

GÖTEBORGS UNIVERSITET
SPRÅKDATA
Institutionen
~~Institutionen~~ för språklig databehandling
Department of Computational Linguistics

GU-ISS-77-01

Bo Ralph, Jerker Järberg, Sture Allén

SVENSK ORDBOK OCH
LEXIKALISK DATABAS

Förstudierapport

1977-12-20

Postadress
Språkdata
Norra Allégatan 6
402 22 GÖTEBORGS

Telefon
031 - 11 13 60
11 13 70
11 13 80

1. Sammanfattning	5
2. Bakgrund	5
2.1. Projektets ursprung	5
2.2. Organisatorisk uppläggning	8
2.3. Hittillsvarande förlöpp	11
2.4. Förhållandet Lexikalisk databas-Svensk ordbok	14
2.5. Förhållandet till officiella intressenter	15
3. Rapportens syfte	16
4. Projektens teoretiska aspekter	20
4.1. Teorins roll i ordboksarbetet	20
4.2. Lexikonets roll	21
4.3. Den lexikala enheten	23
4.4. Morfologi i lexikonet	26
4.5. Definition och beteckningsinnehåll	29
4.6. Definitionssystemets struktur	33
4.7. Lexikon och encyklopedi	37
4.8. Urvalsproblematiken	40
5. Målsättning för ordboken	43
5.1. Översikt över forskningsläget	43
5.2. Nålgrupp och användarkrav	45
5.3. Informationens sammansättning	47
5.4. Uppläggning och hantering	49
6. Överföring av ordboken	51
6.1. Ordbokens organisering	51
6.2. Tilläggsinformation i ordboken	52
6.3. Uppslagsord	54
6.4. Ordmodeller för uppslagsord	55
6.5. Ordlistans struktur	56
6.6. Ordboksartiklarnas innehåll	59
6.7. Ordboksartiklarnas uppläggning	61
6.8. Exempel på ordboksartiklar	62
7. Datateknisk metod	67
7.1. Lagringsform	67
7.2. Interaktiva och automatiska arbetsmoment	69
7.3. Maskinell utrustning	70
7.4. Programmering	72
7.5. Sättning	72
Litteraturlösteckning	73

Bo Ralph, Jerker Järborg, Sture Allén

SVENSK ORDBOK OCH
 LEXIKALISK DATABAS

Förstudierapport

1977-12-20

INNEHÅLLSFÖRTECKNING

1. Sammanfattning	3
2. Bakgrund	5
2.1. Projektets ursprung	5
2.2. Organisatorisk uppläggning	8
2.3. Hittillsvarande förlopp	11
2.4. Förhållandet Lexikalisk databas—Svensk ordbok	14
2.5. Förhållandet till officiella intressenter	15
3. Rapportens syfte	16
4. Projektets teoretiska aspekter	20
4.1. Teorins roll i ordboksarbetet	20
4.2. Lexikonets roll	21
4.3. Den lexikala enheten	23
4.4. Morfologi i lexikonet	26
4.5. Definition och teckeninnehåll	29
4.6. Definitionssystemets struktur	33
4.7. Lexikon och encyklopedi	37
4.8. Urvalsproblematiken	40
5. Målsättning för ordboken	43
5.1. Översikt över forskningsläget	43
5.2. Målgrupp och användarkrav	45
5.3. Informationens sammansättning	47
5.4. Uppläggning och hantering	49
6. Uformning av ordboken	51
6.1. Ordbokens organisation	51
6.2. Tilläggsinformation i ordboken	52
6.3. Uppslagsordens status	54
6.4. Ordmodeller för uppslagsord	55
6.5. Huvudlistans struktur	58
6.6. Ordboksartiklarnas innehåll	59
6.7. Ordboksartiklarnas uppläggning	61
6.8. Exempel på ordboksartiklar	62
7. Datateknisk metod	67
7.1. Lagringsform	67
7.2. Interaktiva och automatiska arbetsmoment	69
7.3. Maskinell utrustning	70
7.4. Programmering	72
7.5. Sättning	72
Litteraturförteckning	73
Arbetsrapporter (bilagor)	73
Övrig citerad litteratur	74

1. SAMMANFATTNING

Vid Språkdata har sedan 1977-01-01 bedrivits förstudier för projektet Lexikalisk databas. Projektets teoretiska syfte är att etablera ett välstrukturerat lexikon lagrat i form av ett länkat nätverk (databas). Det praktiska syftet är primärt att databasen skall läggas till grund för en ny svensk enspråkig definitionsordbok, Svensk ordbok. Projektet bedrivs i samarbete med Esselte Studium.

Förstudiens fas a, som avslutas i och med denna rapport, har ägnats åt teoretiska överläggningar, experimentella undersökningar, språkvetenskaplig och datateknisk metodutveckling och mångsidig kontakt- och samarbetsverksamhet. De teoretiska övervägandena har resulterat i en genomarbetad modell för lexikonet. Denna modell har mottagit viktiga impulser från de experimentella undersökningarna, som inriktats på centrala områden; å andra sidan har teoriutvecklingen givit incitament till förnyade experiment. Den parallellt med experimenten fortgående lingvistiska metodutvecklingen har inriktats på kategorivisa bedrivna analyser och bearbetningar. På datasidan har flexibla programsystem för delundersökningar och inkodning utarbetats; arbetet på den länkade lagringsformen utvecklas successivt.

Bland de många viktiga kontakter som har tagits under året skall här nämnas dels Svenska språknämnden, som förklarar sig beredd till konkret samarbete, dels Skolöverstyrelsen, vars planerade invandrarordböcker får viktiga beröringspunkter med vår verksamhet. Esselte Studiums ordboksredaktion har framfört viktiga praktiska synpunkter på projektet under flera arbetssammanträden. Ett nära samarbete med Logoteket har fortlöpande ägt rum med avseende på materialinsamling och materialurval.

Under slutskedet av fas a har Svensk ordbok stått i centrum. Genom att de teoretiska aspekterna konfronterats med praktiska krav har tämligen klara riktlinjer för ordboken kunnat presenteras, främst i denna rapport. Både teoretiskt och praktiskt är projektet väl rustat inför fas b.

Sture Allén och Kjell Johansson bildar ordboksprojektets organisatoriska ledningsgrupp. Vetenskaplig ledare för Lexikalisk databas/Svensk ordbok är Sture Allén. I det dagliga arbetet har Bo Ralph fungerat som vetenskaplig koordinator. Inom projektet har vidare arbetat Sture Berg, Jerker Järborg och Christian Sjögreen. Rolf Gavare och Martin Gellerstam har genom sin ställning inom Logoteket varit kontinuerligt engagerade i planeringen och (delvis) det praktiska arbetet. Rapporten grundas således på flera personers insatser. I huvudsak har den dock formulerats av Jerker Järborg och Bo Ralph.

Till att projektet igångsattes är det faktum att det saknas en fullgod ordbok över modern svenska. Den enda omfattande ordbok som finns över huvud taget, Illustrerad svensk ordbok, börjar bli föråldrad och svarar endast delvis mot de krav som vill ställa på en enspråkig ordbok av modernt snitt. Övriga verk som närmast faller inom genren är starkt specialiserade till sin inriktning och/eller synnerligen begränsade till omfånget (se projektbeskrivningen s. 4 f. för en översikt och kortfattad utvärdering).

Vad (vad som nu heter) institutionen för språklig databehandling i Gøteborg (hädanefter kallad Språkdata) har en rad projekt av lexikologisk natur genomförts under en tioårsperiod. Principiellt betydande arbete har också utförts i viss omfattning av personal vid Språkdata utanför ramen för Språkdatas normala verksamhet. Under denna arbete har idéer om en svensk ordbok väckts på ett naturligt sätt.

Det mest omfattande lexikologiska projektet redovisas i Svensk frekvensordbok (NFO) 1-4. Undersökningen gäller aktuellt svenskt språkbruk såsom det representeras i fem stora muntaltidningar år 1981. Del 1-2 (1979-1980) tar sikte på ordnivån. Den första delen behandlar graford och morfemkomponenter, den andra lemmans morfemsteg. Den spelar en framträdande roll vid analysen såväl som vid presentationen av resultaten. Till ett och samma lemma räknas ordpar med gemensam uttrycksdel och gemensamt flexionsstätt. (För en utförligare beskrivning av lemmabegreppet se Allén, "Förhållandet mellan tal och skrift" och inledningen till NFO 1.) I den tredje delen undersöks ordens kombinatoriska egenskaper. Tonvikten är lagd på det egentliga frekvensbegreppet, dvs. de språkligt mest relevanta kombinationerna (i NFO 2 kallade konstruktioner). NFO 3 ger vid

2. BAKGRUND

Bakgrunden till projektet Lexikalisk databas/Svensk ordbok har tidigare beskrivits i Allén, "Lexikalisk databas. Projektbeskrivning". Här skall några punkter betonas särskilt. Det är också lämpligt att göra några kompletteringar nu när fas a av förstudien har genomförts.

2.1. Projektets ursprung

Den yttersta orsaken till att projektet igångsatts är det faktum att det saknas en fullgod ordbok över modern svenska. Den enda omfattande ordbok som finns över huvud taget, Illustrerad svensk ordbok, börjar bli föråldrad och svarar endast delvis mot de krav man vill ställa på en enspråkig ordbok av modernt snitt. Övriga verk som närmast faller inom genren är starkt specialiserade till sin inriktning och/eller synnerligen begränsade till omfånget (se projektbeskrivningen s. 4 f. för en översikt och kortfattad utvärdering).

Vid (vad som nu heter) institutionen för språklig databehandling i Göteborg (hädanefter kallad Språkdata) har en rad projekt av lexikologisk natur genomförts under en tioårsperiod. Principiellt besläktat arbete har också utförts i viss omfattning av personal vid Språkdata utanför ramen för Språkdatas normala verksamhet. Under detta arbete har idéer om en svensk ordbok väckts på ett naturligt sätt.

Det mest omfattande lexikologiska projektet redovisas i Nusvensk frekvensordbok (NFO) 1-4. Undersökningen gäller aktuellt svenskt skriftspråk såsom det representeras i fem stora morgontidningar år 1965. Del 1-2 (1970-1971) tar sikte på ordnivån. Den första delen behandlar graford och homografkomponenter, den andra lemmen. Lemmabegreppet spelar en framträdande roll vid analysen såväl som vid presentationen av resultaten. Till ett och samma lemma räknas enheter med gemensam uttrycksdel och gemensamt flexionssätt. (För en utförligare beskrivning av lemmabegreppet se Allén, "Förhållandet mellan tal och skrift" och inledningen till NFO 1.) I den tredje delen undersöks ordens kombinatoriska egenskaper. Tonvikten är lagd på den egentliga fraseologin, dvs. de språkligt mest relevanta kombinationerna (i NFO 3 kallade konstruktioner). NFO 3 ger vid

handen att orden i en godtycklig text är frasbundna långt mer än väntat. I ordbokens fjärde del slutligen behandlas ordleden (morfemen) och deras betydelse. Ordleden redovisas på flera nivåer. På de två lägsta definieras ordleden i princip formellt. På den tredje nivån samklassificeras de på fortfarande formella grunder men under iakttagande av ett negativt semantiskt kriterium. På den högsta nivån differentieras olika betydelser på grundval av positiva semantiska kriterier. Vid differentieringsarbetet har en användbar metod utvecklats som syftar till att bestämma (synkronisk) grundbetydelse, varifrån sekundära betydelser kan härledas på ett systematiskt sätt.

I anslutning till detta projekt och med Illustrerad svensk ordbok som utgångspunkt har ett svenskt homograflexikon upprättats. Psalm-boken har genomarbetats efter liknande riktlinjer. Personnamnsstudier bedrivs fortlöpande på bred bas. Resultat från dessa arbeten har publicerats eller är under tryckning.

Andra projekt vid institutionen koncentrerar sig på speciella aspekter av lexikologin. Projektet Svensk definitionsvokabulär har givit en preliminär minimalordlista med vars hjälp andra svenska ord skall kunna definieras entydigt. Projektet Algoritmisk textanalys uppmärksammar särskilt morfologi och syntax inom ramen för datamaskinell analys av autentisk text. Här har bl.a. en formaliserad svensk ordböjnings- och ordbildningsgrammatik utvecklats.

Av annat pågående projektarbete som kan vara av intresse i detta sammanhang kan särskilt nämnas Språksociala studier i en talspråklig databas. Som framgår av namnet kommer talat språk här att stå i fokus. Till det arbete på projektbasis som hittills nämnts kommer så utvecklingsarbetet inom Logoteket, ett nationellt serviceorgan som är en gren av Språkdata. Logoteket tar årligen in maskinläsbara texter på i runt tal tio miljoner löpord. Med dessa texter som grund håller en ordbank på att skapas. I takt med att ordbanken växer ytterligare och programsystemet förfinas erhålls en allt kraftfullare databas som kan spela en central roll vid ordboksframställning.

Det arbete som kort skisserats har i första hand vetenskaplig karaktär. De publikationer som resulterat riktar sig visserligen till allmänheten, men deras natur gör att de ändå har mest att erbjuda den språkvetenskapligt orienterade. Inriktning på mer allmänpraktisk användning har emellertid några ordlistor haft. Detta gäller Alléns "Svensk ordlista" och "Våra viktiga ord". Vi känner nu ett behov av att i vidare omfattning ta till vara den erfarenhet av lexikologiska frågor som vunnits under åren och göra den nyttig för en större grupp medborgare. En god svensk definitionsordbok av rimligt omfång förefaller därvid vara den angelägnaste uppgiften. Denna stämning är en viktig del av projektets bakgrund.

En annan betydelsefull aspekt är Språkdatas tekniska utrustning. Den metodologi som utvecklats för lexikologisk undersökning av stora datamängder förutsätter att datamaskin utnyttjas. Institutionens egen dataanläggning medger avancerad språkvetenskaplig behandling av språkligt material. De tekniska möjligheterna beskrivs närmare i kapitel 7.

Självva den utveckling som Språkdatas organisation genomgått har likaledes påverkat projektplanerna. År 1972 etablerades Språkdata som universitetsinstitution under benämningen avdelningen för språklig databehandling (fortfarande dock med viss administrativ bindning till institutionen för nordiska språk och dåvarande avdelningen för allmän språkvetenskap). Sedan dess har egen dataanläggning installerats (1973) och Logoteket inrättats (1975). Under 1977 har Språkdata blivit självständig institution vid Göteborgs universitet. Den relativa stadga som garanteras härigenom förbättrar naturligtvis ytterligare förutsättningarna för att driva ett omfattande projekt framgångsrikt.

De allt tätare kontakter med Esselte Studium som förekommit på senare tid har blivit ett incitament till att planerna på en fysisk ordbok konkretiserats. Databastänkandet har därmed fått en direkt praktisk målinriktning. Till de antydda språkvetenskapliga erfarenheterna kan nu läggas den praktiska lexikografiska sakkunskap som Esselte Studium representerar.

2.2. Organisatorisk uppläggnig

I projektbeskrivningen (s. 21 f.) gavs en tentativ *tidsplan* för projektet enligt följande:

Förstudie fas a	1977-01-01 - 1977-09-30
Planeringskonferens	1977-09-29/30
Förstudie fas b	1977-10-01 - 1977-12-31
Huvudfas	1978-01-01 - 1982-12-31
Tryckningsfas	1983-01-01 - 1983-06-30

Denna plan presenterades med reservationer. Det har under året visat sig lämpligt att modifiera den i några avseenden. I samförstånd med våra kontaktmän vid förlaget utsträcktes fas a av förstudien till att räckta året ut. Planeringskonferensen uppdelades på tre olika sammankomster med skiftande inriktning. Organisationsfrågor, policyfrågor, vissa lexikografiska frågor och andra frågor av mera principiell natur diskuterades i en trängre krets 1977-11-08 varvid representanter för Esselte Studium och Språkdata deltog. Delvis samma personer samrådde med Bertil Molde, Svenska språknämnden, vid ett sammanträde 1977-11-30 om samarbetsmöjligheter (främst utnyttjande av nämndens material och granskning av lexikografisk output) och några grundläggande lexikografiska frågor. Positiva beslut om samarbete fattades. Ett tredje sammanträde, som snarast får karaktären av konferens med särskilt inbjudna specialister, är utsatt till 1978-02-03.

Fas b sträcker sig nu över första kvartalet 1978. Följaktligen startar huvudfasen 1978-04-01.

Under den gångna perioden har en rad *kontakter* tagits. Det har redan nämnts att Bertil Molde från Svenska språknämnden deltagit i överläggningar med Esselte Studium och Språkdata. Det har som antyttts framkommit, att språknämnden förfogar över intressanta material som kan göras disponibla, och att vissa tjänster kan erbjudas. Det är angeläget att det intresse från språknämndens sida som ådagalagts, liksom nämndens lexikografiska erfarenhet i allmänhet, kan komma projektet till godo på bästa sätt. Tillfogas kan att informella kontakter med företrädare för redaktionen för Svenska Akademiens ordlista förekommit.

Skolöverstyrelsen planerar en omfattande verksamhet (det s.k. LEXIN-projektet) med sikte på att producera en serie tvåspråkiga invandrarordböcker för diverse språkområden (högprioriterade är turkiska och serbokroatiska). Handläggare hos SÖ är Hardy Hedman. Språkdata och institutionen för lingvistik i Stockholm har tillsammans fått i uppdrag att utreda förutsättningarna för ett sådant projekt. Det står redan klart att vad som på längre sikt behövs är tillgång till en databas med egenskaper som ligger nära dem vår databas måste förete. SÖ:s betydelsefulla möjligheter att bidra till materialinsamling samt viss utvärdering och praktiskt arbete ställer LEXIN-projektet i en intressant relation till Lexikalisk databas/Svensk ordbok. Dessutom har de kontakttillfällen som yppats givit möjlighet att testa idéer om den lexikaliska databasen. De synpunkter som framkommit har i några fall visat sig mycket värdefulla.

Sedan flera år upprätthåller Språkdata på många plan intensivt vetenskapligt utbyte med institutionen för semantik vid universitetet i Bielefeld, Västtyskland. Rent allmänt är ordbokstänkandet livligt för närvarande i Västtyskland - flera lexikologiskt orienterade projekt är i full gång, vissa av ansenligt omfång; bl.a. planeras en modern 20-bandig ordbok. Informationsflödet har gått i bägge riktningar, och professorn i semantik i Bielefeld, János Petöfi, har under hösten gästat Språkdata. Därvid har databasprojektet diskuterats. Personlig kontakt har också tagits med Gerhard Wahrigs forskargrupp i Mainz. Wahrig är en välkänd ordboksförfattare och utarbetar för närvarande ett system för datamaskinellt baserade lexikaliska databanker. Medan vi har det internationella perspektivet kan också nämnas att delar av Språkdatas personal knutit viktiga internationella kontakter i samband med en arbetskonferens om språklig databehandling som höllits i Palo Alto, USA, i juni 1977. Vid de "nordiska datalingvistikdagar" som anordnades i Göteborg under oktober månad presenterades vidare huvuddragen av projektet Lexikalisk databas i föredragsform (för en sammanfattning av föredraget se Ralph, "Projektet Lexikalisk databas"). De reaktioner som därvid framkom har beaktats under det fortsatta arbetet. Slutligen har Språkdata nyligen gästats av en ledande brittisk språksociolog, Peter Trudgill, universitetet i Reading, varvid en del specialfrågor diskuterades.

Härtill kan fogas att Språkdata under året representerats vid ett stort antal relevanta konferenser på såväl internationell som nationell basis. Dessa har gällt både språkvetenskapliga och data-tekniska eller informationsvetenskapliga ämnen, i några fall en kombination därav. De personliga kontakter som knyts vid sådana tillfällen kan erfarenhetsmässigt få mycket stor betydelse. På åtskilliga punkter har förarbetet för det här aktuella projektet direkt påverkats i gynnsam riktning.

Den modifierade planen implicerar att produktionsfasen tar sin verkliga början 1978-04-01. Fram till dess, från årsskiftet räknat, bör en del avgränsade förundersökningar utföras, som en del av förstudiearbetet fas b. (Exempelvis har vissa kompletterande klassifikationer och otraditionella men användbara markeringar i den maskinläsbara versionen av SAOL₁₀ planerats. En del andra genomgångar av lexikaliska kategorier är också önskvärda.)

Det är att märka att också produktionsfasen kommer att innehålla flera typer av utvecklingsarbete. Framför allt kan fastställandet av definitionsvokabulären (se avsnitt 4.6) nämnas. En av huvudpoängerna med den tänkta metodologin är emellertid just att arbetet t.o.m. på produktionsstadiet kan bedrivas "på tvären". Det blir inte frågan om att skriva en ordboksartikel i taget, från början till slut, fullständig i varje detalj. I stället kan vissa typer av information med fördel implementeras i ett sammanhang, rakt igenom hela materialet. En första uppgift - vid sidan av materialurvalet - blir att avgöra vilka typer av information som särskilt har egenskaper som lämpar sig härför. Det är med andra ord ingen traditionell form av lexikografi - med papper och penna utbytt mot textskärm och dator - som kommer att bedrivas, utan hela förfarandet kommer att genomgripande förändras. Därför kan produktionsfasen ta sin början redan innan alla detaljer i tillvägagångssättet slutgiltigt fastlagts. Huvudsaken är att strategin i stort står klar och att lösningar finns inom synhåll. Det kan tilläggas att vissa experimentella arbeten under förstudien givit resultat som säkert kommer att betraktas som definitiva och som därför föreligger redan vid produktionsfasens start.

2.3 Hittillsvarande förlopp

Proportionellt sett har en mycket stor del av de muntliga diskussioner som förts internt vid Språkdata gällt projektets övergripande filosofi. Därvid har både språkvetenskapliga och mera datorinriktade problem uppmärksammats. Många preliminära idéer framkom redan i Alléns projektbeskrivning. Andra har berörts i olika underhandsrapporter som tillkommit under förstudien. När det gäller rent språkteoretiska frågor förs ett ganska utförligt resonemang i kapitel 4 nedan. I de första avsnitten av kapitel 7 tas några data-tekniska premisser upp. Vissa allmänt teoretiska och principiella frågor är också huvudämnet i promemorior som Ralph, "Förutsättningar" och "Ordböcker och språkmodeller", Allén, "Om lexikografiska definitioner" och "Förslag till struktur för en lexikalisk databas". Våra gemensamma samtal, som utgjort en stor del av det samlade arbetet, har inte direkt dokumenterats skriftligt men har kanaliserats i den lingvistiska underström som säkert kan urskiljas såväl i flera delrapporter som i föreliggande framställning.

Diskussionerna har mest kretsat kring tre huvudfrågor: (1) urval och handhavande av *material*, (2) urval och organisation av *information*, (3) *presentation* av resultat.

När det gäller materialsidan spelar Logoteket en central roll. Dess verksamhet kan delvis anpassas efter databasens/ordbokens behov. Inom Logotekets ram har en inventering av materialproblemet genomförts med sikte på databasen/ordboken (se Gellerstam, "Välja sina ord"). Om materialurval se vidare avsnitt 4.8. Till handhavandet av materialet räknas främst lagringen och presentationen vid analysarbetet. Om lagringsformen resonerar Allén i "Förslag till struktur för en lexikalisk databas". Se även avsnitt 7.1 nedan. På programmeringssidan har sonderingar i befintliga (kommersiella) databassystem utförts. Preliminära ståndpunkter har också fattats beträffande våra speciella behov (jfr problemkatalogen Gavare & Sjögren, "Databaser" och avsnitt 7.1 nedan). Visst förberedande programmeringsarbete pågår (se avsnitt 7.4).

Det mesta arbetet har ägnats informationsaspekten: vilka informationstyper skall uppmärksammas; vilken teoretisk bas skall väljas för urvalet, analysen och organisationen av informationen? För att

få ett fast beslutsunderlag har tre ganska omfattande experimentella arbeten utförts.

För det första har ett ansevärt antal verb inmatats i datamaskin och analyserats efter vissa givna principer. Syftet har ytterst varit att göra en framstöt mot definitionsvokabulären, som vi anser skall utgöra den principiella stommen i definitionerna. Goda indikationer har erhållits angående det konvertibilitetssystem som skisseras i avsnitt 4.6. Undersökningen anvisar också en framkomlig metodologisk väg för definitionsarbetet i allmänhet. Ytterligare resultat av helt annat slag kan så småningom utvinnas från denna undersökning, men målsättningen har för tillfället inskränkts till den just angivna. För vidare upplysningar om verbanalys se Ralph, "Verb för verb", "Format för verbparafraser" samt "Från verparafraser till konvertibilitetssystem".

En andra verbundersökning har likaså avkastat en del intressanta resultat, låt vara att den bör drivas vidare under fas b av förstudien. Några hundratal verb har försetts med maximala (nominala) argumentramar och inmatats i en välstrukturerad form i datorn. Argumenten har klassats syntaktiskt och semantiskt. Genom att lagringsformen är smidig kan materialet inspekteras från flera olika synvinklar med hjälp av enkla omsorteringar. På detta sätt har kunskap vunnits om syntaktiska konstruktioner m.m., men framför allt har en god inkörspport till definitionernas format i allmänhet upptäckts. Denna undersökning beskrivs utförligare i Järborg, Ralph & Sjögren, "Undersökning av argumentramar" och Ralph, "Kasusbeteckningar för ytargument".

Den tredje undersökningen är än så länge något blygsammare till omfånget men har visat sig lovande. I den tas mera direkt sikte på ordens definitioner. Dels är syftet att så småningom inringa de återkommande enheterna, dvs. de enheter som till slut kommer att ingå i definitionsvokabulären, dels att precisera relationen mellan de ytstrukturella konstruktionsegenskaperna och den djupare liggande semantiska strukturen. Som ett särskilt resultat väntas också en provkarta på definitionsformat för olika typer av ord utfalla. Härom se vidare Järborg & Ralph, "Mot definitionen".

Under senare delen av perioden har presentationsfrågorna allt mer ryckt fram i förgrunden. En synnerligen viktig punkt är därvid uppslagsordens status i ordboken. Olika alternativ har diskuterats i Berg, Järborg & Ralph, "Synpunkter på stickord" och Allén, "Om uppslagsord i Svensk ordbok/Lexikalisk databas". Vissa allmänna presentationsproblem tas upp i Järborg & Ralph, "Ordboksartiklarna: ordbildnings- och ordstruktursynpunkter". Dessa frågor har framför allt aktualiserats av att man på Esselte Studiums håll sysselsatt sig med dem (jfr Svensén, "Några funderingar om typografi m.m. i Svensk ordbok"). Från databasens synpunkt är dessa frågor tidsmässigt egentligen inte akuta, rent tekniskt sett, inte från ordbokens heller. Strängt taget skulle man (fortfarande rent tekniskt sett) kunna bestämma uppslagsordens form, vissa indelningsgrunder och presentationsordningar när arbetet i övrigt börjar fullbordas, tack vare den välstrukturerade lagringsformen. Alla nödvändiga data finns under alla omständigheter tillgängliga i databasen. Att beslut skall fattas har framför allt varit ett önskemål från förlagets sida. För tillfället får vi dock konstatera att det från språkvetenskaplig synpunkt finns några tänkbara, till stor del likvärdiga, förfaranden. För ett mer detaljerat resonemang hänvisas till avsnitt 6.7.

Som ett mycket betydelsefullt moment i det provarbete som skisserats ovan framstår metodutvecklingen. Det har visat sig helt nödvändigt att gå fram kategorivis vid analysen under produktionsfasen, något som enkelt medges inom ramen för den planerade metoden i stort. En lista på nödvändiga specialundersökningar har också utkristalliserats. Vidare har ett mycket smidigt programsystem utvecklats (se Sjögren, "Dokumentation av programsystem för datorbaserad registerhantering"). Det kan förutses få vidsträckt användning under projektets produktionsfas. Vi börjar också få en god uppfattning om hur det praktiska arbetet måste organiseras rent konkret.

Till arbetets hittillsvarande förlopp hör noga taget även de kontakter som etablerats. Det är ingen överdrift att säga att mycken möda lagts ner på att konfrontera projektplanerna med andra forskares erfarenhet. Många värdefulla synpunkter har därvid erhållits. De viktigaste kontakter som tagits nämns ovan i avsnitt 2.2.

2.4. Förhållandet Lexikalisk databas-Svensk ordbok

I denna rapport kastas blickarna än mot den lexikaliska databasen, än mot den (svensk-)svenska ordboken. Det är då viktigt att komma ihåg att vi talar om två projekt: *Lexikalisk databas* och *Svensk ordbok*. Språkdatas projekt Lexikalisk databas är ett vetenskapligt projekt som syftar till att etablera en databas med lexikalisk information. Esselte Studiums projekt Svensk ordbok har som direkt målsättning att producera en ordbok på basis av ett magnetbandsmanuskript genererat ur databasen. I samarbete mellan Språkdata och Esselte Studiums ordboksgrupp fastställs *ordbokens* innehåll och organisation; rent principiellt är ordboksunderlaget en delmängd framtagen ur databasen. Språkdata tar ansvar för den språkvetenskapliga (inklusive datatekniska) sidan av uppgiften, Esselte Studium för den trycknings- och förlagstekniska. I praktiken fortsätter naturligtvis arbetet i den anda av samförstånd och ömsesidig lyhördhet som i högsta grad rått hittills.

Rent praktiskt kommer databasen under överskådlig tid att sammanfalla med ordboksunderlaget i allt väsentligt. (Eventuellt kan enheterna i databasen redan från början behöva förses med i vissa avseenden fylligare uppgifter än vad som kommer till uttryck i ordboken, men detta är en detaljskillnad; extrauppgifterna kommer dessutom indirekt ordboken till godo.) Med tiden torde databasen och ordboken komma att få olika omfång. Ordboken behåller i stort sett sin givna målsättning, medan databasen kontinuerligt utökas med ny information. Databasen blir ett mångsidigt användbart redskap som garanterar förenklad förnyelse även för ordbokens del. I det längre perspektivet möjliggörs också framtagningar av strängt taget vilka delmängder som helst genom den välstrukturerade lagringsformen.

Detta är inte rätt plats att diskutera det exakta förhållandet mellan Lexikalisk databas och Svensk ordbok i alla upptänkliga avseenden. Ovan har några överordnade principer berörts. Mera utpräglade kontraktsfrågor tas inte upp här.

2.5. Förhållandet till officiella intressenter

Det har nämnts ovan (se avsnitt 2.2) att det föreligger en del beröringspunkter mellan Lexikalisk databas/Svensk ordbok å ena sidan och vissa myndigheter å den andra. De som märks i första rummet är Skolöverstyrelsen, Svenska språknämnden och Svenska Akademien (främst SAOL:s redaktion).

Här skall endast slås fast att det är mycket angeläget att utveckla de kontakter som redan finns. Det är viktigt att vi får tillstånd en god avstämning mellan olika projekt. Vi har kunnat urskilja en positiv inställning till informationsutbyte av olika slag. Eftersom det dessutom finns möjlighet att material och i någon mån hjälp med visst arbete kan påräknas från en del håll, finns det all anledning att noga observera dessa relationer.

För huvudrapporten gäller generellt att den inte fyller gängse vetenskapliga krav på källhänvisningar. (Detsamma gäller i växlande grad för de enskilda arbetsrapporterna.) Påståenden bekräftar normalt inte genom litteraturreferenser. Vanligen görs hänvisningar endast av praktiska skäl, nämligen till andra arbeten som vi själva skrivit, mestadels målläsningsmaterial inom projektets ram.

Gången hela rapporten upprätthåller principiellt den tidigare nämnda bedömningen mellan Lexikalisk databas och Svensk ordbok, som teoretiskt är mycket strikt. Rapporten skall emellertid läsas med den förutsättningen att uppmärksamheten hela tiden är praktiskt inriktad mot ordboken. Arbetet måste läggas upp så att ordbokenns behov täcks i första hand. Ordboken ter sig direkta utgångspunkt i den lexikaliska databasen och är i sista hand helt beroende av denna. Inriktningen på ordboken förser projektet med en naturlig ram och anger automatiskt en lämplig arbetsordning.

3. RAPPORTENS SYFTE

Föreliggande rapport är en lägesbeskrivning vid utgången av förstudiens fas a (i den modifierade planen). Den ger en bakgrund till projekten Lexikalisk databas/Svensk ordbok (kapitel 2), en sammanfattning av den språkteoretiska grundsyn som projekten vilar på (kapitel 4), vissa förutsättningar för ordboken (kapitel 5), en så långt möjligt samlad bild av ordboksplanernas mera konkreta sidor (kapitel 6) samt, slutligen, en antydan om datorns roll i arbetet jämte en presentation av den erforderliga metodiken (kapitel 7).

Rapporten utgör i stora drag en resonerande framställning. På många punkter har emellertid detaljstudier bedrivits under året, vars resultat redovisats i delrapporter. Dessa bifogas och listas som en särskild avdelning i den för övrigt sparsamt hållna litteraturförteckningen. Allmänt kan sägas att diskussionen i denna skrift görs mer summarisk närhelst specialrapporter finns att referera till. Proportionellt mer plats har ägnats åt sådana beslut som annars dokumenterats knapphändigt eller inte alls.

För huvudrapporten gäller generellt att den inte fyller gängse vetenskapliga krav på källhänvisningar. (Detsamma gäller i växlande grad för de enskilda arbetsrapporterna.) Påståenden *beläggs* normalt inte genom litteraturreferenser. Vanligen görs hänvisningar endast av praktiska skäl, nämligen till andra arbeten som vi själva skrivit, mestadels målinriktade skrifter inom projektets ram.

Genom hela rapporten upprätthålls principiellt den tidigare nämnda bodelningen mellan Lexikalisk databas och Svensk ordbok, som teoretiskt är mycket strikt. Rapporten skall emellertid läsas med den förutsättningen att uppmärksamheten hela tiden är praktiskt primärt inriktad mot ordboken. Arbetet måste läggas upp så att ordbokens behov täcks i första hand. Ordboken tar sin direkta utgångspunkt i den lexikaliska databasen och är i sista hand helt beroende av denna. Inriktningen på ordboken förser projektet med en naturlig ram och anger automatiskt en lämplig arbetsordning.

Nästan alla existerande ordböcker saknar explicita uppgifter om vilket teoretiskt underlag vederbörande ordbok har. Detta är en allvarlig brist som Svensk ordbok ej skall vara behäftad med. Det är en uppenbar styrka om en välgrundad och genomgående modell kan åberopas på en rad punkter. Varje ordbok behöver en teoretisk bas, men behovet accentueras naturligtvis i takt med ambitionerna. Det står klart att Svensk ordbok är en ambitiös ordbok. Bland de uppgifter som särskilt kräver enhetliga grepp märks främst *definitioner* och *konstruktionssätt*, men även *morfologisk* och *uttrycksmässig struktur*. Vidare förefaller det oss i högsta grad önskvärt att hela *presentationen* ingår i ett väl integrerat system.

Med den utveckling den teoretiska lingvistikens genomgått på senare tid blir lexikografi inte enbart ett förståndigt tillämpande av språkvetenskaplig kunskap på en informationsmängd som ordnas för i första hand praktiskt bruk. I och med att lexikologin har fått en centralare ställning i språkteoretiska sammanhang ställs det krav på språkteorierna att de skall rymma lexikologiska aspekter. Sedan detta synsätt börjat anammas har det utvecklats teorier vilka framstår som bättre instrument vid lexikografiskt arbete, eftersom teorierna numera avses erbjuda en helhetsbild över språket. Samtidigt står det klart att lexikonet vetter mot flera komponenter inom den totala språkmodellen. Detta senare är ett teoretiskt grundat argument för att införa flera typer av information i ordboken (som i någon mening bör spegla det teoretiska lexikonet). Lexikografins teoretiska aspekter har uppmärksamrats under förstudien men har dessutom varit av centralt intresse för Språkdatas personal under lång tid. (Vi kan inte i denna skrift generellt anföra den ganska långa rad av vetenskapliga publikationer i vilka vi kan sägas ha dokumenterat vårt intresse på denna punkt.) Flera hithörande frågor av grundläggande natur utreds relativt detaljerat i kapitel 4.

Då denna rapport trots de ganska utförliga teoretiska resonemangen (som ansetts nödvändiga) är emfatiskt inriktad på den praktiska slutprodukten, skall några tänkta karakteristika för ordboken sammanfattas här. Projektbeskrivningen (s. 10 ff.) gav några preliminära riktlinjer för ordboken. De som här ges är en vidareutveckling av dem och kan anses mindre preliminära, om också inte slutgiltigt bindande. Rapporten, sådan den är utformad i det följande, är ämnad

att närmare utveckla och motivera dessa riktlinjer. (Det språk-
teoretiska resonemanget har som antyttts särskilt koncentrerats
till kapitel 4. Den fysiska ordboken fokuseras i kapitel 5-6.
Slutligen diskuteras metodologin ytterligare i kapitel 7. Punkter-
nas inbördes ordning i uppställningen på följande sida följer i
stora drag samma relativa ordning.)

- (1) Sammanfattande arbetet i ordboken principieilt ett språk-
teoretiskt arbete (varigenom förutom ord även morfer och fraser kan
inledas).
- (2) En tydlig beskrivningsnivå för uppslagsarbetet inom
lexikonmodellens ram.
- (3) Beskrivning vid accepterandet av sammansättningar och av-
ledningar; gäller i synnerhet förutsägbara bildningar, som i
alla händelser får rudimentär behandling.
- (4) Särskilt alfabetisk ordning.
- (5) En uppslagsform per artikel.
- (6) Särskild ordning mellan underavdelningar.
- (7) Särskilt deklarerad teoretisk grund.
- (8) Dessa språkliga regelbundenheter sammanfattade i inledning
eller appendix.
- (9) Stark tonvikt på definition; preciserade definitionstyper.
- (10) Särskild definitionsvokabulär om 3000-4000 ord, som redovisas
explisit.
- (11) Förstärkt morfologisk och syntaktisk information, koncentrerat
avdelad.
- (12) Användande av fraseologi (Lilienalvik).
- (13) Särskild formell notationsapparat.
- (14) Överskådlig typografi.
- (15) Särskilt omfattande (ett band).
- (16) Automatisk fotosättning via magnetband.
- (17) Magnetbandet framtaget på grundval av information som lagrats
i en välstrukturerad databas.
- (18) Inkodning i fixt format direkt på textskärm.
- (19) Möjlighet till kontinuerlig uppdatering av ordboken genom
den direkta association till den flexibla databasen.

Några karakteristiska egenskaper hos Svensk ordbok

- (1) Ett centralt allmänt ordförråd på upp till 100 000 explicita uppslagsord i ett integrerat presentationssystem.
- (2) Utnyttjande av autentiskt material, tillhandahållet genom Logotekets försorg.
- (3) Den grundläggande enheten i ordboken principiellt ett språkligt tecken (varigenom förutom ord även morfem och fraser kan introduceras).
- (4) Preciserad beskrivningsnivå för uppslagsenheten inom lemma-lexemmodellens ram.
- (5) Restriktivitet vid accepterandet av sammansättningar och avledningar; gäller i synnerhet förutsägbara bildningar, som i alla händelser får rudimentär behandling.
- (6) Strikt alfabetisk ordning.
- (7) En uppslagsform per artikel.
- (8) Reglerad ordning mellan underavdelningar.
- (9) Explicit deklarerad teoretisk grund.
- (10) Vissa språkliga regelbundenheter sammanfattade i inledning eller appendix.
- (11) Stark tonvikt på definition; preciserade definitionstyper.
- (12) Särskild definitionsvokabulär om 3000-4000 ord, som redovisas explicit.
- (13) Förfinad morfologisk och syntaktisk information, koncentrerat meddelad.
- (14) Angivande av fraseologi (idiomatik).
- (15) Enkel formell notationsapparat.
- (16) Överskådlig typografi.
- (17) Hanterligt omfång (ett band).
- (18) Datorstyrd fotosättning via magnetband.
- (19) Magnetbandet framtaget på grundval av information som lagrats i en välstrukturerad databas.
- (20) Inkodning i fixt format direkt på textskärm.
- (21) Möjlighet till kontinuerlig uppdatering av ordboken genom dess direkta association till den flexibla databasen.

4. PROJEKTETS TEORETISKA ASPEKTER

Etableringen av en lexikalisk databas är primärt en teoretisk uppgift, eftersom det i princip rör sig om att konstruera en modell av språkbrukarens interna lexikon. En dylik modellkonstruktion är emellertid ett stort arbete, som ytterligare försvåras av att inga självklara angreppspunkter anmäler sig. Av denna anledning är det ett stöd för projektet Lexikalisk databas att även ha en praktisk målsättning i form av en svensk ordbok. Den praktiska inriktningen kan motverka teoretisk handlingsförlamning på punkter där indicier saknas för vad som är "rätt" teori. Den praktiska tillämpningen framtvingar ett beslut och bestämmer delvis dess innehåll, liksom den teoretiska uppläggningsen kan ge en erforderlig konsekvens och stadga i ordboksarbetet.

De problemtyper som diskuteras nedan kommer följaktligen att i stor utsträckning vara gemensamma för Lexikalisk databas och Svensk ordbok. I stort sett tas sådana teoretiska frågor upp som aktualiseras genom den praktiska målsättningen. De preliminära avgöranden som träffas kommer däremot delvis att vara alltför ambitiösa i förhållande till ordbokens omedelbara konkreta behov. För tillämpningen av de teoretiska ställningstagandena hänvisas till avsnitt 6.

4.1. Teorins roll i ordboksarbete

Utarbetandet av en ordbok aktualiserar praktiskt taget alla språkvetenskapliga problemställningar, och därtill på ett sätt som gör att problemen inte kan angripas ett i sänder utan måste lösas i ett sammanhang. De viktigaste problemtyperna är knutna till *ordboksenhetens utformning, urvalet av ord och definitionssystemet*. För vart och ett av dessa områden (som icke är oberoende av varandra) krävs ett samlat teoretiskt grepp för att vinna enhetlighet och konsekvens.

Det räcker emellertid inte att internt ha teoretisk medvetenhet. Åtskilliga ordböcker har åstadkommit av framstående forskare med betydande språkkänsla och hög teoretisk medvetenhet. Resultatet kan ändå verka oöverskådligt och inkonsekvent, beroende på att den teoretiska modellen aldrig redovisats tillräckligt explicit. Detta kan medföra dels att framställaren av ordboken icke upptäcker sina egna inkonsekvenser, dels att läsaren får svårigheter att sätta

sig in i redigeringsprinciperna. Vi menar alltså att en ordbok måste byggas på en *explicit* teoretisk modell, som i den färdiga produkten bör redovisas på lämpligt sätt, dels direkt i inledning och/eller appendix, dels indirekt genom systematiskt genomförda och lättförståeliga redigeringsprinciper. (Se vidare avsnitt 6 nedan samt Allén, "Lexikalisk databas. Projektbeskrivning" s. 13 och Ralph, "Förutsättningar".)

En väldefinierad teoretisk utgångspunkt kan vara ett stöd också med avseende på läsarnas förväntningar. Det är uppenbarligen ett faktum att många av en ordbok med vetenskapliga pretentioner väntar sig definitiva, objektivt "sanna" svar på sina frågor. Å andra sidan kan fullständig objektivitet och neutralitet givetvis icke uppnås vad beträffar urvalet eller definitionerna av (t.ex. värde-laddade) ord. Användandet av en explicit teoretisk modell tvingar språkforskaren att specificera sina motiv för ett visst val av ord och definition, främst inför sig själv men i princip även inför läsaren. Härigenom skapas förutsättningar för att ordboken skall kunna fylla de mera grundläggande vetenskapliga kraven på saktighet och kontrollerbarhet.

4.2. Lexikonets roll

Inom modern lingvistik tänker man sig den abstrakta språkförmågan ("competence" enligt Chomsky) som dels ett system av regler, en *grammatik*, dels en lista över enheter som kan ingå i grammatiken, ett *lexikon*. Härav är det grammatiken, och då särskilt syntaxen, som tilldragit sig det största intresset. Enligt vissa forskare bör lexikonet systematiskt rensas från alla regelbundenheter, vilka förtecknas särskilt i ett speciellt regelsystem ("redundancy rules"). Målet är givetvis att vinna största möjliga ekonomi i beskrivningen, vilket i den gängse vetenskapstraditionen anses äga både ett vetenskapligt egenvärde och psykologisk relevans. I en språkbeskrivning av angivet slag är alltså lexikonet närmast en lista över oregelbundenheterna – det oförutsägbara, byggstenarna – i språket, medan alla större eller mindre regelbundenheter ingår i grammatiken.

Det sagda gäller i första hand den ideala individuella språkbrukaren (som egentligen inte kan existera). I praktiken är det väl så

viktigt att abstrahera fram en kollektiv grammatisk modell som gäller för ett större språksamfund. I det senare fallet kan man inte göra anspråk på psykologisk relevans i enskilda detaljer; samfund har ingen "själ". Denna reservation bör man ha i minnet då det i det följande talas om psykologisk relevans. I praktiska sammanhang (som här) betyder detta att pragmatiska hänsyn trots allt ofta blir avgörande.

Det är viktigt att lägga märke till att den ovan angivna språksynen innebär att orden uppfattas som helt konstanta och välbestämda enheter vilka kan kombineras fritt (enligt grammatikens regler) både i syntaktiskt och semantiskt avseende. (Undantag brukar göras för fraseologiska förbindelser med säregen betydelse, s.k. idiom.) En dylik statisk och atomistisk uppfattning av lexikonets roll verkar dock mindre rimlig ur språkpsykologisk synvinkel. Det finns åtskillig närsyntax som är starkt beroende av det exakta ordvalet; jämför t.ex. de kända svårigheterna att i ett främmande språk välja rätt prepositionsfras till ett givet huvudord.

Uppenbarligen finns också hela skalan av semantiska bindningar mellan ord, från den absoluta semantiska självständigheten (normalfallet) över den partiella avhängigheten till det totala beroendet (i vissa idiom). Ett välkänt exempel på partiell semantisk bindning är *köpa för(/mot) x*, där argumentet *för x* har den för verbet *köpa* (och vissa andra, t.ex. *sälja*) specifika uttolkningen 'slags värde = x'. Liknande resonemang kan föras t.ex. om vissa verb + partikelkonstruktioner. Således verkar det som om en psykologiskt rimlig lexikonmodell inte kan avgränsas till att vara en lista över ord utan även måste innehålla syntaktiska uppgifter. Den skarpa gränsen mellan det statiska lexikonet och den dynamiska (generativa) grammatiken kan och bör ej upprätthållas. (Man kan lägga märke till att den mänskliga hjärnans processmöjligheter är underlägsna dess lagringskapacitet, något som direkt bör påverka språkmodellens utformning på den just nämnda punkten.)

Det sagda innebär att det lexikon som konstrueras inom projektet måste baseras på en syntaktisk modell, vilken i enlighet med resonemanget i avsnitt 4.1 skall anges explicit. För vårt syfte förefaller en variant av s.k. kasusgrammatik mycket lämplig (se Ralph,

"Förutsättningar" och "Kort om kasus"). Kasusgrammatik är förmånlig på grund av att den på det syntaktiska planet har överskådliga samband mellan djupstruktur och ytstruktur, samtidigt som dess semantiska distinktioner (semantiska "roller") är lagom detaljerade för en praktisk analys. (Vid verkligt djup-semantisk analys däremot har kasusgrammatik vissa svagheter.) Sammanfattningsvis är det just orienteringen mot ytstrukturen som i våra ögon gör kasusgrammatiken praktiskt överlägsen andra modeller. All information i en ordbok för allmänt bruk bör likaså vara ytorienterad eftersom det är den fysiskt manifesterade sidan av språket som språkbrukaren normalt är medveten om.

En för svenskan fungerande kasusgrammatisk modell är under utarbetande och kan väntas föreligga inom kort. Vid utarbetandet, som försiggår genom såväl teoretiska som pragmatiska undersökningar (för de senare se Järborg, Ralph & Sjögreen, "Undersökning av argumentramar" och Järborg & Ralph, "Mot definitionen"), eftersträvas en kasusuppsättning som kan utgöra en kompromiss mellan hanterbart omfång och semantisk diskriminationsförmåga (se Ralph, "Kasusbeteckningar för ytargument").

Den syntaktiska modellen skall i första hand användas för beskrivning av lexikonenheternas konstruktionssätt. Härjämte torde den vara betydelsefull dels för behandlingen av semantiskt bundna konstruktionssätt (jfr ovan), dels, indirekt, för den syntaktiska utformningen av definitionstyperna. Definitionerna formuleras på vanligt svenskt språk, med ord, och alltså inte i något symbol-språk. Den kasusgrammatiska analysen bör därför ge en direkt indikation om definitionens format i varje enskilt fall.

Här må inskjutas att en enkel typ av kasusgrammatik ofta implicit har anammats i traditionella ordböcker (typ: *ge ngn ngt* el. *ge ngt till/åt ngn*). Användningen i Lexikalisk databas avses emellertid bli mera utförlig, upplysande och konsekvent.

4.3. Den lexikaliska enheten

I princip bör den enhet som förtecknas i lexikonet vara det språkliga tecknet i de Saussures mening: föreningen av ett (begreppsligt sammanhållet) innehåll med en (relativt enhetlig) uttrycksform.

För tecknets uppträdande i produktiv ordbildning se avsnitt 4.4. I praktiken motsvarar det språkliga tecknet oftast ett ord, men undantag finnes (jfr nedan).

För både innehålls- och uttryckssidan gäller att avgränsningsproblem uppträder mellan två potentiella enheter och sammanhållningsproblem uppträder inom en potentiell enhet. Sådana problem är givetvis lättare att lösa på uttryckssidan än på innehållssidan. När det gäller att avgöra huruvida böjningsformerna av ett ord tillhör samma enhet krävs det viss språkvetenskaplig skolning för att inse problemets existens. Ett närbesläktat men något svårare problem är om samma enhet *tänk-* kan anses föreligga i *tänka*, *tänkesätt*, *tänkbar*. Om man svarar ja på denna fråga innebär detta dels att enheten *-sätt* igenkännes och identifieras med ordet *sätt*, dels att den morfologiska kombinationen av enheterna betraktas som regelbunden. Men det innebär också att avledningsändelsen *-bar* 'möjlig att ...' igenkännes och att avledningen betraktas som regelbunden. Härmed har också avledningsmorfem principiellt erkänts som lexikala enheter i teoretisk mening, dvs. avgränsade enheter med *ett* begrepp kopplat till *ett* uttryck. Resonemanget kan givetvis även genomföras för prefix etc. Så långt semantiska och morfologiska regelbundenheter kan urskiljas finns alltså argument för att uppta även vissa enheter mindre än ord (= morfem) som lexikala.

I motsatta ändan av skalan ligger det icke ovanliga fallet att en viss kombination av ord — som är helt regelbunden i syntaktiskt avseende och sammansatt av *ord* vars betydelse är otvetydig och som ändå kommer att få redovisas — dock har en samlad betydelse som ej kan härledas enligt normala regler. Som exempel på dessa s.k. idiom kan nämnas *kasta yxan i sjön* 'resignera'. Här uppträder föreningen av uttryck och innehåll först på flerordsnivån. Följaktligen måste i sådana fall enheter större än ord erkännas som lexikala. Den teoretiska distinktionen är av vikt, men därmed inte sagt att idiomerna skall bilda alldeles egna stickord i ordboken.

Enligt resonemanget omfattar nu lexikala enheter dels ord (vanligast), dels morfem (vissa regelbundna affix), dels vissa ordkombinationer (idiom). Emellertid tillkommer också det faktum att

lexikonenheten icke alltid har välbestämda gränser gentemot sin syntaktiska närmkontext (som påpekades i föregående avsnitt). För åtskilliga ord gäller att de kan ha ett konstruktionssätt som ger ordet i fråga en speciell, svårförutsägbar semantisk specificering. Fallet *köpa för (/mot) x* är redan nämnt: att *för x* betecknar det värde som finns implicit i definitionen av *köpa* är svårförutsägbart. (Normalbetydelsen av *för x* är 'som gynnar x'.) Konstruktionen *höra upp* borde "algebraiskt" sett kunna ha något dussin betydelser (tre huvudbetydelser av *höra* gånger kanske fyra huvudbetydelser av *upp*, om man följer den morfologiska semantiska analys som utförts i NFO 4), men endast två är realiserade: 'lyssna ordentligt' respektive 'sluta'. I konstruktionen [*x har*] *för avsikt* [*att*] är det inte självklart att *för avsikt* har betydelsen 'som sitt (= x:s) syfte' etc. I dessa exempel, som lätt kunde mångfaldigas, rör det sig om ett semantiskt "tungt" ord (verb, substantiv eller adjektiv), utökat med ett "relationsord" (preposition, verbpartikel etc.). Skillnaden gentemot idiomtypen består i att det tunga ordet nu har en av sina huvudbetydelser och relationsordet åtminstone en av sina bibetydelser. Det förefaller därför som om den diskuterade typen, hädanefter kallad *specifika konstruktioner* bör betraktas som ett lexiko-syntaktiskt *tillägg* till huvudordets lexikala enhet. Man skulle kunna tala om "en-och-en-halv-enheter". I en ordbok som (enligt 4.2) vill beakta den dynamiska aspekten kan denna kategori ej nonchaleras. Detta är inte mer anmärkningsvärt än att presentationen av derivationsmorfologi implicit ger en redovisning av delar mindre än ordet och sambandet dem emellan.

Sammanfattningsvis måste vi alltså i semantiskt avseende räkna med följande typer av lexikala enheter: *ord* (av varierande komplexitet), *morfem*, *idiom* och (som tillägg till ord) *specifika konstruktioner*. De återgivningsproblem som dessa enheters olika status ger upphov till kommer att diskuteras i avsnitt 6, där också förslag till den praktiska utformningen ges.

Inledningsvis berördes uttryckssidan av det språkliga tecknet, och det konstaterades att förändringar i uttrycket på grund av böjning av typen *stolen - stolar* är så triviala att de på intet vis ger anledning ansätta olika lexikala enheter. Tvärtom kan ett visst böjningsmönster (paradigm) begagnas som ett särskiljande kriterium

på uttryckssidan. Böjningsmönster är jämte syntaktiskt uppträdande huvudkriteriet för lemmabegreppet, vid sidan av allmän uttrycksmässig likhet. Vid återgivningen av de lexikala enheterna kan antingen de särskiljande dragen på uttryckssidan (lemma eller motsvarande) eller de särskiljande dragen på innehållssidan (huvudbetydelser, lexem) rankas högst. För en diskussion i frågan se Berg, Järborg & Ralph, "Synpunkter på stickord" och Allén, "Om uppslagsord i Svensk ordbok/Lexikalisk databas".

4.4. Morfologi i lexikonet

Språkbrukarens lexikon får antagas innehålla en uppsättning *språkliga tecken*, enligt den modell vi försökt uppställa i 4.3.

Även om en ordbok inte kan vara en genomförd konkretisering av det abstrakta lexikonbegreppet, bör detta betraktelsesätt hållas i minnet när man betraktar sådana exempel som

- (i) *tjänsteman vid järnvägen*,
- (ii) *järnvägstjänsteman*.

Hur många språkliga tecken förekommer i exemplen? Åtskilliga ordböcker förtecknar samtliga olika ord, vilket gör (i) ordrikare än (ii). Det verkar dock absurt att föreställa sig att språkbrukaren skulle betrakta *järnväg*, *tjänsteman* och *järnvägstjänsteman* som tre *olika* språkliga tecken. Tvärtom synes det rimligt att antaga att språkbrukarens lexikon dels innehåller en lämplig grunduppsättning av tecken, dels regler för hur tecken kan sammansättas och vilka egenskaper (på uttrycks- och innehållssidan) som sammansättningarna får. Eftersom ordboken under inga förhållanden kan förteckna ens de "vanligaste" sammansättningarna (kanske *järnvägstjänsteman* men knappast *järnvägstjänstemannaorganisation* etc) bör hänsynen till den teoretiska lexikonmodellen väga tungt vid urvalet av sammansättningar. Härvidlag får man sikta på den ideala språkmodellen, eftersom olika individer säkerligen har olika språkförmåga i praktiken (jfr avsnitt 4.5 och Ralph, "Ordböcker och språkmodeller"). Principiellt bör en sammansättning förtecknas endast om den, gentemot sina sammansättningsled, är ett *nytt tecken*, i den meningen att dess innehåll eller form ej kan förutsägas. Vid den praktiska tillämpningen får man dock styras av en viss generositet med tanke på att enskilda språkbrukare säkerligen har en relativt

bred grå zon mellan de två kategorierna (jfr avsnitt 4.6 om perifer vokabulär).

Motsvarande resonemang kan genomföras i fråga om avledningssuffix och -prefix. Fallet *tänka - tänkbar* berördes ovan; bland övriga vanliga suffix kan nämnas *-ning* (*önska - önskning*) *-are* (*meta - metare*) *-sk* (*glömma - glömsk*). Typiska och okontroversiella prefix är t.ex. *o-* (*tur - otur*) *anti-* (*demokrat - antidemokrat*). Affix sådana som dessa torde kunna anses som fullgoda språkliga tecken (ehuru bundna), varför huvudprincipen bör gälla även här: förutsägbara avledningar bör ej redovisas som egna enheter.

I denna lexikonmodell är således språkbrukarens lexikon uppbyggt kring morfemen i lika hög grad som orden, vilket givetvis förutsätter att språkbrukaren förfogar över en uppsättning morfologiska regler för att producera och tolka nya ord. Som en komplettering till vårt konkreta lexikon fordras alltså en morfologisk modell, vilken har följande huvuduppgifter: (1) beskriva *hur* enheterna kan kombineras, (2) beskriva *hur kombinationens uttryck* ser ut, (3) beskriva *hur kombinationens innehåll* skall tolkas. Liksom förut är det väsentligt att den morfologiska modellen görs explicit och att den på lämpligt sätt återges i ordboken.

En ordbok uppbyggd enligt de angivna principerna kan spara mycket utrymme genom att majoriteten av sammansättningar och åtskilliga avledningar behandlas ytterst schematiskt (eller i vissa fall möjligen ej förtecknas alls). Vi återkommer i avsnitt 4.8 till de konsekvenser modellen får för det slutliga urvalet av lexikonenheter. Här skall ytterligare påpekas att det faktum att ett sammansatt

eller avlett ord icke betraktas som ett eget, nytt tecken (och alltså icke företrädes i lexikonet) *ej* är liktydigt med att ordet "icke finnes" i svenska språket. Sammansättningar i svenskan har en högst varierande status, från mycket fasta (*strykjärn*) till helt tillfälliga bildningar (*avtalsrörelsestart*); samtliga är givetvis helt acceptabla svenska ord, ehuru endast en bråkdel av dem kan förtecknas i en ordbok. Motsvarande gäller, i tillämplig mån, för avledningar (*avtalsrörelsestartande* subst. -*t*). Det är viktigare att upplysa om *att* ett givet ord kan bilda sammansättningar och *hur* dessa bildas och uttolkas, än att förteckna en tämligen godtycklig uppsättning självklara sammansättningar.

Ovan fastlades att kriteriet för att låta ett morfologiskt komplext ord utgöra en egen lexikonenhets bör vara att dess betydelse *ej* är förutsägbar relativt beståndsdelarnas betydelser. Här uppkommer vid praktisk tillämpning tolkningsproblem beträffande vad som skall förstås med "förutsägbar betydelse". Den viktigaste funktionen hos den morfologiska modellen blir därför att beskriva de semantiska reglerna för ordbildning. Här bör de olika betydelse-typerna för sammansättning upptagas (lokalbestämning som i *ankdam*, funktionsbestämning som i *metspö* etc.). Det kan också nämnas att de flesta inhemska verb kan bilda verbalabstrakter (på *-ning*) och nomen agentis, ibland nomen instrumentalis (på *-are*) etc. I och med uppställandet av den morfologiska modellen finns alltså regler för de produktiva uttolkningsmöjligheterna för morfologiskt komplexa ord. Emellertid kan det inträffa att ett givet ords betydelse å ena sidan kan beskrivas som summan av sina delars betydelser, å andra sidan har erhållit en metaforisk eller specialiserad tilläggsbetydelse som saknas hos dess delar. Så är t.ex. *järnväg*

förvisso ett slags väg av järn, men i dess normala betydelse 'rälsbundet transportsystem' tillkommer faktorer som icke finnes i järn och väg men som vid sammansättningen tar överhanden. Ett mindre outrerat fall är *sommarstuga* med tilläggsbetydelsen 'fritidsbostad'. För dylika avgränsningsproblem fordras tillgång till den semantiska modell för produktiv metaforik, specialiseringar etc. som diskuteras i 4.6. I teoretiskt avseende har vi här att göra med "delvis egna" tecken, vilka därför bör förse med uppgift om tilläggsbetydelsen i lexikonet, medan huvuddefinitionen kan härledas ur delarnas definitioner. Denna princip kan ganska enkelt överföras till ordbokens representationssystem.

4.5. Definition och teckeninnehåll

Om lexikonenheten principiellt skall antagas bestå av ett tecken = föreningen av ett innehåll med ett uttryck, så står det klart att beskrivningen av tecknets innehåll är såväl den viktigaste som den mest krävande delen av arbetet. Många olika typer av innehållsbeskrivningar kan uppställas: för vårt syfte torde de grundläggande typerna vara *definition* och *parafras*.

Med definition menar vi nedan: en noggrann språklig beskrivning, i mindre komplexa termer, av ett teckens allmängiltiga innehåll. Med parafras menar vi: en kortfattad ungefärlig motsvarighet till tecknet. Definitionen måste i allt väsentligt täcka innehållet hos tecknet och kan därför bli begreppsmässigt komplicerad i vissa fall. Ehuru en god definition skall finnas för varje enhet i lexikonet, förutses att ordboken av utrymmesskäl kan behöva ersätta definitioner med parafraser för en viss del av lexikonet. Det bör understrykas att dessa ord egentligen inte diskrimineras härigenom, eftersom det ändå rör ord som inte tillför språket så mycket nytt semantiskt gods. Observera att enligt vår terminologi en synonym eller en översättning (i en tvåspråkig ordbok) normalt är att betrakta som parafras.

Eftersom tecknets innehåll teoretiskt är ekvivalent med definitionen, diskuteras lämpligen de semantiska grundproblemen med innehållsavgränsning etc. i detta sammanhang. Till en början kan vi konstatera att definitionen avser tecknets *allmängiltiga* innehåll. Den självklara (sociolingvistiska) aspekten är att innehållet skall

vara gemensamt för språksamfundet. Allmängiltighet kan dock krävas också i ett annat *ävseende*, nämligen i fråga om de *betydelse-skiftningar* som ordet kan undergå när det används i olika typer av syntaktiska konstruktioner.

I traditionell lexikografi brukar betydelseskiftningarna beaktas genom ett ofta osystematiskt bruk av extra betydelseenheter, *bi-betydelser* och allmänna anvisningar (t.ex. "överfört" eller "metaforiskt"), kompletterat med en likaledes osystematisk exemplifiering. Denna tradition har två allvarliga teoretiska nackdelar. Dels kan det ymniga bruket av nya betydelseenheter dölja den teoretiskt centrala skillnaden mellan äkta huvudbetydelser (som bestämmer egna tecken) och *bi-betydelser* av ett och samma tecken; dels är det principiellt omöjligt att förteckna samtliga möjliga metaforiska användningar, utvidgningar eller specialiseringar från en given betydelse, eftersom detta icke är ett statistiskt lexikaliskt faktum utan ligger inherent i samspelet mellan den lexikaliska huvudbetydelsen och språkbrukarens kreativitet. För vårt lexikon (och därmed ordboken) avser vi att begagna en annan modell för att komma ifrån dessa nackdelar.

Modellen bygger på principer som redan i viss utsträckning prövats (på morfemnivå) i arbetet med NFO 4. Stommen utgörs av ett regelsystem för semantisk utvidgning, specialisering etc. från en given *kärnbetydelse*. Regelsystemet kan sägas definiera en potentiell *semantisk räjong* till kärnbetydelsen, inom vilken de acceptabla utvidgningarna etc. måste ligga. Mellan kärnbetydelse och utvidgning måste råda semantisk kontinuitet. Observera att såväl tillägg som bortfall av en väsentlig betydelsekomponent bryter kontinuiteten även om resten är stabil. Omfånget av och konkretionsgraden hos den semantiska räjongen beror såväl av regelsystemet som av den givna kärnbetydelsens karaktär. Olika ord(familjer) får på så sätt olika status i fråga om precision och abstraktion, men detta förhållande är egentligen bara ett utslag av språkets allmänna flexibilitet. I idealfallet består lexikonet på innehållssidan uteslutande av väl skilda kärnbetydelser med väldefinierade räjonger som ej överlappar.

Vid den praktiska etableringen av databasens enheter kommer vissa avsteg från idealet att behöva göras, särskilt vad beträffar betydelser i periferien av en semantisk räjong. Här kan kompletterande semantisk anvisning fordras. Metodologiskt försiggår etableringen och den slutgiltiga definieringen av lexikonenheter (lexem) genom en ständig växelverkan mellan å ena sidan en preliminär definition och det semantiska regelsystemet, å andra sidan det semantiska fält som preliminärt betraktas som den semantiska räjongen till den sökta kärnbetydelsen. Härmed får definitionen status av metodologiskt hjälpmedel och ej blott slutlig output. Med andra ord: en preliminär definition används som prövosten när nya potentiella lexem testas.

För att bli användbar på sätt som avses måste definitionen anpassas till det semantiska regelsystemet så att målområdet kan täckas i största möjliga utsträckning. Ett trivialt exempel: *lejon* kan definieras som (a) 'ett slags kattdjur' eller (b) 'ett stort, starkt kattdjur'. I åtskilliga språkliga användningar (*kämpa som ett lejon, lejonparten, en men ett lejon* etc.) föreligger betydel-senyanser vilka spontant uppfattas som "överförda" från kärnbetydelsen. Med definition (a) kan inga rimliga semantiska utvidgningsregler förklara de överförda betydelserna ("den part som tillhör ett slags kattdjur" och "en men ett slags kattdjur" är helt nonsens och "kämpa som ett slags kattdjur" missar något väsentligt); med (b) blir de överförda betydelserna lättare genomskådliga. Uttryckt på annat sätt är kravet på definitionerna alltså bl.a. att de skall kunna förklara produktiv metaforik. Som en följd kan viss asymmetri uppstå mellan i övrigt jämförbara enheter inom en kategori, men detta kostar inte så mycket som det ger.

För mer komplicerade exempel än ovan (verb, abstrakta substantiv etc.) fordras därjämte att hänsyn tas till enhetens totala syntaktiska uppträdande i satser eller motsvarande. Dessutom bör närbesläktade enheter i förekommande fall lösas i ett sammanhang. (Detta är också viktigt för centralvokabulären; se nedan avsnitt 4.6.) Med denna utvidgade metodik kan följande grupp definitioner erhållas:

- (i) köpa; a köper b för c av/från d; 'a lämnar värde (= c) till d i syfte att erhålla *B*'
- (ii) sälja; a säljer b för c till d; 'a lämnar b till d i syfte att erhålla värde (= c)'
- (iii) byta; a byter b mot c med d; 'a lämnar b till d i syfte att erhålla c'.

De väsentliga delarna av definitionerna är kursiverade. Observera att skillnaden mellan *köpa/sälja* och *byta* är förekomsten av ett värde i de förras definitioner. Lägg också märke till den relativa komplexiteten hos definitionerna, vilken är en följd av det syntaktiska helhetsperspektivet. Ur definitioner av denna typ bör metaforiska uttryck som *sälja någon billigt* o.dyl. lätt kunna förklaras. (Det exemplifierade definitionsformatet kommer till stor del att reduceras i ordboksartiklarna; se vidare avsnitt 6.8.)

De anförda exemplen aktualiserar frågor kring definitionens syntaktiska form. Som allmän princip bör gälla att definitionen (åtminstone med någon ansträngning) skall vara syntaktiskt utbyttbar mot den definierade enheten. För vissa ordklasser, eller åtminstone stora underklasser (så troligen för verb), innebär detta att definitionen erhåller en helt enhetlig syntaktisk form. Däremot torde icke så olika substantivtyper som artnamn och verbalabstrakter kunna definieras på samma sätt. Adverbena utgör här, som ofta, en särskilt elakartad ordklass. Målsättningen bör därför vara att ställa upp ett begränsat antal undergrupper till varje ordklass, allt efter ordens syntaktiska och semantiska kombinationsmöjligheter. Inom varje undergrupp definieras sedan på enhetligt sätt. Det kan tilläggas att själva avgränsningen av de traditionella ordklasserna ibland kan diskuteras. Bestämmandet av undergrupperna är en empirisk fråga som lämpligen angrips med hjälp av ett stort antal preliminära "definitioner i kontext", som i exemplen i föregående stycke. Över huvud taget löses många problem bäst kategorivis.

I exemplen *köpa/sälja* angavs implicit en möjlighet att redogöra för specifika konstruktioners betydelser, i det att argumentet *c*, förekommande i den syntaktiska konstruktionen *köpa/sälja ... för c*, uttolkades som betecknande ett värde. Detta är en viktig kompletterande information som blir åtkomlig med den metodik som ovan angivits. En annan typ av definitionskomplettering gäller den be-

tydelsemodifikation som verb undergår vid tillfogande av vissa partiklar. Härvid är framför allt en noggrann definition av själva verbalhandlingen nödvändig för kompletteringens kvalitet (se Ralph, "Verb för verb" och "Från verbparafraser till konvertibilitets-system"). En tredje typ av komplettering har redan antytts ovan; den gäller betydelser i periferien av den semantiska räjongen till kärnbetydelsen. Denna tredje typ har icke samma teoretiska status som de övriga, utan är mera ett sätt att, för det enskilda fallet, dokumentera hur långt den semantiska räjongen kan utsträckas. Därför kan ej heller samma krav ställas på nära anslutning till huvuddefinitionen som i de två förra fallen.

För de parafraser som, särskilt i ordboken, kan komma att användas i stället för (eventuellt jämte) de fullständiga definitionerna kan inga definitiva riktlinjer ännu uppställas. Ett väsentligt krav bör dock vara att de är *distinktiva*, i den meningen att de i kombination med uttrycksdelen kan identifiera tecknets innehålls-sida gentemot andra tecken. En metodik för utformning av distinktiva parafraser finns redan utarbetad på morfemnivå. En idealisk parafras är givetvis en s.k. synonym, i den mån en sådan står att uppbringa. Utformandet av och samspelet mellan definition, parafras och synonym är ännu icke fastlagt.

4.6. Definitionssystemets struktur

Enligt föregående avsnitt skall en definition vara en noggrann språklig innehållsbeskrivning av ett tecken, *i mindre komplexa termer*. De kursiverade orden antyder att definitionerna i viss mening skall vara "riktade", nämligen från det komplexa till det enklare. Genom att definitionerna är riktade (vilket i och för sig oftast är fallet i traditionella ordböcker) får definitionssystemet nödvändigtvis en viss (partiell) struktur. Hittills har man inom lexikologien oftast stannat på en punkt där ordens begreppsnehåll är partiellt och implicit strukturerat. Inom Lexikalisk databas önskar vi däremot utveckla definitionssystemet dithän att det erhåller en total struktur som kan redovisas explicit. Härvid ansluter vi oss till en allmän riktning som har gamla rötter, t.ex. i Lockes, Leibniz' och Descartes' filosofiska system, och som fått teoretiskt lingvistiskt uttryck hos bl.a. János Petöfis forskargrupp som nämndes ovan i avsnitt 2.2. Tillvägagångssättet har även

praktiskt tillämpats i ordböcker sådana som Michael West, *An international reader's dictionary* och den snart publicerade "Learner's dictionary" som Longmans utvecklat under flera års tid.

För databasen planerar vi följande uppläggning. I centrum finns en uppsättning grundenheter, hädanefter kallad definitionsvokabulär. Definitionsvokabulärens element är axiom i logikens mening, dvs. de är förutsatta och principiellt ej definierbara. Vi avser dock icke att reducera definitionsvokabulären till dess logiska minimum; detta bl.a. för att dess enheter, icke minst för ordbokens behov, också skall kunna (kvasi)definieras. Definitionsvokabulären består av *ord*, och dessa skall behandlas som andra ord. Det bör dock märkas att "definitionerna" inom definitionsvokabulären är cirkulära och icke kan uppfylla kravet på reducerbarhet till mindre komplexitet. I en teoretisk definitionsvokabulär kunde enheterna lämpligen definieras i logiskt symbolspråk el. dyl.

För varje lexikonenhet utanför definitionsvokabulären gäller nu att den direkt eller indirekt skall kunna definieras av definitionsvokabulären. Med direkt definiering menas naturligtvis att enhetens hela definition ingår i definitionsvokabulären. Med indirekt definiering menas att (vissa av) orden i den första definitionen i sin tur kan definieras (etc.), så att definitionsvokabulärens element, och endast dessa, uppnås efter ett ändligt antal steg. Annorlunda uttryckt skall varje lexikonenhet kunna *successivt dekomponeras* till axiomen. Genom detta grundläggande krav erhåller databasen en fast semantisk struktur.

Det kunde möjligen synas enklare att låta samtliga lexikonenheter definieras direkt av definitionsvokabulären. Ett sådant förfaringsätt har dock både teoretiska och praktiska nackdelar (låt vara att det praktiskt använts i t.ex. Wests ovan nämnda ordbok). En definition enbart i termer av axiomen tenderar att utplåna delar av begreppsstrukturen hos den definierade enheten, såvida icke parentessystem eller liknande tillgrips. Visserligen kan en noggrann avpassning av definitionsvokabulärens storlek delvis råda bot på denna brist. Dock förefaller den successiva dekompositionen mycket rimligare ur språkpsykologisk synvinkel än direkt definiering, inte minst därför att definitionsvokabulärens interna struktur därigenom är inbyggd i systemet. I praktiska tillämpningar har direkt defi-

niering visat sig vara en, icke helt elegant och tämligen utrymmeskrävande form. Principen om successiv dekomposition saknar de angivna nackdelarna.

Som redan antytts kan det förväntas att ordboken ej har utrymme för fullständiga definitioner för alla enheter, speciellt om definitionerna skall uppfylla fordringarna i 4.5. En del av materialet kan behöva parafraseras i stället för definieras. En utgångspunkt för gränsdragningen gentemot de mer summariskt behandlade enheterna är det morfologiska betraktelsesättet (se 4.4). En annan utgångspunkt kan erhållas genom analys av definitionssystemets struktur. Följande principresonemang kan tillämpas. För varje lexikon-enhet kan anges ett ungefärligt "avstånd" till definitionsvokabulären. Antag t.ex. att detta "avstånd" sätts lika med antalet "steg" i den successiva dekompositionen som måste tas för att alla ord i den givna lexikonenheten skall ha reducerats till definitionsvokabulären. Man kan nu välja en gräns för avståndet och låta enheter innanför denna gräns bli fullt definierade, medan enheter utanför endast blir parafraserade. Det är icke orimligt att tänka sig att den enskilde språkbrukarens lexikon skulle vara strukturerat på liknande sätt, med varierande semantisk precision för olika typer av lexikonenheter.

Slutligen kan gränsdragningen också behöva ta hänsyn till förhållandet mellan språkliga fakta och fakta om världen (se vidare 4.7). Om man beslutar sig för att inom något fält av vetande avstå från detaljredovisning så följer det att detaljerna endast kan parafraseras. Typfallet är artnamn inom biologin, t.ex. *musseron* 'slags svamp', där *svamp* blir den yttersta fullt definierade enheten. Systemet äger tillämpning även inom andra delar av ordförrådet där över- och underordning kan urskiljas.

För att förenkla det fortsatta resonemanget inför vi förkortningarna FDV för den *fullständigt definierade* vokabulären och PV för den *parafraserade* vokabulären. Det understryks att beteckningarna ej avses vara normativt eller pedagogiskt värderande. Följande relationer gäller alltså:

DV innehåller enheter som endast kan cirkeldefinieras och vars betydelse förutsätts känd.

FDV innehåller fullt definierade ord. Varje FDV-enhet kan successivt dekomponeras, ofta via ytterligare FDV-enheter, till DV-enheter.

PV, som främst är avsedd direkt för ordboken, innehåller parafraserade enheter. Parafraserna bör på enahanda sätt ytterst vara nedbrytbara till DV-enheter.

Vid etableringen av de olika vokabulärtyperna måste teoretiska överväganden och praktiska språkliga experiment ständigt komplettera varandra. Det praktiska arbetet kommer snarast att fokuseras på FDV. Härvid bör följande punkter beaktas (jfr ovan): (1) till en FDV bör det vara möjligt att upprätta en tentativ DV så att ingen FDV-enhet ligger väsentligt "längre bort" än någon annan; (2) gränsen FDV-PV beror delvis av hur specifik man vill vara i fråga om fakta om världen; (3) hänsyn måste tas till de morfologiska aspekterna, särskilt vid gränsdragningen FDV-PV. Den färdiga FDV skall alltså vara ett välstrukturerat system, där de nämnda synpunkterna invägs. Ur en väl definierad FDV kan olika preliminära DV extraheras genom olika sätt att integrera definitionerna. De sålunda empiriskt erhållna DV kan testas mot teoretiska modeller eller mot på annat sätt utvunna vokabulärer (se framför allt Ralph, "Projektet Svensk definitionsvokabulär").

PV kan icke vara lika homogen som DV och FDV. Här ryms dels ord som är fullgoda tecken, dels sammansättningar/avledningar med specialbetydelser, dels kanske vissa helt förutsägbara sammansättningar etc. Den slutliga omfattningen av PV är icke så mycket en teoretisk fråga för databasen som en praktisk konsekvens av det omfång som fastställs för ordboken.

Det kan tilläggas att en DV såsom här skisserats inte nödvändigtvis är liktydig med språkets kortaste, mest frekventa och lättast förståeliga ord. Sådana ord är tvärtom ofta mångtydiga eller semantiskt vaga och lämpar sig därför mindre väl för detta ändamål. Så preciserar t.ex. verbet *åstadkomma* det väsentliga innehållet i det mångtydiga *göra* och är följaktligen en starkare kandidat för DV.

4.7. Lexikon och encyklopedi

I modern språkvetenskap skiljer man mellan den kunskap som fordras för att språkbrukaren skall kunna systematiskt begagna språkets teckenuppsättning och den kunskap om världen som finns kodifierad i teckensystemet. Den förra typen brukar betecknas som *lexikonet* (i teknisk mening; se avsnitt 4.2), den senare som *encyklopedien*. Exakt samma distinktion brukar göras inom praktisk lexikografi mellan ordböcker och uppslagsböcker. Trots de icke obetydliga gränsdragningsproblemen måste skillnaden mellan lexikalisk och encyklopedisk information upprätthållas, såväl inom databasen som i ordboken; detta av både teoretiska och praktiska skäl. Det kan icke vara en språkvetenskaplig uppgift att ge en fullständig beskrivning av världen, lika litet som en ordbok kan usurpera uppslagsbokens funktion. (För en delvis avvikande mening se flera rapporter från projektet Kernlexikon i Bielefeld, lett av János Petöfi; jfr ovan avsnitt 2.2 och 4.6.) Det rimliga encyklopediska kravet på en ordbok är att den inom den egna ramen skall rymma så mycket information om världen att den är självförklarande.

Vid ytterlighetsfallen föreligger inga svårigheter att skilja mellan lexikon och encyklopedi. Så är t.ex. ord sådana som *alla*, *bliva*, *och*, *som* etc. typiska bärare av enbart lexikalisk information, medan egennamn är typiska bärare av encyklopedisk information. Man kan här lägga märke till den grundläggande skillnaden: lexikonet svarar för den systematiska och i språkpsykologisk mening permanenta informationen, vilken också har egenskapen att vara tämligen väl beskrivbar med rent språkliga medel. Encyklopedien däremot informerar om de språkliga etiketter som tilldelats existerande företeelser. En språklig encyklopedisk beskrivning kan fordra komplettering (t.ex. genom en karta för geografiska namn, fotografi för personnamn).

Emellertid står det klart att själva språksystemet, även i snävaste mening, innehåller en implicit beskrivning av vissa grundläggande egenskaper hos världen. Ett intressant exempel är det subsystem som de lokativa prepositionerna utgör (se vidare Ralph, "Prepositioners presuppositioner"). Prepositionssystemet beskriver genom sin karaktär indirekt den geometriska strukturen i världen med god approximation. Substantiv såsom *hörn*, *kant*, *sida* kompletterar

geometrien och verb såsom *hålla*, *lyfta*, *släppa* introducerar vissa naturlagar, i och med att de presupponerar gravitationen. På liknande sätt presupponerar ord sådana som *kontrakt* och *rabatt* en viss ekonomisk struktur och *befogenhet*, *myndighet*, *ålägga* en viss politisk struktur. Andra världsorienterade kategorier som mycket ofta är språkligt representerade är deiktiska och talaktsberoende uttryck. Lexikonet måste alltså under alla omständigheter innehålla viss kunskap om världen. Vår uppgift blir därför att ange riktlinjer för upprättandet av en *encyklopedisk gräns* som bestämmer hur specifik kunskap lexikonet får innehålla. Nedan ges några preliminära kriterier för urskiljning av ej accepterade enheter.

Enheter bör utspärras ur lexikonet bl.a. på följande grunder:

- Om enheten utgör en (relativt språkssystemet) speciell beteckning för en företeelse i världen av typen egennamn.
- Om enheten endast kan definieras med *heta* eller liknande.
- Om enheten endast kan definieras med hjälp av annan utspärrad enhet. (Avser invånarnamn o.dyl. Däremot tolereras specifik information i komplettering till definitionen: *beduin* 'medlem av ett nomadiskt ökenfolk' spec. i Nordafrika och Västasien.)
- Om definitionen måste göra bruk av exakta koordinater i tid och/eller rum.
- Om den företeelse i världen som enheten betecknar är unik.
- Om enheten ej kan definieras tillfredsställande utan hjälp av ickespråkliga medel. (Här avses sådant som diagram, kartor etc.)

De uppräknade kriterierna kan upphävas på systematiska språkliga grunder. (Detta betyder att om man har tagit med *måndag*, *tisdag*, *onsdag*, *torsdag* måste *fredag*, *lördag*, *söndag* också inkluderas.)

En viktig typ härvidlag är produktiv metaforik utgående från enheten. Som goda lexikonenheter betraktas numera *bojkott*, *mackadam*, *raglan* m.fl., ehuru de uppkommit ur egennamn. Från ganska sen tid (1940) har vi *quisling*. "Stora" namn som *Einstein*, *Hitler*, *Napoleon* har icke lexikaliserats fullständigt, men utgör dock en bas för långtgående metaforisk användning. För sådana gränsfall kan specialutredningar fordras. En annan typ av gränsdragningsproblem uppstår vid beteckningar på samhällseliga institutioner o.dyl.

Tydliggen är *direktoriet* (observera formen) 'styrande församling under senare delen av franska revolutionen' på fel sida om den encyklopediska gränsen, medan däremot *parlament*, *underhus* får anses lexikaliserade. Moderna gränsfall skulle kunna vara *Hagarunda* och *privattjänstemannakartell(en?)*, där väl det sistnämnda ännu ej uppnått samma status som *landsorganisation* i språkligt hänseende.

Det anförda batteriet av kriterier har tagit sikte på beteckningar av namnkaraktär som avser enstaka företeelser i världen. Det finns emellertid ytterligare en typ, där beteckningen visserligen avser klasser av företeelser men behåller namnkaraktären och den bristande integreringen i språkssystemet. Det bästa exemplet är kanske de mera abstrusa biologiska artnamnen. Gränsfallen i denna grupp kan i ordboken redovisas schematiskt i den parafraserade vokabulären, om de i övrigt uppfyller urvalskriterierna (se härom 4.8). Den avsedda gruppen, som kan sägas ligga inom den encyklopediska gränss-zonen, kan utskiljas enligt följande:

- Om enheten endast kan definieras som underbegrepp och det ej finns systematiska språkliga skäl att utskilja enheten från det överordnade begreppet.
- Om enheten ej ingår naturligt i något språkligt subsystem.
- Om enheten måste definieras formellt (enligt något utomspråkligt system), ej funktionellt.

Det bör åter understrykas att de uppställda kriterierna är preliminära. Vid tillämpningen är det viktigt att det systemtänkande som redovisats framför allt i 4.6 får styra besluten. Det gäller att hela tiden hålla i minnet att syftet är att beskriva de lexikaliska aspekterna av *språkets* dynamik, ej att klassificera verkligheten enligt några filosofiska principer.

4.8 Urvalsproblematiken

Det konkreta urvalet av lexikonenheter angår på lång sikt ordboken i högre grad än databasen, där ej samma omfångsrestriktioner föreligger. I uppbyggnadsskedet sammanfaller urvalskriterierna nästan helt. I det längre tidsperspektivet kommer databasen att växa på egna villkor, medan ordurvalet för speciella ordböcker förblir snävt målinriktat.

Ordurvalet för en ordbok kan betraktas från så många synpunkter att det kan vara svårt att identifiera de underliggande grundfrågorna. Även om vi från början utesluter alla normativa aspekter (de har ingen plats i ett projekt som vill *beskriva* språksystemets teckenförråd) kan urvalsproblematiken behandlas ur minst tre aspekter: den språksociologiska (i vid mening), den systematiskt semantiska och den morfologiska. Vi skall nedan kort belysa vilka effekter hänsynstagandet till dessa aspekter kan ha för ordurvalet.

Det språksociologiska grundproblemet är givetvis *vilket* språk som skall beskrivas i lexikonet. Till att börja med måste språket avgränsas i tiden och rummet. (Skall äldre svenska beaktas? Skall regionala varianter beaktas?) Sedan dessa yttre gränser har fixerats gäller det att bestämma fördelningen över de språkvarianter som kan upprättas på grundval av olika sociala funktioner. Här måste man åtminstone ta hänsyn till dimensionen högtidligt-vardagligt språk, dimensionen fackspråk-allmänspråk och dimensionen värdeladdat-neutralt språk. Härvidlag kan avgörandena grundas på bl.a. frekvens. Den praktiskt viktigaste frågan torde gälla hur mycket fackspråk, inom vilka fack, som skall medtagas.

I språkteoretiskt avseende är fackspråk närmast att betrakta som allmänspråk med vissa specialiserade användningar av somliga ord samt med viss helt unik terminologi (vanligen mest substantiv). Semantiskt utmärks terminologin därav att den är språkligt isolerad (ingen metaforik etc.), att definitionerna är mycket regelbundna (konstruerade begrepp, ej spontant framvuxna) och av att den enligt resonemanget i 4.7 i stor utsträckning tillhör den encyklopediska gränzonen. För inlärning av ett nytt fackspråk skulle alltså fordras dels god allmänspråklig kunskap, dels en liten uppsättning nyckeltermer. Konkret kan saken formuleras så, att genom-

snittsläsaren med hjälp av ordboken skall kunna läsa visst slag av facktext (konversationslexikon, läroböcker, specialtidsskrifter etc.), varvid ordboken skall ge den första terminologiska nyckel som fordras för konstruktiv inlärning. Vilka fack som måste täckas och vilken svårighetsgrad "måltexten" skall ha är en sociologisk fråga snarare än en språklig.

Den systematiskt semantiska aspekten på urvalsproblematiken anknyter till diskussionen i 4.6. Genom att lexikonenheterna avses bilda ett strukturerat begreppssystem kan potentiella nytillskott undersökas med avseende på i vilken grad de berikar systemet. Ett nytillskott kanske endast skulle bidra med en nyansering av något redan förefintligt begrepp, eller tvärtom öppna vägen till ett nytt subsystem av begrepp. Slutsatsen är alltså att ordurvalet måste genomföras lika mycket under hänsynstagande till ordbokens *innehållsliga täckningsgrad* som till själva ordens *formella täckningsgrad*.

Det bör påpekas att hänsyn till den semantiska strukturen spelar in i fråga om *två* urval: dels urvalet av enheter i allmänhet, dels - och framför allt - urvalet av FDV-enheter. Vi kan illustrera beslutsprocessen genom att åter kort betrakta fackspråksproblematiken. Antag att vi (på utomspråkliga grunder) bestämt att ett visst fackområde bör täckas i ordboken. Fackområdets terminologi underkastas nu en semantisk analys som syftar till att utvälja de enheter (de må vara specialanvändningar av vanliga ord eller nya termer) som är begreppsmässigt grundläggande för området i fråga. Dessa grundläggande enheter förs till FDV, där de erhåller fullständiga definitioner. Ur den resterande terminologin utväljes en grupp kompletterande enheter, avsedda att ge huvuddragen av fackområdet. Dessa enheter förs till PV, där de parafraseras, ofta med hjälp av de grundläggande enheterna. Det andra valet styrs dels av inre kriterier (huvuddragen av området), dels av yttre kriterier, sådana som fackområdets relativa väsentlighet, balansen gentemot det övriga materialet, morfologiska aspekter (se nedan) m.m.

När det gäller den morfologiska aspekten, slutligen, erinras om resonemanget i avsnitt 4.4. Där framhölls att det språkliga tecknet ofta motsvaras av morfemet snarare än ordet, varför det ofta kan

räcka att definiera tecknets innehåll på endast en plats i lexikonet (nämligen vid det kortaste eller mest typiska ord i vilket morfemet ingår), medan sammansättningar/avledningar redovisas summariskt (i PV) om alls. Denna princip inses lätt äga relevans även för ordurvalet, dels givetvis genom att styra fördelningen på FDV och PV, dels på ett annat, något subtilare sätt.

På grund av den genomförda morfologiska principen har ordboken en implicit morfologisk täckning av förutsägbara sammansättningar och avledningar. Varje nytillskott kan därför bedömas med avseende på i vad mån det ökar den morfologiska täckningsgraden. Det finns tre huvudmöjligheter: (1) att den nya enheten redan är täckt, genom att den består av känt morfologiskt stoff (eventuellt med viss specialiserad betydelse), (2) att den nya enheten innehåller nya morfem, som icke förekommer i språket i övrigt, (3) att den nya enheten innehåller nya morfem som dock förekommer i andra ord utanför ordbokens ram. Under i övrigt lika omständigheter bör alltså ökning av ordbokens morfologiska täckning vara ett kriterium i urvalsproceduren.

Vi kan ännu en gång anknyta till fackspråksproblematiken. Åtskilliga fackområden (t.ex. kemi) har en morfologiskt utomordentligt regelbunden terminologi. Om ett visst fackområde skall täckas och om de semantiska urvalsprinciperna ej talar för någon annan lösning, kan ett strategiskt urval av terminologi skapa en mycket omfattande morfologisk täckning av ifrågavarande fackspråk. Här föreligger i själva verket något så intressant som en teoretiskt väl-motiverad möjlighet att både täcka stora delar av språkssystemet och spara utrymme i ordboken.

5. MÅLSÄTTNING FÖR ORDBOKEN

Efter den övergripande teoretiska diskussionen i kapitel 4 övergår vi nu till att ställa *ordboken* mera direkt i centrum för uppmärksamheten. Mot bakgrund av vad som hittills sagts anmäler sig några frågor naturligt. De kan sammanföras under fyra rubriker: *översikt över forskningsläget, målgrupp och användarkrav, informationens sammansättning, uppläggning och hantering.*

5.1. Översikt över forskningsläget

Med forskningsläge menar vi gängse idéer om hur ordböcker skall utformas. Man kan säga att forskningsläget, om det förstås på detta sätt, speglas dels i de ordböcker som är allmänt gångbara, dels i resonerande framställningar om lexikografi.

Det är naturligt att fråga sig om det finns någonting i det befintliga ordboksbeståndet som kan ge oss utvecklingsbara impulser. Denna rapport är att betrakta som ett inlägg i en dialog mellan språkforskare och förlagsfolk, och det är därför meningslöst att här göra en detaljerad genomgång av de verk av ordbokskaraktär som finns. Bägge parter är väl förtrogna med området, låt vara att vi har något olika infallsvinklar. Det kan här räcka med några antydningar.

Allén ger i projektbeskrivningen (s. 8 f.) några exempel på relevant pågående arbete i Sverige och utomlands. I samma skrift (s. 4 f.) får de mest centrala svenska verk som kommer i fråga några sammanfattande omdömen. Gellerstam har studerat en anseelig mängd utländska arbeten av växlande omfång. De som givits mest presenteras översiktligt i rapporten "Välja sina ord". Huvudintresset har gällt materialurval, men även den allmänna organisationen av informationsmängden samt den underliggande metodiken har uppmärksammats.

Dessutom finns givetvis åtskilligt av värde att hämta i den språkvetenskapliga litteraturen i allmänhet. Vi torde tillsammans ha en sådan överblick över denna, att vi bör komma allt väsentligt på spåren vid den systematiska inläsning som fortlöpande äger rum. I stället för att ge någon utförlig värderande genomgång av allt som producerats skall vi här dra några slutsatser av våra studier hittills.

Det finns ingen ordbok på något språk som vi kan lägga som mönster för Svensk ordbok. (Detta gäller innehållet; för den fysiska utformningen kan säkert ganska goda förebilder väljas.) På vissa punkter kan man få inspiration, men någon *arbetsbesparing* kan förmodligen inte uppnås, eftersom de konkreta detaljerna oftast är språkberoende.

Longmans stora ordboksprojekt (med publicering av en ordbok under 1978 som mål) innehåller många intressanta grepp. Detta projekt måste ha givit en del erfarenheter som vi skulle vara betjänta av att känna till. Det vore säkert mycket givande om man kunde få kontakt med någon eller några av de forskare som varit inblandade i projektet, något som nog skulle kunna åstadkommas. Som forskningsinstitution skulle Språkdata knappast uppfattas som konkurrent, men inte heller förlagen emellan behövde något konkurrensförhållande föreligga. (Longmans planerar veterligen ingen svensk enspråkig ordbok!)

De egenskaper hos Longmans ordbok som gör den speciellt intressant för oss är *dels* den konsekventa användningen av en definitionsvokabulär, *dels* den systematiska lingvistiska analys som genomförts och som resulterat i genomtänkta kategoriseringar, av vilka somliga tycks ha haft i första hand interna funktioner, dvs. inte kommer till explicit uttryck i själva ordboken annat än i form av organisation av stoffet, *dels*, och på sätt och vis viktigast, dess datoranknutna uppläggning. De kortfattade uppgifter om kodsystém vid den lingvistiska kategoriseringen som påträffats tyder på att intressanta sökningar och omsorteringar för lingvistiska ändamål varit möjliga.

Longmans ordbok använder en definitionsvokabulär, och det finns andra utländska ordböcker som också gör det. Vi kan förmodligen inte dra någon nytta härav, eftersom definitionsvokabulären måste uppbyggas systematiskt — på det egna språkets grund, självklart. I övrigt gäller att vissa befintliga analyser kan utnyttjas punktvis. Sålunda väntar vi oss att väsentliga delar av det morfosemantiska arbetet bakom NFO 4 skall komma till användning.

5.2. Målgrupp och användarkrav

Att skriva en bok är en kommunikativ handling. Man skriver normalt med tanke på en läsekrets. Situationen är egentligen densamma vid ordboksframställning. En svårighet i det senare fallet är emellertid att man vill nå många olika syftemål med en ordbok. Användarna kan tänkas vilja få många *olika typer* av frågor besvarade. Problemet fördjupas ytterligare för ordbokskonstruktören genom att ordböcker har en extra pedagogisk dimension; det finns så att säga en kategori potentiella användare på förhand, men man kan också vilja stimulera nya, normalt passiva, grupper till användning. Ordboken måste utformas så att den är tillgänglig även för dessa ovana läsare. En angelägenhet för åtkilliga. Målgruppen bör därför vara den vidast möjliga, snarast alla medborgare med grundskole-

För all läsning gäller att den bästa effekten uppnås om läsaren är engagerad, dvs. läser aktivt. Det är lättare för skribenten att hitta rätt tonläge ju snävare syftet är. Men därför behöver man inte avfärda möjligheten att åstadkomma en ordbok som syftar till att täcka "allmänhetens behov".

I det här sammanhanget skall man minnas att "allmänhetens" sammansättning drastiskt ändrats på senare tid, i och med att stora grupper som har främmande språk som förstaspråk flyttat in i landet. Härmed tillkommer en kategori som tidigare varit i stort sett negligibel, nämligen folk som inte har någon självklar intuition om det svenska språksystemet. Till yttermera visso utgör invandrarerna själva en mycket sammansatt grupp; inte minst behärskar olika individer svenska i mycket olika utsträckning.

Ett alternativ till en medelstor "universalordbok" vore kanske att konstruera ett ordbokspaket innehållande flera band som representerade delsyften. Detta kan låta idealiskt, men vem avgör hur fördelningen skall tillgå? Hur vet en användare att han skall välja en viss del snarare än en annan. Sådana bedömningar är förmodligen omöjliga att verkställa i praktiken, fränsett att de kan bli politiskt diskutabla.

Troligen är det omöjligt att på förhand ha en säker uppfattning om vilka som kommer att använda en ganska omfattande enspråkig svensk

ordbok. I själva verket finns det knappast några uppgifter om vilka som använder redan existerande föregångare. Det enda som kan redovisas är mer eller mindre välgrundade gissningar, vid sidan av mer precisa upplagesiffror och försäljningsbesked.

Det är lätt att ge exempel på vissa institutioner som med största sannolikhet hyser potentiella avnämare: kontor, skolor och myndigheter av olika slag etc. I ett land som Sverige med å ena sidan en komplicerad samhällsstruktur som kräver mycken språklig – inte minst skriftlig – kommunikation och å andra sidan ett välutvecklat utbildningssystem som praktiskt taget alla berörs av kan en ordbok antas bli en angelägenhet för åtskilliga. Målgruppen bör därför vara den vidast möjliga, snarast *alla medborgare med grundskolekompetens*.

Det bör ha framgått av kapitel 4 att en del sofistikerad information övervägs för ordbokens del. Det är uppenbart att vissa tillgängliga uppgifter kommer att lämnas outnyttjade av rätt många användare om målgruppen blir den avsedda. Är det då riktigt att förse ordboken med en mängd information som aldrig efterfrågas av många, kanske majoriteten? Det synes försvarbart att inkludera även den mer sofistikerade informationen (eftersom den säkert används av åtskilliga) under förutsättning att den inte i orimlig grad komplicerar ordboken i övrigt. Dessutom vet vi inte hur vissa typer av information kommer att användas, därför att traditionella ordböcker hittills underlåtit att ge sådan information. I någon mån blir det fråga om att vänja användarna vid de nya möjligheterna. Utfallet går kanske inte att värdera förrän efter en viss tid.

Bland de krav användarna kan tänkas ställa på boken märks både praktiska-estetiska och innehållsmässiga-funktionella. Ordboken skall rimligen rymmas i *ett band*. Vidare skall den ha ett *behändigt format*; framför allt måste proportionen mellan bredd, höjd och tjocklek noga övervägas. Härvid spelar givetvis papperskvalitén in, men den är dessutom ett argument i sig. (Enligt mångas spontana vittnesbörd betyder just papperskvalitén oerhört mycket för hur boken uppfattas.) En annan avvägningsfråga är stilen. Graden bestäms delvis i relation till utrymmet, men *lättläsheten* beror inte

bara på bokstävernas storlek. Alla dessa frågor liksom layout o.dyl. blir i första hand förlagets angelägenhet. I inte så få fall samverkar dock de innehållsmässiga aspekterna med de mera trycktekniska (t.ex. typografi!), och det är viktigt att dialogen förs aktivt mellan förlagsmän och språkmän härvidlag.

När det gäller de mer innehållsinriktade frågorna, torde kraven gå isär rätt kraftigt beroende på användarkategori. Allmänt fordrar man emellertid att den information boken tillhandahåller skall vara ordnad på ett överskådligt sätt. Att alla ingående ord på något vis går att finna i *strikt alfabetisk ordning* är helt nödvändigt. Ett annat önskemål som ofta framförs innebär att informationen skall återfinnas där man först söker den, dvs. man skall inte bli hänvisad till annan plats i boken. Ovan har flera gånger gjorts antydningar om referenssystem, appendix m.m. som tycks svära mot dessa fordringar (jfr även i synnerhet kapitel 6). Vad man då, som så ofta, får eftersträva är en god avvägning. Vad söker man i de flesta fall då man slår upp ett ord? Om det alls finns ett svar på denna fråga (och vi har våra aningar), bör man försöka ge så mycket information på det "ordinarie" stället att normalbehovet (= (av)stavning? böjning? uttal? kort semantisk specifikation?) täcks i allt väsentligt. Sedan kan man ha olika ambitionsnivåer. Idealet är att även de mer raffinerade kraven tillgodoses, men den som ställer dem får göra sig mera omak (dvs. följa en hänvisning till annat uppslagsord eller till appendix). Slutligen måste det krävas att den metaspråkliga informationen presenteras på ett lätt-tillgängligt sätt; förkortningar och abstrakta symboler är alltför ofta svårforcerade.

5.3. Informationens sammansättning

En framträdande faktor vid profileringen av Svensk ordbok är det underliggande *materialet*. Existerande ordböcker (med redan bearbetat material i någon form) måste beaktas. Den typen av korpus tycks vanligen utgöra stommen vid ordboksarbete – man *utgår* från det som finns och kompletterar eller reducerar det. Därigenom tenderar en viss lexikografisk tradition att utbildas inte bara vad beträffar hanteringen av stoffet utan även själva avgränsningen av materialunderlaget. Vi vill lägga mer tonvikt vid *autentiskt* material, dvs. sådant språkligt material som verkligen manifesterats i autentiska

kommunikativa situationer. Hypotesen är att de flesta ordböcker å ena sidan innehåller mycken umbärlig barlast, å andra sidan saknar täckning för vissa aspekter av språket i funktion. Uttryckt på annat sätt är existerande ordböcker sällan representativa för det språk som brukas i realiteten.

Om de existerande ordböckernas ordförråd får vi en god uppfattning genom Riksspråkligt datalexikon, som framställts vid Språkdata. Detta är ett maskinlagrat ordindex som bygger på en samkörning av uppslagsorden i Focus uppslagsverk, Nusvensk frekvensordbok och Svenska Akademiens ordlista, tionde upplagen. Självfallet kan denna bas utvidgas med flera verk om de bara har en maskinläsbar version.

Det autentiska materialet ställer i första hand Logoteket till förfogande. Härvid urskiljs två grundtyper: löpande text av olika slag och excerpter. Excerptsamlingar med olika tyngdpunkt finns på flera håll. Viktigast i detta sammanhang är de hos Svenska språknämnden i Stockholm och Svenska Akademiens ordboksredaktion i Lund. Det är angeläget att samarbete etableras med dessa institutioner. (Ett principiellt muntligt samarbetsavtal med Svenska språknämnden har redan träffats.) För att excerpter av antytt slag skall kunna omfattas av den förutsatta metodiken krävs inkodning. Logotekets texter föreligger redan i maskinläsbart skick.

Insamling av löpande text sker genom *stratifiering*. Härmed avses täckning av centralt språkbruk av olika specialiseringsgrad. Visst fackspråk måste vara representerat (exempelvis skolböcker, populärvetenskaplig journalistik, encyklopediartiklar m.m.). Urvalet är en grannlaga uppgift. *Bearbetningen* i sin tur kan innefatta moment som lemmatisering och morfologisk analys. Givetvis kräver alla dessa uppgifter personalresurser och tekniska resurser. Det förtjänar emellertid att påpekas, att en ansenlig mängd autentiskt material (i den angivna meningen) nu finns bearbetad. Den har därmed tagit steget över till den andra kategori av material som nämndes ovan (dvs. existerande ordboksmaterial).

Olika kriterier för inkludering av lexikaliskt material i Svensk ordbok kan tänkas. För accepterande fordras i första hand att den information som kan ges är av primärt *lexikalisk* - ej encyklopedisk

- natur (jfr avsnitt 4.7). Vidare skall uppläggningsen vara deskriptiv, ej normativ. Av det senare följer att visst material som normalt ignoreras eller stigmatiseras i ordböcker mycket väl försvarar sin plats, under förutsättning att det är väl integrerat i språksystemet. Gränsen mot dialektalt stoff får kanske dras ganska snävt, eftersom det i det fallet gäller strängt taget *autonoma språksystem* med viss överlappning i förhållande till riksspråket. Skönsavgöranden blir stundom nödvändiga. Allmänt gångbara vardagliga uttryck eller rentav slangord bör självklart accepteras. Genom inriktningen på det integrerade språksystemet faller dock den mer efemära slangvokabulären utanför fokus. Behandlingen av vad som ibland betraktas som stilistiskt "lägre" språkbruk rymmer problem som liknar dem som uppkommer vid behandlingen av värdeladdade uttryck (jfr ord som *arbetsvillig, demokrati, frimickel*).

Hela språkliga kategorier påverkas av beslutet om förkastande av encyklopedisk information. Namn och namnliknande ord, t.ex. initialförkortningar, får en klart underordnad status - om de inte utsluts helt. Problematiskt är att namnrelaterade ord visar språk-systematiska sidor (jfr t.ex. *kongo-les, nigeri-an, tanzani-er* dels sinsemellan, dels med *hurrit-er, zigen-are*, de senare utan påvisbar associering till något nomen proprium). Viss flexibilitet medges på dessa och en del andra punkter om appendix utnyttjas. Det kan också tilläggas här som i andra sammanhang, att man inte bundit sig vid sina beslut i alla detaljer tack vare den behändiga lagringen av ordboken. Illustrerad svensk ordbok har kommit ut i tre upplagor och åtskilliga *oförändrade nytryck* därtill. Svensk ordbok skulle utan teknisk svårighet kunna revideras fortlöpande.

5.4 Uppläggningsen och hantering

Det senast sagda leder oss över till den tekniska proceduren. "Ordboksmanuskriptet" är ett magnetband som framställs på grundval av en principiellt oberoende databas. Produktionen av den första versionen av Svensk ordbok sammanfaller delvis med etablerandet av själva databasen. När väl databasen existerar, kan emellertid kontinuerlig uppdatering ske. Ordboken kan med enkla medel aktualiseras, revideras eller förbättras på diverse sätt. Härigenom faller ett alldeles särskilt ljus över några av de frågor som tagits upp ovan. Man kan a priori tycka att modeord o.dyl. skall hållas utanför

en ordbok. De är ofta kortlivade och påverkar sällan systemet nämnvärt. Å andra sidan kan de vara högst frekventa och i viss mening livaktiga när de förekommer. Sådana ord skulle kunna inkluderas tentativt för att omprövas vid utarbetandet av en eventuell ny upplaga. Ett sådant förfarande kunde sägas skänka ordboken samma prägel av dynamik som är typisk för språket självt. Häre ligger onekligen en möjlighet till profilering.

Databastankens viktigaste implikationer har redan antytts: flexibilitet, mångdimensionalitet, ständig aktualitet. De tekniska egenskaperna och möjligheterna belyses mer i kapitel 7. Här skall bara återigen understrykas att en databas såsom den här åsyftade låter sig formas snart sagt hur som helst. Det är inte fråga om ett en gång för alla slutet helt, utan utvecklingsmöjligheterna är legio. Rent *principiellt* behöver inte ens det enspråkiga perspektivet behållas i all framtid. De restriktioner databasen känner är främst av teknisk art (minneskapacitet, programkomplexitet, utrustningsresurser). På avdelningarna och den inbördes ordningen mellan dem är naturligtvis sjö definitiva. Varje avdelning bör fylla sin speciella funktion för läsaren. Fördelningen av information på de olika avdelningarna är kanske främst en pragmatisk fråga, där erfarenheter och skattningar av läsarförväntningar är de mest vägande argumenten.

Vi föreslår att de olika avdelningarnas funktioner specificeras enligt följande. Inledningen skall vara en bruksanvisning till ordboken. Huvudlistan skall, som namnet anger, innehålla huvuddelen av den lexikaliska informationen; därtill skall den i görligaste mån vara användbar isolerad (utan inledning eller appendix). Appendix skall innehålla tilläggsinformation som kompletterar huvudlistan för den särskilt intresserade läsaren.

Eftersom inledningen skall vara en bruksanvisning till de övriga delarna kan dess innehåll knappast specificeras fullständigt ännu. Troligen fördras dock dels ett elementärt avsnitt för ovana läsare, närmast avsett som hjälp för att återfinna ord i huvudlistan, dels ett mera utförligt avsnitt som beskriver de olika möjligheterna att utvinna ytterligare information såväl ur huvudlistan som appendix. Det elementära avsnittet måste innehålla lättförståeliga redogörelser för alfabetiseringsprinciperna, behandling av stavnings-

6. UTFORMNING AV ORDBOKEN

Innevarande kapitel baseras dels på de teoretiska övervägandena i kapitel 4, dels på den allmänna målsättning som fastlades i kapitel 5. Eftersom kapitlet är inriktat på i högsta grad praktiska frågor försöker vi presentera konkreta förslag på alla de punkter där så är möjligt. Detta innebär dock ej att vi därmed definitivt bundit oss till de framlagda förslagen; fastmer kommer åtskilliga att behöva ändras i huvuddrag eller detaljer allteftersom arbetet fortskrider; och detta av både praktiska och teoretiska skäl. Ibland markeras den preliminära karaktären hos besluten genom att alternativa förslag framlägges.

6.1. Ordbokens organisation

En ordbok består åtminstone redigeringstekniskt av två huvudavdelningar: *inledning* och *huvudlista*, För Svensk ordbok finns det också anledning räkna med en tredje avdelning: *appendix* (rättare: *appendices*). Namnen på avdelningarna och den inbördes ordningen mellan dem är naturligtvis ej definitiva. Varje avdelning bör fylla sin speciella funktion för läsaren. Fördelningen av information på de olika avdelningarna är kanske främst en pragmatisk fråga, där erfarenheter och skattningar av läsarförväntningar är de mest vägande argumenten.

Vi föreslår att de olika avdelningarnas funktioner specificeras enligt följande. Inledningen skall vara en bruksanvisning till ordboken. Huvudlistan skall, som namnet anger, innehålla huvuddelen av den lexikaliska informationen; därtill skall den i görligaste mån vara användbar isolerad (utan inledning eller appendix). Appendix skall innehålla tilläggsinformation som kompletterar huvudlistan för den särskilt intresserade läsaren.

Eftersom inledningen skall vara en bruksanvisning till de övriga delarna kan dess innehåll knappast specificeras fullständigt ännu. Troligen fordras dock dels ett elementärt avsnitt för ovana läsare, närmast avsett som hjälp för att återfinna ord i huvudlistan, dels ett mera utförligt avsnitt som beskriver de olika möjligheterna att utvinna fylligare information såväl ur huvudlistan som appendix. Det elementära avsnittet måste innehålla lättförståeliga redogörelser för alfabetiseringsprinciperna, behandling av stavnings-

varianter, hänvisningssystemet, sökning vid ej återfunna sammansättningar/avledningarna samt uttolkning av den typografiska koden i ordboksartiklarna.

Läsaren kan tänkas använda ordbokens, framför allt huvudlistans, information på två sätt: dels för analys (sökning efter information om i text påträffade ord), dels för syntes (sökning efter information för produktion av egen text). Det är rimligt att räkna med större beredskap för egen aktivitet i det senare fallet. Målsättningen för huvudlistan bör vara att den ensam skall kunna tillgodose analysens behov och tillsammans med inledningen tillgodose syntesens behov. Detta innebär att all den information som fordras för ordigenkännande och ordförståelse måste finnas i huvudlistan; vidare att dispositionen och den typografiska utformningen av artiklarna måste vara så enkel att den kan genomskådas utan hjälp av inledningen (enligt samstämmiga uppgifter når inledningen i en ordbok icke någon majoritet av läsarna). Likaledes bör de summariskt behandlade sammansättningarna/avledningarna redovisas på ett sådant sätt att den läsare som ej återfinner en sökt sammansättning/avledning uppmuntras att söka efter ordets delar.

Till appendix förs två typer av uppgifter: dels sådan information om ordbokens underliggande grammatiska modell som ej är *nödvändig* för användandet av ordbokens centrala delar (dylik information redovisas i inledningen), dels kompletterande information som kan underlätta språklig produktion och avancerad språkförståelse eller helt enkelt kan vara intresseväckande av allmänna skäl. Appendix måste dock förbli *kort* i förhållande till huvudlistan.

6.2. Tilläggsinformation i ordboken

Här beskrivs några grundläggande typer av uppgifter som kan eller bör återfinnas i ordboken, ehuru de uppenbarligen inte hör hemma i huvudlistan. Uppgifterna i fråga kan redovisas i inledning eller appendix. Om de klart har sin plats i inledningen påpekas detta särskilt.

- Ortografisk variation. Alla stavningsvarianter av alla ord kan inte förutses. Det fordras alltså uppgifter om möjliga stavningsätt (t.ex. *c-k*, *e-ä*, *sj-sch-sh*). Därtill kan man behöva redovisa

relevanta delar av svenskans ortografi, såsom reglerna för dubbel-skrivning av konsonant och avstavningsregler. Ätminstone den förra typen av uppgifter måste stå i inledningen, eftersom de är väsentliga för möjligheten att återfinna ord i huvudlistan.

- Uttalsprinciper och uttalsvariation. Eftersom svenskans ortografi trots allt är tämligen ljudnära kan fullständig uttalsangivelse för alla ord i huvudlistan inbesparas. Uttalsangivelsen reserveras till de inte ljudenliga fallen. I stället fordras dock att de allmänna uttalsprinciperna redovisas. Systematisk uttalsvariation bör också anges i detta sammanhang. Som exempel kan nämnas s.k. uttalslättnader (*d*-bortfall i *röd*, *sned* etc.), uttal av romanska ord (*j:sj* i *genialisk* etc.) samt tryckförflyttning vid böjning (*'doktor : dok'torer*). Uppgifter av denna allmänna typ hör snarast till inledningen.

- Flexionsmorfologi. Ehuru den viktigaste ordböjningen måste anges noga i huvudlistan kan det vara värdefullt med en samlad presentation, både för att demonstrera det totala böjningssystemet och för att ge en indikation om böjningen av oredovisade ord.

- Derivationsmorfologi. Huvudlistan avses göra utrymmesbesparingar genom en tämligen summarisk behandling av självklara typer av sammansättningar och avledningar. Derivationsmorfologin redovisas lämpligen genom att informationen uppdelas på två avdelningar. I den ena avdelningen beskrivs främst hur man återfinner delar av ett sökt (sammansatt eller avlett) ord i huvudlistan, vartill viss upplysning om det sökta ordets sannolika sammansättnings- eller avledningsbetydelse ges. I den andra avdelningen beskrivs ordbildningsprinciperna tämligen utförligt, såväl formellt som innehållsligt. Den förra avdelningen måste ingå i inledningen.

- Syntaktisk modell och syntaktiska konstruktionssätt. Den syntaktiska modell som ordbokens grammatiska information bygger på bör specificeras. En viktig del av den grammatiska informationen är uppgifter om syntaktiska konstruktionssätt, som i huvudlistan redovisas med kodbeteckningar, vilkas nyckel måste ges i ordboken. (Syntaktiskt konstruktionssätt är en utvidgning av den typ av information som i äldre ordböcker återges med *vb tr. pred. adj.* och liknande.)

- Semantisk härledningsmodell. Ordbokens definitioner inriktas på s.k. kärnbetydelser (se avsnitt 4.5), varför samtliga belagda eller potentiella betydelsenyanser hos orden ej kommer att redovisas i detalj. I stället ges allmänna anvisningar om hur betydelser kan överföras, utvidgas eller specialiseras utgående från kärnbetydelser. Viktigt är att dessa anvisningar presenteras på ett lättillgängligt sätt.

- Egennamn m.m. Egennamn och egennamnsavledda ord faller utanför gränsen mot encyklopedisk information (se avsnitt 4.7), men på grund av att avledningarna ofta är svårförutsägbara kan det finnas skäl att ge vissa formella uppgifter om en del egennamnsavledning- ar. En möjlighet är att lägga denna information i appendix.

- Etymologisk information. Denna information, som får klassas som allmänt intresseväckande kan infogas i artiklarna eller i utrym- mesbesparande syfte brytas ut ur huvudlistan och redovisas separat.

6.3. Uppslagsordens status

I kapitel 4 beskrivs en semantiskt motiverad uppdelning av vokabu- lären i tre klasser (DV, FDV, PV; se närmare 4.6). Teoretiska ar- gument har också givits för en icke-homogen behandling av vokabu- lären på morfologiska grunder (4.4) och som följd av gränsdrag- ningen mot encyklopedien (4.7). Nedan anges hur de olika ordtyper- na tänkes bli representerade i ordboken.

I de flesta avseenden behandlas ord inom DV och FDV lika, dvs. de redovisas fylligt och med fullständig definition. Eftersom ord inom DV dels har en teoretiskt central ställning inom ordboken, dels i sista hand måste definieras med inbördes cirkularitet, kan det vara befogat att typografiskt markera DV-ord på något diskret sätt.

Inom PV kommer många ord att behandlas lika fylligt som inom FDV, så när som på att den fullständiga definitionen ersätts med para- fras. Andra ord behandlas mer summariskt, av formella och/eller innehållsliga skäl. För vissa ord anges endast de formella oregel- bundenheterna (avvikande uttal, speciell böjningsform, oförutsäg- bar morfologisk uppbyggnad); i vissa andra fall utelämnas formella

uppgifter helt, eftersom informationen lätt kan återvinnas inom ordbokens ram. På motsvarande sätt parafraseras vissa ord genom direkta hänvisningar till sina beståndsdelar, i förekommande fall kompletterade med uppgifter om specialanvändning eller dylikt. Några ord parafraseras genom indirekt hänvisning, i den meningen att den morfologiska uppbyggnaden antyds typografiskt (härom se Järborg & Ralph, "Ordboksartiklarna: ordbildnings- och ordstruktur-synpunkter").

Av de enheter som ej är egentliga ord (se 4.3) redovisas morfem med egna artiklar enligt modellen för FDV. Fall av typen verb med partikel behandlas under vederbörande verb. Bland flerordsenheter-na redovisas den bundna, syntaktiskt ostrukturerade typen (ex. *au pair*) som flerordsenhet, medan den öppnare, strukturerade typen (ex. *kasta yxan i sjön*) redovisas under något eller några av sina huvudord (se vidare avsnitt 6.5).

6.4. Ordmodeller för uppslagsord

Vid upprättandet av den primära lexikografiska enheten, uppslagsordet, beaktas det språkliga tecknets formella såväl som innehålls-liga egenskaper (se utredningen i avsnitt 4.3). Beroende på hur man ordnar kriterierna inbördes erhåller man olika modeller, med konsekvenser även för ordningen mellan uppslagsorden. Vi beskriver här ett par olika modeller och diskuterar kortfattat deras för- och nackdelar. Modellerna exemplifieras i avsnitt 6.8. För en diskussion se vidare Berg, Järborg & Ralph, "Synpunkter på stickord" och Allén, "Om uppslagsord i Svensk ordbok/Lexikalisk databas".

Den första modellen (se 6.8, lista 1) är ett försök till precisering av den s.k. *etymologiska* principen i traditionella ordböcker. Orden indelas i grupper efter grov semantisk och morfotaktisk samhörighet. Det primära uppslagsordet är det semantiskt-morfologiskt enklaste ordet i gruppen. Sekundära uppslagsord (har momentsiffror i exemplet) upprättas på formella grunder. I sista hand klassificeras betydelseskilnader.

Vi finner denna modell vara olämplig av följande skäl. Den bryter i onödan mot den alfabetiska ordningen; ett genomfört kompletterande alfabetiskt hänvisningssystem kostar mycket utrymme. Ordboks-

artiklarna blir svåröverskådliga med sina många delmoment, där också formellt och innehållsligt betingade moment växlar på ett föga läsarvänligt sätt. Slutligen är det första kriteriet (grov semantisk samhörighet) mycket svårt att praktiskt tillämpa. I exemplet har *grundare* förts till ²*grund*, men så kunde även ha skett med *grundlig* eller med *intetdera*.

Som alternativ till vad man kunde kalla den kvasiatymologiska modellen uppställs *lemma-lexemmodellen*, som kan utföras i olika varianter. Den utgår från att såväl språklig form som språkligt innehåll måste beaktas på ett systematiskt sätt. Härvid gäller å ena sidan att likhet i fråga om form såväl som likhet i fråga om innehåll verkar sammanhållande, å andra sidan att olikhet i respektive avseenden verkar särskiljande.

På grundval av formella kriterier etableras *lemman*, på grundval av innehållsliga kriterier *lexem*. Lexemen kan i sin tur innehålla betydelsenyanser.

Normalfallet är att lemma och lexem täcker varandra, så att inget problem uppstår beträffande den lexikaliska enheten (ex. *krinolin*).

I vissa fall drar form och innehåll åt olika håll. Detta kan illustreras med exempelvis det starkt böjda *sticka*. Här har vi å ena sidan den sammanhållande böjningen *sticka*, *stack*, *stuckit*, å andra sidan åtminstone tre olika betydelseenheter: 'stinga', 'ge sig av' och 'skifta (om färg)'. Inom lemma-lexemmodellens ram har man här alltså möjlighet att ansätta ett uppslagsord *sticka* med tre tillhörande lexem eller tre uppslagsord med samma formella egenskaper och innehållsligt svarande mot lexemen. Presentationsmässigt kan flera mellanformer tänkas.

För att klargöra att den språkliga formen utövar inflytande på lexikonets struktur också om antalet uppslagsord baseras på lexemen skall vi fästa uppmärksamheten på det svagt böjda *sticka*. Detta har ungefärligen betydelsen 'binda garn i maskor'. En rimlig ordning mellan de fyra verbala *sticka*-lexemen vore kanske 'stinga', 'binda garn i maskor', 'skifta färg', 'ge sig av'. På formsidan skulle detta ge en växling starkt, svagt, starkt, starkt. En sådan

växling är inte acceptabel genom sin brist på stringens. Den är antagligen också psykologiskt sett dålig. Gemensamma formella egenskaper verkar som nämnts sammanhållande.

Lemma-lexemmodellen samlar i sig upp de relevanta teoretiska aspekterna, möjliggör ett systematiskt utnyttjande av dem och ger förutsättningar för en ändamålsenligt konsekvent presentation. Den kan i ordboken utformas enligt någon av flera olika varianter. För att visa bredden i dess tillämpbarhet har två avsiktligt polariserade varianter oförbindligt exemplifierats i 6.8, listorna 2 och 3. Den ena ger ytligt sett en mera framträdande roll åt formen, den andra åt innehållet.

Den lemmaorienterade varianten i lista 2 är troligen den enklaste att tillämpa. Den har fördelen att ge snabb och koncis presentation av de formella uppgifterna, vilket bl.a. innebär att böjningsformer lätt kan återfinnas. En nackdel är att uppslagsorden i vissa fall kan innehålla lexem med mycket skiftande semantisk karaktär, medan semantiskt närbesläktade ord kan tillhöra olika lemman; se t.ex. verbet *grunda*.

Den mera lexemoriterade varianten i lista 3 har fördelen att endast kräva en indelningsnivå under uppslagsordet. Den ger också delvis ett närmare samband mellan form och innehåll. En nackdel är att antalet uppslagsord blir stort, varför man förlorar utrymme, och att i samband härmed informationen om de formella egenskaperna blir oekonomisk genom upprepad återkomst. Se verben ¹*grunda* - ⁴*grunda* i lista 3.

Den kvasietymologiska modellen får anses avfärdad. För närvarande diskuteras vilken variant av lemma-lexemmodellen som är mest fördelaktig. En fråga av vikt är hur många besvärliga fall som uppkommer inom olika varianter. Detta är en sak som det konkreta arbetet kommer att ge besked om. Avgörandet behöver inte träffas nu. Genom databasförfarandet kan tvärtom valet av slutlig modell uppskjutas till ett ganska sent stadium. Det viktiga är att databasen fylls med sådan information att endera modellen lätt kan framtagas.

6.5. Huvudlistans struktur

På huvudlistan ställs vissa grundläggande krav, som delvis är motstridiga. Å ena sidan bör ordbokens vokabulär organiseras så att varje enhet är lätt att återfinna samt att den viktigaste informationen (särskilt för analys; jfr 6.1) för en sökt enhet erhålles direkt ur listan, utan hänvisning till andra enheter. Å andra sidan vinner ordboken givetvis på varje tillkommande ord. Det både teoretiskt (se 4.4) och praktiskt lämpligaste sättet att spara utrymme är att behandla självklara sammansättningar och avledningar summariskt, vilket förutsätter ett hänvisningssystem av något slag.

För att enheten lätt skall återfinnas i huvudlistan fordras dels en god modell för uppslagsorden, dels en genomförd strikt alfabetisk ordning mellan dem. Den alfabetiska ordningen kompletteras med grammatiska och semantiska ordningskriterier för underavdelningarna.

I artikelinnehållet redovisas nedan i punktform. För presentationsfrågor hänvisas generellt till 6.7 samt till Allén, Morfem (nämligen prefix och suffix) införs på sina alfabetiska platser i listan, varvid dock deras status markeras genom efterställd (*anti-*) respektive framförställd (*-ning*) divis. För exempel jämför Järborg & Ralph, "Ordboksartiklarna: ordbildnings- och ordstruktursynpunkter" (fallet *-sam*). Den ostrukturerade flerordstypen (*au pair*) alfabetiseras med negligering av blankt typrum (man får t.ex. *aula - au pair - autograf*). Idiomtypen redovisas däremot under något eller några av sina huvudord, sist i artikeln (se Allén, "Om utformningen av artiklarna i Svensk ordbok" samt nedan avsnitt 6.8). Som huvudord betraktas de ord i idiomet som är semantiskt tunga, som knappast kan utbytas utan att idiomet går förlorat och/eller som tillhör ordklasserna adjektiv, verb och substantiv. I tveksamma fall överregistreras idiom.

I samma artikel som huvudordet, efter vederbörligt delmoment, förtecknas tillägg till ordet (s.k. "en-och-en-halv-enheter": verb + partikel, verb + reflexiv etc.). Dessa delvis nya enheter, som kan ses som en speciell typ av fraseologi, har ofta avvikande syntaktiskt konstruktionssätt, vilket redovisas samtidigt. Se exemplet *grunda upp* i avsnitt 6.8. Om flera "en-och-en-halv-enheter" föreligger, alfabetiseras i vanlig ordning. I vissa fall hänvisas i stället till det sammansatta verbet.

Exempel på hur hänvisningssystemet kan läggas upp finns i Järborg & Ralph, "Ordboksartiklarna: ordbildnings- och ordstruktursynpunkter" samt nedan i 6.8. Det bör understrykas att något beslut om hänvisningssystemets omfattning och om vad som skall förstås med "självklar" sammansättning etc. ej har fattats och ej kan fattas förrän en preliminär FDV har etablerats. Det pragmatiska målet kan vara att kompositionen av FDV tillsammans med hänvisningssystemet skall kunna bygga upp en totalvokabulär på upp till 100 000 ord. Det kan allmänt fastläggas att hänvisning bör ske till det morfologiskt mest närliggande ordet (analogt med principen om successiv dekomposition; se avsnitt 4.6). I lista 2 och 3 i 6.8 hänvisas alltså från *grundare* till *grunda* och endast indirekt till det primära *grund*.

- Definition, parafras. Här hänvisas till avsnitt 4.5. I 6.8, liksom i 6.6. Ordboksartiklarnas innehåll användning som identifikation av

Det maximala artikelinnehållet redovisas nedan i punktform. För presentationsfrågor hänvisas generellt till 6.7 samt till Allén, "Om utformningen av artiklarna i Svensk ordbok" och Järborg & Ralph "Ordboksartiklarna: ordbildnings- och ordstruktursynpunkter".

- Ortografi. Denna gives omedelbart av stickordet. Ortografisk variation kan i vissa fall redovisas på sin alfabetiska plats i huvudlistan. Se för övrigt avsnitt 6.1.

- Uttal. Detta markeras endast i svårförutsägbara fall (accent och kvantitet anses normalt förutsägbara). Den viktigaste uttalsuppgiften är huvudtryckets lokalisering, vilket ofta är tillräcklig hjälp

- Avstavning. En enkel avstavningsuppgift kan vara befogad för morfologiskt enkla ord. Därmed införs i viss mån ett normativt moment, i och med att ordboken faktiskt tycks rekommendera viss avstavning. Fördelarna med att ange en avstavningsmöjlighet överväger dock.

- Formell morfologi. För analytiskt bruk bör uppbyggnaden av morfologiskt komplexa ord redovisas på något enkelt sätt. För syntesen fordras att ordbildningsstammen anges, för de fall den avviker från den normala ordformen (se vidare Järborg & Ralph, "Ordboksartiklarna: ordbildnings- och ordstruktursynpunkter").

- Ordklass, konstruktionssätt. Ordklassuppgiften är grammatiskt central och nödvändig bl.a. som nyckel till böjningsangivelsen. Konstruktionssätt är en syntaktisk precisering av ordklassuppgiften (liksom böjningsangivelsen är en morfologisk precisering). Ordklasssystemet och konstruktionssättsmodellen är inbördes beroende i den meningen att en knapp uppsättning ordklasser kräver mer detaljerade konstruktionssättsangivelser.
- Böjning. Alla tänkbara böjningar kan icke redovisas; därför är det viktigt att centrala böjningsformer utväljs. Man kan eventuellt tänka sig hänvisningar vid s.k. suppletiv böjning (*god ... komparation se bättre*).
- Definition, parafra. Här hänvisas till avsnitt 4.5. I 6.8, lista 3, har parafrasen en utökad användning som identifikation av uppslagsordet.
- Synonymi, antonymi. Om en god, klarläggande synonym kan påträffas, bör den anges. Antonymer kan förväntas bli stundom förekommande kompletteringar av definitioner eller parafrafer.
- Specifika konstruktioner. Dessa beskrivs framför allt i avsnitt 4.3. Eftersom det rör sig om en svårförutsägbar semantisk specificering i huvudordets närmkontext, bör den specifika konstruktionens form och betydelse i möjligaste mån integreras med huvudordets definition. För den systematiska inordningen i huvudlistan se 6.5.
- Specialbetydelse, användningsområde. Denna typ av uppgift är ytterligare en form av komplettering av definitionen/parafrasen. I princip anges de applikationsområden som inte logiskt följer av definitionen. Den vanligaste uppgiften av denna typ gäller överförd betydelse o.dyl. (se 6.8). Ett särfall är implicita värderingar som ibland fordrar tämligen långa utredningar (t.ex. arbetsvillig ... från arbetsgivarhåll om den som arbetar trots strejk).
- Stilnivå, bruklighet. Information av detta slag ges endast i flagranta fall. Distinktionen formell : informell stil är kanske tillräcklig. Stor försiktighet måste under alla omständigheter tillrådas.

- Region, dialekt. Eftersom ordboken skall återge svenskt riksspråk blir denna typ av uppgift sällsynt. (Ett exempel finns i avsnitt 6.8.) Möjligen kan icke-naturaliserade utländska ord betecknas på något motsvarande sätt, som en varningssignal och en uppmaning att beakta uttals- eller böjningsangivelsen.
- Etymologi. Meningsfull etymologisk information kan knappast göras alltför kortfattad. Det kunde därför eventuellt vara lämpligt att i vissa fall avstå från etymologi och i andra fall göra etymologisk hänvisning (se *gryнна* i avsnitt 6.8) för att på detta sätt vinna utrymme för en fylligare etymologisk utredning vid centrala ord.
- Idiom. Dessa är egna enheter, vilka endast av tekniska skäl redovisas under ett vanligt uppslagsord. De erhåller aldrig någon full definition, endast parafras, men behandlas för övrigt analogt med normala ord.

6.7. Ordboksartiklarnas uppläggning

Detta avsnitt har nära anknytning dels till avsnitten 6.6 och 6.8, dels till Allén, "Om utformningen av artiklarna i Svensk ordbok". Den senare skriften utgör allmänt underlag för resonemanget. Delar av uppläggningsen är givetvis beroende av den slutligen valda modellen för uppslagsord; i dessa fall kan alltså ännu inga beslut fattas.

- Ortografi, avstavning, morfologisk uppbyggnad. Stickordet ger ortografin, men kan även användas för en ytterligare uppgift. Här föreslås att avstavningsmöjlighet anvisas för morfologiskt enkla ord, genom "kil" nedifrån. För morfologiskt komplexa ord anges i stället morfologiska huvudgränser (före avledningssuffix, efter foge-s etc.) genom upphöjd (diskret) punkt, vilken samtidigt tjänar som primär avstavningsgräns.

- Uttal. Anges enligt fonetisk praxis inom hakparentes. Uttalsbeteckning bör ske enligt något av de alternativ som ges i Beeck, Santesson & Witthoff 1969. I många fall blir det endast fråga om att ange ett eller ett par ljud.

- Ordklass. Genomskinliga förkortningar används.

- Böjning, ordbildning. Centrala böjningsformer och ordbildningsstam skrivs i följd efter ordklassuppgiften. Böjningsformer som kan läggas direkt till stickordet skrivs med divis vänster, i övriga fall skrivs det böjda ordet ut. Ordbildningsstam skrivs alltid med divis höger. Ett format för angivande av alternativ (snedstreck? el.?) måste utarbetas.

- Konstruktionssätt. Detta bör anges med typografiskt klart avvikande sifferbeteckning, för att markera den metaspråkliga karaktären hos informationen. Av systematiska skäl bör konstruktionssättet följa lexemen.

- Specifika konstruktioner, idiom. Man kan överväga om "en-och-enhalv-enheter" och idiom bör utskrivas i sin helhet eller med utnyttjande av upprepningsslinga. De bör följa närmaste relevanta lexem eller betydelsenyans, där så är möjligt, eljest följa i slutet av artikeln efter semikolon.

- Etymologi, etymologisk hänvisning. Sådana uppgifter bör stå sist i artikel respektive moment beroende på den valda modellen för uppslagsord. Härigenom markeras informationens illustrerande karaktär; dessutom kan om så önskas hakparentes användas utan risk för sammanblandning. Ord med hänvisning bör icke ha etymologiska uppgifter

- Specialbetydelse, användningsområde, stilnivå, bruklighet, region, dialekt. Dessa uppgifter (och andra här ej förutsedda) har naturligen karaktären av fria kommentarer och skrivs, i petit, på den i förekommande fall lämpligaste platsen. Vanligen gäller väl uppgiften lexemet. Den mest frekventa kommentaren torde bli någon form av utvidgning eller specialisering av definitionen/parafrasen. I samtliga fall är det viktigt med genomskinliga förkortningar i den mån sådana används.

6.8. Exempel på ordboksartiklar

Det skall än en gång understrykas att inga exempel som hittills producerats inom Lexikalisk databas är definitiva. Detta gäller såväl exemplen i tidigare skrifter som de här föreliggande. Exem-

plen är att uppfatta som illustrationer.

För notationssystemet hänvisas till Allén, "Om utformningen av artiklarna i Svensk ordbok" samt till avsnitt 6.7. Här skall endast noteras några speciella faktorer:

- ' ' i stickord markerar kil (avstavningsgräns).
- • i stickord markerar morfologisk gräns.
- ' ' (enkla citationstecken) markerar parafras i lista 1 och lista 3.
- (3) (siffra inom rund parentes) har i exemplen använts som markering för fingerat konstruktionssätt. Olika siffror antyder klar skillnad i syntaktiskt uppträdande.
- I ett fall anges i klartext, inom hakparentes, att verbpartikeln är betonad.
- Etymologi schablonmarkeras utom i ett fall där etymologisk hänvisning ges.

Lista 1

- ¹ grund 'föga djup' 1 adj. *grunt -are -ast*, (3), med liten utsträckning nedanför ytan spec. om vatten, även överfört
- 2 subst. *-et ∅*, (5), grund punkt, grynna, västsv. både
- 3 grund, da verb *-r -de -t -d*, *grund-*, (11), bli grundare; särskilt grunda upp bli grunt, (10), se ²upp 1 [etymologi]
- ² grund 'underlag' 1 subst. *-en -er*, (5), a existerande underlag för möjlig verksamhet ofta överfört; på grund av beroende på b tillverkat underlag för byggnadsverksamhet även överfört
c förberedande målningsskikt
- 2 grund, da verb *-r -de -t -d*, *grund-*, a stödja på viss grund överfört, (12), se 1 a b åstadkomma grund mest överfört, (11), se 1 b c grundmåla, (11), se 1 c
- 3 grund, are subst. *-n el. grundarn ∅, grundar-*, (6), en som grundar överfört, se 2 b [etymologi]
- grund, lig 'noggrann' adj. *-t -are -ast*, (3), som arbetar utgående från grunden överfört, se ²grund 1 a
- gryn, na 'bränning' subst. *-n grynnor, grynn-*, (5), undervattensklippa [jfr ¹grund]
- ² upp 'fullbordande' verbpartikel (betonad), ofta i sammansättning, (37),
- 1 så att förloppet slutföres perfektiviskt

Lista 2

- ¹grund adj. *grunt -are -ast* med liten utsträckning nedanför ytan
spec. om vatten, även överfört, (3), [etymologi]
- ²grund subst. *-en -er*
1 existerande underlag för möjlig verksamhet ofta överfört, (5);
på grund av beroende på
2 tillverkat underlag för byggnadsverksamhet även överfört, (5)
3 förberedande målningsskikt, (5) [etymologi]
- ³grund subst. *-et* \emptyset grund punkt, se ¹*grund*, (5), *gryнна, västsv.*
både
- grundda verb *-r -de -t -d, grund-*,
1 stödja på viss grund överfört, se ²*grund 1*, (12)
2 åstadkomma grund mest överfört, se ²*grund 2*, (11)
3 grundmåla, se ²*grund 3*, (11)
4 bli grundare, se ¹*grund*, (11); särskilt grunda upp bli grunt,
(10), se ²*upp 1*
- grundare subst. *-n el. grundarn* \emptyset , *grundar-*, en som grundar
överfört, se *grunda 2*, (6)
- grundlig adj. *-t -are -ast* som arbetar utgående från grunden
överfört, *noggrann*, se ²*grund 1*, (3)
- grynna subst. *-n grynnor, grynn-*, undervattensklippa, bränning,
(5) [jfr ¹*grund*]
- ²upp verbpartikel (betonad) ofta i sammansättning
1 så att förloppet slutföres perfektiviskt, (37)

Lista 3

- ¹grund 'föga djup' adj. grunt -are -ast, (3), med liten utsträckning nedanför ytan spec. om vatten, även överfört [etymologi]
- ²grund 'grund punkt' subst. -et ∅, (5), se ¹grund, grynna, västsv. både
- ³grund 'underlag' subst. -en -er, (5)
- 1 existerande underlag för möjlig verksamhet ofta överfört; på grund av beroende på
- 2 tillverkat underlag för byggnadsverksamhet även överfört
- 3 förberedande målningsskikt [etymologi]
- ¹grundda 'stödja på viss grund' verb -r -de -t -d, grund-, (12) överfört, se ³grund 1
- ²grundda 'åstadkomma grund' verb, böjning se ovan, (11), mest överfört, se ³grund 2
- ³grundda 'grundmåla' verb, böjning se ovan, (11), se ³grund 3
- ⁴grundda 'bli grundare' verb, böjnings se ovan, (11), se ¹grund; särskilt grunda upp, bli grunt, se ²upp 1
- grundare 'en som grundar' subst. -n el. grundarn ∅, grundar-, (6), överfört, se ²grunda
- grundlig 'noggrann' adj. -t -are -ast, (3), som arbetar utgående från grunden överfört, se ³grund 1
- grynna 'bränning' subst. -n grynnor, grynn-, (5), undervattensklippa [jfr ¹grund]
- ²upp 'fullbordande' verbpartikel (betonad) ofta i sammansättning, (37)
- 1 så att förloppet slutföres perfektiviskt

7. DATATEKNISK METOD

För projektet gäller en del datatekniska premisser. De har huvudsakligen två aspekter: den metodologiska och den utrustningsmässiga.

7.1. Lagringsform

Databasen lagras i en välstrukturerad form. Som grundmodell har vi tänkt oss ett *länkat nätverk* (se Allén, "Förslag till struktur för en lexikalisk databas"; jfr även Allén, "Language network as linguistic data base"), som förefaller tilltalande på flera sätt. För språkforskare med teoretiska intressen framstår alltid själva modellbygget lockande. Nätverkstanken är i sammanhanget särskilt fascinerande, eftersom det åtminstone ytligt sett verkar som om människans inbyggda lexikon har ungefär en sådan struktur. Det tycks som om flera olika dimensioner aktiveras samtidigt vid språkliga processer. En lexikonmodell på nätverksstrukturens grund borde kunna fånga något av den psykologiska realiteten.

Trots att den teoretiska sidan av saken är synnerligen spännande, överflyglas den av de praktiska implikationerna. Givet en databas lagrad som ett länkat nätverk, har man en oerhörd kontroll över sitt material. Genom att följa de olika länkarna kan databasen via textskärmsterminalen genomsökas flerdimensionellt utan egentliga väntetider. Snabb access till materialet åstadkoms genom lagring på skivminne. Den fullständiga överblick över materialet som lagringsformen möjliggör ger ytterligare försäkring om konsekvent och genomgripande bearbetning.

Flera databassystem av varierande raffinemang har utvecklats inom den kommersiella sektorn. Även för språkliga ändamål har databaser lagts upp (t.ex. termbanken i Bergen). Efter att ha tagit del av ett antal tillgängliga system, och efter att ha överlagt med flera experter på området, har vi kommit till slutsatsen att det lönar sig att bygga upp ett eget system, direkt anpassat till de egna kraven, snarare än att försöka ta över eller modifiera något existerande system. Det kan nämnas att Kolbjørn Heggstad, föreståndare för Projekt for datamaskinell språkbehandling, som är en pendang till Språkdata i Norge, var gästforskare hos Språkdata under läs-

året 1976-1977. Bergen gruppen har sysselsatt sig mycket med databasproblem, och det var synnerligen värdefulla inblickar som under denna tid förmedlades av Heggstad. Vissa förarbeten för vår egen databas har utförts (jfr nedan avsnitt 7.4).

Databasen såsom ett länkat nätverk kan under uppbyggnadstiden betraktas ur två synvinklar: *dels* kan dess etablering övervägas som ett mål i sig, dvs. den välstrukturerade databasen blir resultat av att ett (del)projekt genomförs, *dels* kan den smidiga lagringsformen vara ett hjälpmedel, ett sätt att underlätta och förbättra organisationen av en informationsmängd. Det skall betonas att detta dubbla synsätt endast gäller under initialskedet; i fortsättningen framstår nätverksstrukturen i första hand som ett överlägset instrument.

Nu befinner vi oss emellertid på det första stadiet. Sedan vi fått mera grepp om problematiken förefaller det lämpligt att bygga upp enklare delstrukturer i första omgången. Dessa kan efter hand utvecklas och förenas med varandra till ett komplexare system. Den alternativa strategin - att direkt konstruera ett komplicerat totalt databassystem - rymmer inte bara svårigheter utan även risker. Om databassystemet etableras samtidigt som material ackumuleras och analyseras, kan korrektions- och uppdateringsarbetet ställa till problem. De verkliga riskerna består i att inmatat material kan utplånas eller förloras ur sikte genom att en länk påverkas. Det ligger i sakens natur att kedjereaktioner lätt utlöses.

Härav framgår att målsättningen skall vara följande. Databasen etableras i första hand innehållsmässigt, vilket betyder att material inmatas (för proceduren se närmare nedan avsnitt 7.2). I samband med analysen av materialet upprättas delstrukturer, som i ett senare skede kan förfinas och sammanföras till komplexare nätverk. Detta tillvägagångssätt gynnas av ett annat förfarande, om vilket beslut fattats på oberoende grunder, nämligen det att materialet i stor utsträckning kommer att analyseras kategorivis. De olika delstrukturerna associeras till varandra efter hand, varvid genomgående länkar som möjliggör omedelbar sofistikerad sökning läggs in. Som följd härav blir det länkade nätverket snarast ett mål som nås mot senare hälften av projekttiden.

7.2. Interaktiva och automatiska arbetsmoment

Att det länkade nätverket blir en realitet först på ett senare stadium innebär ingen praktisk inskränkning under det uppbyggnads-skede som spänner över de första åren. Allt finns sökbart och kan alltid tas fram via *satsvisa* körningar. Beträffande mycket av informationen är detta fullt tillräckligt, och på det hela taget innebär redan denna möjlighet en betydande rationalisering jämfört med traditionella metoder. Generellt kan sägas att de maskinella bearbetningarna blir praktiskt sett relativt enkelt genomförbara därför att vi kan använda vår egen anläggning.

Hela proceduren är utpräglad maskintillvänd. Lingvisten inkodar materialet i fixt format på textskärm. Lämpligen kan de olika enheterna ur det material som ställs till förfogande presenteras maskinellt en i sänder på skärmen. Lingvisten vidtar de nödvändiga åtgärderna, som kan bestå i olika slag av analys och kodifiering, eller helt enkelt refusering av hela enheten. Man kan diskutera huruvida automatiskt genererade analysförslag skall ges i samband med presentationen. (Det kan nämnas att vid morfemundersökningen bakom NFO 4 ingångsmaterialet hade genomgått en preliminär automatisk ordledsuppdelning redan vid presentationen på skärmen.) Fördelarna får vägas mot inkodningsprogrammets komplicering, process-tidens ökning, belastningen på minnet etc. Det är ett oavvisligt krav att vi effektivt skall kunna arbeta interaktivt vid flera skärmar samtidigt. Sålunda skall exempelvis ordbankens material oavbrutet vara tillgängligt på särskild skärm. Konkordansmaterial kommer förmodligen med fördel att tillhandahållas på mikrokort.

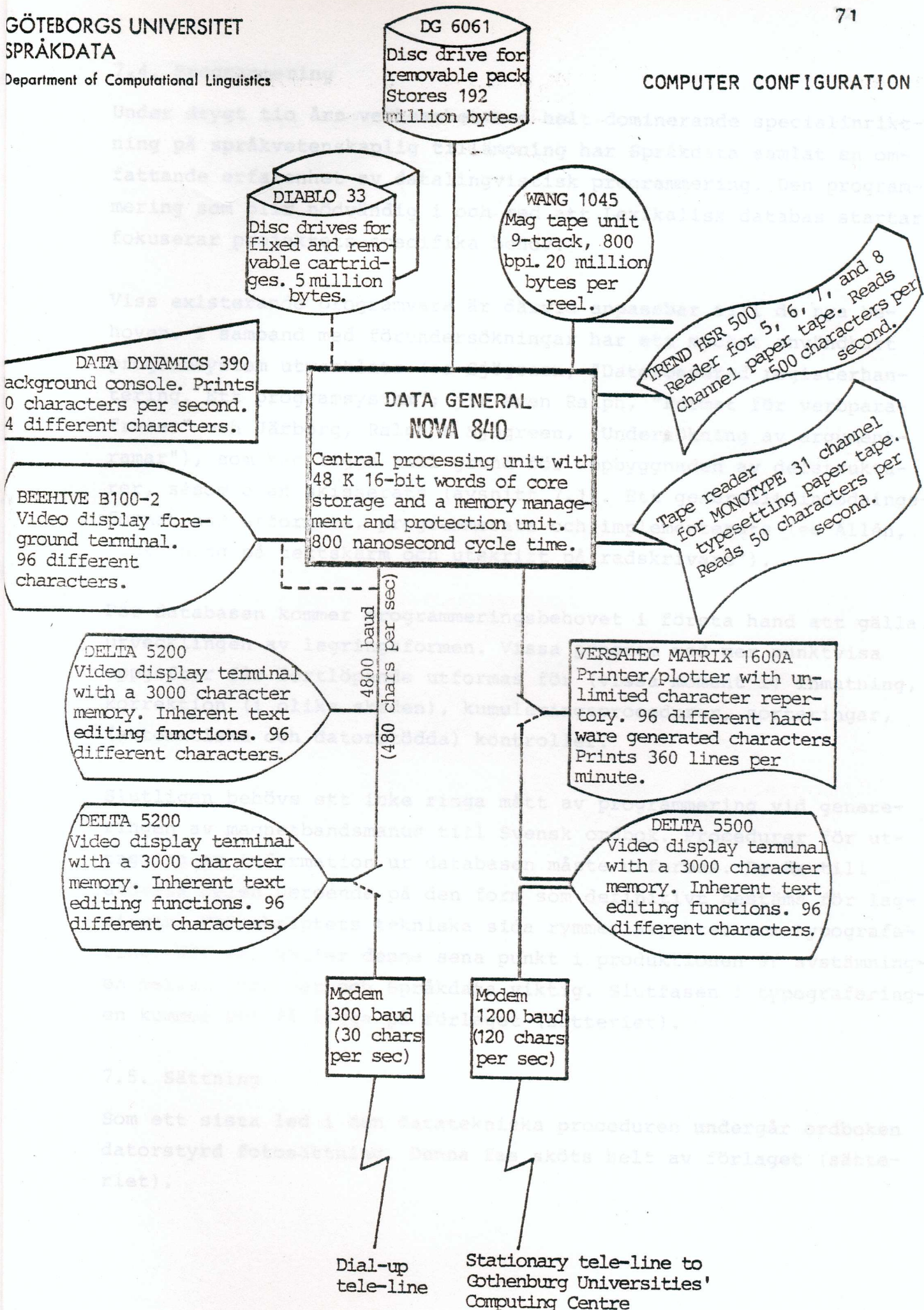
Det här förda resonemanget om databaser med direkt access till hela materialet och interaktiv bearbetning vid textskärm skall inte fördunkla det faktum att vissa moment även i fortsättningen med fördel kan utföras med satsvis maskinell behandling. Flera typer av maskinella kontroller kan tänkas ha den egenskapen. Även pappersutskrifter har delvis kvar sin betydelse. Det kan vara väl så praktiskt att få en felutskrift att kontempera över som att ta ställning till ett fall som blinkar fram på skärmen framför en. Maskinella kontroller kommer att få vidsträckt tillämpning i projektet. Särskilt kommer man att behöva testa på konsekvensfrågor,

tillåtna tecken och förkortningar m.m., men framför allt definitionernas nedbrytbarhet och definitionsvokabulärens användning samt referenssystemet framstår som mycket naturliga mål för automatisk kontroll. Andra typer av utskrifter som erfarenhetsmässigt behövs på papper är olika slags kumulering, delsorteringar m.m. Likaledes får listor tas ut på sådana delmängder av materialet som behöver kategorivis genomgång.

7.3. Maskinell utrustning

Språkdata har en egen dataanläggning som är specialutrustad för behandling av naturligt språk. Centralenheten är en NOVA 840 med ett primärminne på 48 K 16-bitsord (= 98 304 tecken). Cykeltiden är 800 nanosekunder. Kringutrustningen rymmer skivminnen med en sammanlagd kapacitet på 200 miljoner tecken, magnetbandstation, remsläsare för 5, 6, 7, 8 och 31 kanaler, fyra textskärmar varav tre med inbyggda editeringsfunktioner samt en radskrivare/kurvritare med obegränsad teckenrepertoar. Göteborgs datacentrals stora IBM-anläggning (370/148 och 360/65) utnyttjas för särskilt stora satsvisa körningar. En av textskärmsterminalerna kan kopplas via fast förbindelse till datacentralen, en annan via uppringd linje. Dessa terminaler är alltså anslutna via telenätet. Den befintliga datorkonfigurationen framgår mera detaljerat av skissen på nästa sida.

Språkdata är en institution med flera projekt och vissa servicefunktioner. Projektet Lexikalisk databas skulle behöva ha två egna textskärmar till ständigt förfogande. Viss nyanskaffning av utrustning förutsätts därför. Förmodligen behövs inga speciellt avancerade textskärmar för den sortens textbehandling som det rör sig om – det finns flera billiga typer på marknaden. Vid behov kan också de redan existerande mer sofistikerade skärmarna disponeras. Slutligen skulle den interaktiva bearbetningen (särskilt då flera användare utnyttjar anläggningen samtidigt) effektivteras markant om primärminnet utökades med åtminstone 16 K ord. Detta innebär en relativt liten justering av anläggningen som helhet, och man kunde tänka sig att projektet bidrog till denna utrustningsdetalj.



7.4. Programmering

Under drygt tio års verksamhet med helt dominerande specialinriktning på språkvetenskaplig tillämpning har Språkdata samlat en omfattande erfarenhet av datalingvistisk programmering. Den programmering som blir nödvändig i och med att Lexikalisk databas startar fokuserar projektets specifika behov.

Viss existerande programvara är därvid anpassbar till de nya behoven. I samband med förundersökningar har ett mycket användbart programsystem utvecklats (se Sjögren, "Datorbaserad registerhantering. Ett programsystem"; jfr även Ralph, "Format för verbparafrafer" och Järborg, Ralph & Sjögren, "Undersökning av argumentramar"), som kan ligga till grund för uppbyggnaden av delstrukturer, såsom ovan skisserats (avsnitt 7.1). Ett generellt inkodningsformat har utformats, programmerats och implementerats (se Allén, "Inkodning på textskärm och utskrift på radskrivare").

För databasen kommer programmeringsbehovet i första hand att gälla utvecklingen av lagringsformen. Vissa program med mer punktvisa uppgifter får fortlöpande utformas för (vissa moment i) inmatning, korrektion (i olika skeden), kumuleringsprocedurer, sorteringar, (automatiska och datorstödda) kontroller.

Slutligen behövs ett icke ringa mått av programmering vid genereringen av magnetbandsmanus till Svensk ordbok. Procedurer för uttagning av information ur databasen måste utformas. De är till största delen beroende på den form som definitivt bestäms för lagringen. Manuskriptets tekniska sida rymmer aspekter som typografering. När det gäller denna sena punkt i produktionen är avstämningen mellan förlaget och Språkdata viktig. Slutfasen i typograferingen kommer att få ligga på förlaget (sätteriet).

7.5. Sättning

Som ett sista led i den datatekniska proceduren undergår ordboken datorstyrd fotosättning. Denna fas sköts helt av förlaget (sätteriet).

LITTERATURFÖRTECKNING

./ Arbetsrapporter (bilagor)

- Allén, S. 1976-12-08. Lexikalisk databas. Projektbeskrivning.
 --. 1977-03-08. Om lexikografiska definitioner.
 --. 1977-03-25. Förslag till struktur för en lexikalisk databas.
 --. 1977-11-07. Om uppslagsord i Svensk ordbok/Lexikalisk databas.
 --. 1977-11-14. Om utformningen av artiklarna i Svensk ordbok.
 --. 1977-11-23. Inkodning på textskärm och utskrift på radskrivare.
 Berg, S., J. Järborg & B. Ralph. 1977-11-03. Synpunkter på stickord
 Gavare, R. & C. Sjögren. 1977-09-07. Databaser.
 Gellerstam, M. (Under renskrivning.) Välja sina ord.
 Järborg, J. & B. Ralph. 1977-11-24. Ordboksartiklarna: ordbildnings- och ordstruktursynpunkter.
 --. (Under renskrivning.) Mot definitionen.
 Järborg, J., B. Ralph & C. Sjögren. 1977-10-17. Undersökning av argumentramar.
 Ralph, B. 1977-01-13. Förutsättningar.
 --. 1977-01-28. Prepositioners presuppositioner.
 --. 1977-02-16. Ordböcker och ordmodeller.
 --. 1977-03-15. Kort om kasus.
 --. 1977-03-25. Verb för verb.
 --. 1977-04-20. Format för verbparafraser.
 --. Sept. 1977. Projektet Lexikalisk databas. Sammanfattning av föredrag vid Nordiska datalingvistikdagar, Göteborg.
 --. 1977-12-05. Kasusbeteckningar för ytargument.
 --. 1977-12-09. Från verbparafraser till konvertibilitetssystem.
 Sjögren, C. 1977-09-19. Datorbaserad registerhantering. Ett programsystem.

Övrig citerad litteratur

- Allén, S. 1967. Förhållandet mellan tal och skrift. I: Språk, språkvård och kommunikation. Lund: Prisma.
- . 1977-02-24. Language network as linguistic data base. Sammanfattning av föredrag inför Share European Association, London.
- . 1977. Våra viktiga ord. Basordlista med utbytesord. Stockholm: Esselte Studium.
- . 1970-1975. Nusvensk frekvensordbok. 1-3. Stockholm: Almqvist & Wiksell.
- et al. (Under arbete.) Nusvensk frekvensordbok. 4.
- Beeck, Santesson & Witthoff. 1969-04-29. (Intern PM om uttalsbeteckningar.)
- Hässelberg, A. & S. Allén. Svensk ordlista. Stockholm: Esselte Studium.
- Illustrerad svensk ordbok. 1964. Tredje, reviderade uppl. Stockholm: Natur och Kultur.
- Ralph, B. Projektet Svensk definitionsvokabulär. (Slutrapport under arbete.)
- Svensén, B. 1977-10-24. Några funderingar om typografi m.m. i Svensk ordbok.
- Svenska Akademiens ordlista över svenska språket. 1974. Tionde uppl. Stockholm: Norstedts.

