men i olika teoretiska modeller har den föreslagits (t.ex. i flera arbeten av James McCawley), och den finner också språkpsykologiskt stöd i exempelvis språkinläringens progression. Det har framhållits med ökande emfas att lexikonmodellen principiellt bör utformas som ett sofistikerat associativt nätverk (Aitchison 1987, särsk. s. 190 ff.), troligen så sofistikerat att det ännu är långt kvar till en förverkligad fungerande modell. Därmed sätter ordbokens fysiska förutsättningar vissa mycket bestämda gränser för hur nära den teoretiska lexikonmodellen man kan komma med en ordbok. Kraven på linearitet och val av en genomförd indelningsgrund i varje enskilt fall reser oöverkomliga hinder mot att lexikonmodellens lagringsform efterliknas i ordböcker trots aldrig så generösa korsreferenssystem. Naturligtvis kan man då återigen hävda att helt andra möjligheter öppnar sig med ett elektroniskt lexikon. Den frågan berörs något ytterligare avslutningsvis.

## 5 Framtiden

Mycket av det som här framförts kan tyda på att den gamla frågan om ordböckernas närmande till teoretiska språkmodeller för närvarande är en inaktuell frågeställning. Det finns emellertid ytterligare några aspekter som kan vara värda att poängtera. För det första går den tekniska utvecklingen entydigt i riktning mot att kraftfulla persondatorer blir tillgängliga för allt fler personer, och allt större och mer komplicerade datamängder hanteras med allt enklare medel. Därmed kan man fortfarande spekulera i möjligheten att elektroniska "ordböcker" för allmänt bruk kan bli verklighet i en någorlunda överskådlig framtid. Därmed kan flera av de tidigare nämnda problemen åter bli aktuella.

I väntan på denna utveckling är det kanske viktigare att konstatera att ingen har bättre förutsättningar att bidra till den teoretiska lexikonmodellens utveckling i många väsentliga avseenden än erfarna lexikografer. En bodelning mellan de båda verksamheterna skulle därför vara olycklig. Slutligen kvarstår åtskilliga olösta frågor som är av intresse för både lexikografiteori och lexikaliskt inriktad språkteori.

Lexikaliseringsproblemet är en sådan fråga som behöver mer belysning. Det är långt ifrån klart vad som skall betraktas som en lexikaliserad form och hur lexikaliseringsprocessen går till. Troligen måste man räkna med lexikalisering som ett skalfenomen snarare än ett binärt förhållande. Den nyss diskuterade frågan om uppslagsord och kanoniska former är av allmänt intresse, även om man skulle komma till olika resultat för ordböckernas respektive lexikonmodellens del. Det lexikaliska universets omfattning och en rad abstrakthetsfrågor behöver fortfarande preciseras såväl inom ordbokens som lexikonmodellens ram. Metaspråkets utformning är en högintressant fråga för både lexikografisk teori och allmän språkteori. Vi intresserar oss som språkvetare i allt högre grad för folktaxonomi, men vi avfärdar slentrianmässigt folketymologier såsom kontrafaktiva, trots att de tydligtvis representerar folkspråkliga kompetens och kan avslöja en del om deras språkliga slutledningsförmåga. En fråga som nästan inte alls uppmärksammas är folkdefinitioner, dvs. hur man spontant definierar ord. Kan dessa på ett intressant sätt utnyttjas teoretiskt och praktiskt inom den lexikaliska verksamheten såsom en spegling av den mänskliga ordsemantiska kompetensen?

Med detta hoppas jag ha argumenterat något för att lexikografer och språkteoretiker även fortsättningsvis bör ta del av varandras arbete med stor uppmärksamhet.

## Bibliografi

- Adelung, Johann Christoph 1774–86: *Versuch eines vollständigen grammatisch-kritischen Wörterbuchs der hochdeutschen Mundart* [...]. 1.5. Leipzig.
- Aitchison, Jean 1987: *Words in the Mind. An Introduction to the Mental Lexicon*. Oxford/New York: Basil Blackwell.
- Allén, Sture et al. 1970–1980: *Nusvensk frekvensordbok baserad på tidningstext*. 1–4. Stockholm: Almqvist & Wiksell.
- Amsler, Robert A. 1980: *The Structure of Merriam-Webster Pocket Dictionary*. Austin, Texas: University of Austin, Texas.
- Blöndal se Sigfús Blöndal.
- Fjeld, Ruth Vatvedt (utg.) 1992: *Nordiske studier i leksikografi. Rapport fra Konferanse om leksikografi i Norden 28.–31 mai 1991*. Oslo: Skrifter utgitt av Nordisk forening for leksikografi. Skrift nr. 1.
- Garde, Anna/Pia Jarvad (utg.) 1994: *Nordiske studier i leksikografi II. Rapport fra Konferanse om Leksikografi i Norden 11. – 14. maj 1993*. Oslo: Skrifter udgivet af Nordisk Forening for Leksikografi. Skrift nr. 2. (Tr. Århus.)
- Helleland, Botolv 1992: *Proprium og leksikografi*. I: Fjeld 1992, 190–199.

- Ihre, Johan 1769: *Glossarium suiogothicum, in quo tam hodierno usu frequentata vocabula, quam in legum patriarum tabulis aliisque ævi medii scriptis obvia explicantur, et ex dialectis cognatis, moesogothica, anglosaxonica, alemannica, islandica ceterisque gothicae et celticae originis illustrantur*. Upsaliæ: Edmannianis.
- Íslensk orðabók handa skólum og almennungi* 1983. Önnur útgáfa, aukin og bætt. Reykjavík: Bókaútgáfa Menningarsjóðs.
- Longman Dictionary of Contemporary English* 1978. Harlow/London: Longman Group Limited.
- Malmgren, Sven-Göran 1994: Sammansättningsmorfologi och lexikografi. I: Garde/Jarvad 1994, 179–184.
- Mel'čuk, I. A. 1978: A New Kind of Dictionary and Its Role as a Core Component of Automatic Text Processing Systems. I: T. A. *Informations* 2, 3–8.
- Ralph, Bo 1992: Ord på väg. I: Fjeld 1992, 28–39.
- Reiger, Charles J. 1978: Computational Linguistics. I: William Orr Dingwall (ed.): *A Survey of Linguistic Science*. Stamford, CT: Greylock Publishers.
- Sigfús Blöndal 1920–24: *Íslensk-dansk ordbog*. Reykjavík/København/Kristiania: Verslun Þórarins B. Þorlákssonar/H. Aschehoug & Co. (W. Nygaard).
- Sigurd, Bengt 1965: Morfologiska mönster i nationalitetsord. *Arkiv för nordisk filologi* 80, 231–276.
- Smith, Raoul N. 1981: *Conceptual Primitives in the English Lexicon*. Waltham, Massachusetts: GTE Laboratories Incorporated TR 81–401.1.
- Spegel, Haqvin 1712: *Glossarium Sveo-Gothicum eller Swensk Orda-book* [. . .]. Lund: A. Habereger, Gub. och Acad. Carol. Booktr.
- Statsnavne og nationalitetsord* 1994. Oslo: Nordisk spåksekretariat/Novus forlag.
- Svensén, Bo 1987: *Handbok i lexikografi. Principer och metoder i ordboksarbetet*. Stockholm: Esselte Studium/Tekniska nomenklaturcentralen.