



Institutionen för svenska språket

Jerker Järborg

Betydelseanalys och
betydelsebeskrivning
i Lexikalisk databas

Preliminär version

Förord

Föreliggande skrift är avsedd som en redogörelse för de allmänna principer och mer eller mindre detaljerade regler som tillämpats i projektet Lexikalisk databas för att etablera och beskriva de semantiska lexikala enheterna och deras subenheter. Redogörelsen innefattar även en kritisk granskning av reglerna själva och av deras faktiska tillämpning. Det resulterande lexikaliska materialet betraktas primärt ur teoretiskt lexikologisk synvinkel, som en modell av det interna lexikonet (abstraherat över kollektivet av språkbrukare) medan lexikografiska aspekter får en relativt undanskymd plats. Denna behandling beror dels på personlig intresseinriktning, dels på en övertygelse att teoretiskt genomarbetade ordböcker kommer att visa sig vara överlägsna även i praktiskt-pedagogiskt avseende.

Projektet Lexikalisk databas pågick under en följd av år och involverade många medarbetare, varför det nu ej alltid är lätt att veta vem som har bidragit med vad. Då bidragsgivaren kan identifieras nämns han/hon givetvis i texten. I övrigt gäller att redan under arbetet med Nusvensk frekvensordbok 4 aktualiserades vissa grundläggande frågor om lexikonets roll och den lexikaliska betydelsens natur under täta överläggningar mellan Bo Ralph och mig. Liknande fruktbara diskussioner fortsattes och vidgades senare under de tidigare projektåren. Projektledaren Sture Allén deltog givetvis i många av dessa diskussioner, särskilt när det gällde frågor av principiell art. Genom en fullständig, kritisk genomläsning av det framväxande materialet har han också verksamt bidragit till en konsekvent regeltillämpning. Under det praktiska analys- och beskrivningsarbetet fördes en i det närmaste kontinuerlig lexikologisk diskussion, särskilt beträffande detaljregler, mellan de berörda medarbetarna, nämligen (förutom mig själv) i anciennitetsordning Sven-Göran Malmgren, Kerstin Norén och Åsa Abelin. För interpretationen och presentationen av denna projekt-interna, huvudsakligen muntliga, tradition svarar emellertid undertecknad ensam.

Göteborg i december 1989

Jerker Järborg

NB! I denna version har avsnitten 5.3 och 5.4 ej hunnit medtagas.

ett sättband för en tryckt ordbok (varvid olika kategorier "översätts" till olika stilar, eventuella typografiska gränsmarkörer

Betydelseanalys och betydelsebeskrivning i Lexikalisk databas.

1. Material och uppläggnig

Stommen i den blivande databasen utgjordes av en maskinläsbar lista av stickord, som ursprungligen skapats för Svensk baklängesordbok och bestod av unionen av Svenska Akademiens ordlista, 10 upplagan (SAOL10) och Nusvensk frekvensordbok 2 (den fullständiga lemmaförteckningen). Den maskinläsbara listan upptog drygt 140 000 formella enheter (lemman) och fungerade som ett råmaterial, utifrån vilket de lexikaliskt intressanta orden utvaldes och analyserades. Givetvis kompletterades denna lista småningom med lexikaliskt intressanta ord hämtade ur andra källor (se nedan), kanske främst med ord av fackkaraktär. Det totala antalet stickord (semantiskt behandlade eller ej) i den nu existerande databasen torde uppgå till 150 000.

Det viktigaste språkliga referensmaterialet utgjordes av Språkbankens textkonkordanser, särskilt över de två tidningsmaterialen Press 65 (1 million ord) och Press 76 (1,2 millioner ord) samt romanmaterialen Roman 76-77 (4,0 millioner ord) och Roman 80-81 (5,6 millioner ord). De mest använda lexikografiskt bearbetade referensmaterialen var Illustrerad svensk ordbok (ISO), Bonniers svenska ordbok, första upplagan, Svensk handordbok, SAOL10 (knappast för semantisk information), samt vid behov Ordbok över svenska språket, utgiven av Svenska Akademien (SAOB) och Östergren, Nusvensk ordbok. En mellanställning mellan obearbetat och bearbetat referensmaterial hade ett urval läroböcker för gymnasiet, ur vilket de viktigaste fackorden exciperades (till en början lagrade i särskilda fackfiler). Ett utomordentligt viktigt informellt referensmaterial utgjordes av medarbetarnas egen internationaliserade språkliga kunskap, vid behov kontrollerad med hjälp av informanter.

Det framväxande lexikaliska materialet avsågs läggas upp i kategoriserad form, varmed förstås att varje särskild informationstyp (böjning, uttal, definition, exempel, stilkommentar etc.) lagras som en egen variabel under lexikonenheten. Ett dylikt kategoriserat lexikaliskt material kan i princip lätt konverteras till

ett sättband för en tryckt ordbok (varvid olika kategorier "översätts" till olika stilar, eventuella typografiska gränsmarkörer införs, etc.), medan omvändningen (sättband till kategoriserat lexikon) ej nödvändigtvis kan utföras med automatiska rutiner. Det visade sig ändamålsenligt att organisera materialet såsom ett system av filer, bestående av en huvudfil och ett antal underordnade filer. I huvudfilen (kallad ID-filen) lagras endast de lexikaliska grundenheterna (lexemen), vilka i maximalfallet består av en ordkropp, ett lemmanummer och ett lexemnummer. För internt bruk förtydligas därtill lemmanumret med en rudimentär uppgift om ordklass och böjning, samt lexemnumret med en kortfattad parafras (eller synonym). Var och en av de underordnade filerna svarar mot en huvudtyp av lexikalisk information och/eller en grupp av principiella huvudmoment i det lexikaliska analys- och beskrivningsarbetet. Informationen i de underordnade filerna är länkad till de individuella lexikonenheter med hjälp av en särskild pekarfil.¹

De viktigaste underordnade filerna betecknades LEM, LEX, FRAS och FDV. LEM-filen används för information om ordklass, böjning, uttal, alternativa stavningssätt, avstavning o.d. Inmatningen av uttalsangivelser skedde huvudsakligen via en särskild subfil till LEM-filen. Eftersom informationen i LEM-filen endast undantagsvis berördes av semantiska överväganden, lämnas LEM-filen utanför den följande beskrivningen. (Observera dock att ett avgörande huruvida t.ex. ett givet substantiv kan ha plural eller ett givet adjektiv kan kompareras måste träffas på semantiska grunder.) FRAS-filen, som till en början användes för att samla kandidater till lexikaliserade idiom, dvs. flerordiga, mer eller mindre fasta ordsammansättningar, vars betydelser ej rimligtvis kan härledas ur de ingående ordens betydelser, kommer att beröras endast i förbigående, eftersom dess enheter har en avvikande karaktär och är förknippade med en speciell urvalsproblematik (som knappast kan sägas vara slutgiltigt löst inom databasens ram). De följande avsnitten avser nästan uteslutande den semantiska huvudfilen FDV.

¹Filsystemet är internt känt som Arb-systemet; för internationellt bruk har beteckningen The Lexibase System lanserats. Systemet är i sin helhet konstruerat av Christian Sjögreen (se Sjögreen 1988), huvudsakligen på grundval av en skiss av Jerker Järborg (Järborg 1978).

Vid arbetets inledning utkristalliserade sig snart ett antal principiella huvudmoment av semantisk art, nämligen

- a) Utväljande av lexikaliskt (ofta = semantiskt) intressanta lemmen för vidare bearbetning.
- b) Uppdelning av polysema (/homonyma) lemmen i klart urskiljbara betydelseenheter (lexem).
- c) Uppdelning av lexemens betydelsefält i kärnbetydelse och perifer (s.k. transfererade) betydelser.
- d) Excerpering av strategiska exempel som (i) kan bestyrka den föreslagna analysen, (ii) utgör problemfall som måste lösas under det fortsatta analysarbetet, (iii) utgör underlag för utväljande eller konstruktion av de slutgiltiga exemplen.
- e) Excerpering av möjliga lexikaliserade idiom.
- f) Beskrivning av lexemens kärnbetydelser.
- g) Beskrivning av de transfererade betydelserna och deras samband med motsvarande kärnbetydelser.
- h) Utväljande och/eller konstruktion av relevanta exempel som illustrerar kärnbetydelsernas och de transfererade betydelsernas användning.
- i) Utväljande av intressanta lexikaliserade idiom, jämte betydelsebeskrivning av dessa.

Som synes grupperar sig momenten a) - i) ovan på ett naturligt sätt i "förberedande" och "slutgiltiga". Vid den ursprungliga uppläggningsen av arbetet ansågs att de lingvistiska projektmedarbetarna skulle överbelastas av att samtidigt behöva utföra dessa väsentligen skiljaktiga typer av uppgifter. Därför uppdelades momenten a) - i) på en för internt bruk avsedd förberedande fil, LEX, en slutgiltig fil, FDV, samt en idiomfil FRAS, avsedd både för förberedande och slutgiltiga moment. LEX-filen var i princip avsedd för momenten a), b), c) (tentativt) och d), samt i viss utsträckning e). FDV-filen var avsedd för momenten c) (definitivt), f), g) och h), medan FRAS-filen var avsedd för e) och i). Grundtanken var att lexemen skulle etableras i LEX-filen och förses med tillräckligt fyllig kommentar och exemplifiering för att filen skulle kunna utgöra huvudsakligt underlag vid arbetet med FDV-filen. Lexemen skulle sedan noga beskrivas och exemplifieras i FDV-filen. Beskrivningen skulle vara tillräckligt noggrann för att tillfredsställa de vetenskapliga kraven, samtidigt som filen utan

innehållsliga förändringar skulle kunna användas som ordboks- 388, manuskript. Enligt den ursprungliga planen skulle vidare olika medarbetare producera LEX-fil och FDV-fil för en given bokstav, detta för att minimera risken för subjektivitet. FRAS-filen var tänkt att i en första genomgång användas för en extensiv materialsamling, ur vilken senare det definitiva idiomförrådet skulle kunna erhållas huvudsakligen genom strykningar. - Det bör observeras att lexemen i ID-filen normalt etableras först i och med att motsvarande lemman analyseras i LEX-filen (varvid parafras(er) och, vid behov, lexemnummer, inmatas i databasen).

Den här beskrivna uppläggningsen ändrades i slutskedet därhän att alla förberedande moment eliminerades. Lexemen både etablerades och beskrevs i FDV-filen, som även kom att i huvudsak ersätta FRAS-filen. Dessa förändringar, som i princip innebar en viss sänkning av ambitionsnivån, vidtogs främst på grund av tidsbrist. I slutskedet visade det sig också svårt att förena FDV-filens teoretiska och praktiska funktioner. Det framväxande ordboksmanuskript som genererades ur FDV-filen granskades av vissa utomstående intressenter vilka hade en tendens att kritisera åtskilliga teoretiskt välmotiverade definitioner (och andra betydelsebeskrivningar) för svårbegriplighet och läsarovänlighet. Den version av FDV-filen som existerade vid projekttidens utgång kan nog sägas vara i (delvis alltför) hög grad anpassad till sin praktiska funktion som potentiellt ordboksmanuskript. I konsekvens härmed har de två funktionerna numera separerats. Ordboksmanuskriptet, inklusive idiomerna, skall nu genereras från en särskild fil, SOB, som även inkluderar idiomerna, medan FDV-filen reserveras för vetenskapliga användningar. De ursprungligen identiska filerna kommer småningom att divergera, beroende på lexikografiskt motiverade ändringar i SOB-filen och lexikaliskt-semantiskt motiverade ändringar i FDV-filen.

I projektets målsättning ingick från början att definitions-systemet skulle göras semantiskt sökbart, genom att de använda definitionsorden skulle normaliseras (ges sin lexikaliska uppslagsform) och disambigueras (ges korrekt lemma- och lexemnummer). För detta ändamål skapades en särskild fil, kopplad till definitionerna i FDV-filen. Detta arbete medhans icke under projekttiden, men utförs nu (med högre ambitionsgrad) i fortsättningsprojekt, med

hjälp av nya filer inom systemet. För beskrivning se Järborg 1988, Sjögreen 1988.

2. Teoretiska förutsättningar

2.1 Redan i de allra tidigaste projektutkasterna förutsattes att databasens enheter på uttryckssidan skulle etableras enligt den s.k. lemmamodellen. Ett lemma kan kortfattat beskrivas som en grupp ordformer som tillhör samma ordklass, ingår i samma böjningsparadigm och har (så långt möjligt) samma uttal, och på grund härav naturligen kan representeras av en gemensam grundform eller uppslagsform. (För detaljer hänvisas till Allén 1970, Allén 1971.) För att alltså två olika lemman med samma grundform skall etableras fordras skiljaktighet i ordklass, böjningsparadigm eller uttal, medan däremot semantiska skillnader ej beaktas.² Följaktligen kan de olika betydelserna inom ett lemma uppvisa högst varierande grad av semantisk släktskap (ned till total frånvaro av släktskap, som t.ex. vid by 'grupp gårdar'; 'vindstöt'). En sådan semantisk variation inom ett lemma kallas polysemi, vilket strider mot traditionellt lingvistiskt språkbruk. Under varje lemma (enhet på uttryckssidan) etableras de urskiljbara huvudbetydelserna såsom enheter på innehållssidan (kallade lexem); dessa utgör databasens egentliga grundenheter. Denna princip (semantisk indelning av formellt avgränsade enheter) har döpts till lemma-lexem-modellen. (Det bör påpekas att termerna lemma och lexem ej är internationellt standardiserade; de används här och i fortsättningen i den betydelse som etablerats inom LDB.)

Den främsta fördelen med lemma-lexem-modellen är att den är lätt-tillämpad i det praktiska arbetet och enkelt kan transformeras till en överskådlig, strikt alfabetisk lexikografisk representation. Lättillämpligheten är en följd av att endast två, av varandra oberoende, indelningsgrunder används, vilket innebär att enheterna på uttryckssidan relativt enkelt kan etableras på ett tidigt stadium, före den tidskrävande semantiska analysen. Som jämförelse

²Observera dock att ett böjningsparadigm måste anses primärt vara sammanhållet av systematiska semantiska likheter; uttryckslikheterna kan vara minimala (som vid verb som få, etc.). Det är därför knappast korrekt att hävda att lemmat är en rent formellt upprättad enhet.

kan nämnas att den traditionella lexikografin brukar arbeta med dels en (överordnad) historisk-etymologisk, dels en formell, dels en innehållslig indelningsgrund. Dessa indelningsgrunder är givetvis ej oberoende (även ordklasstillhörighet har en semantisk aspekt) och har visat sig svåra att applicera på ett konsekvent sätt.³ Troligen lämpar sig den traditionella modellen bäst för ett lexikon över morfem, eftersom (rot)morfemet är den naturliga enheten i ett etymologiskt perspektiv.

Den främsta nackdelen med lemma-lexemmodellen torde vara att den saknar medel att registrera semantisk släktskap mellan lexemen, där sådan föreligger på det synkrona planet. Härvidlag är den traditionella etymologiska modellen av naturliga skäl överlägsen i vissa fall. Dock behöver en gemensam härstamning för en viss grupp lexem ej nödvändigtvis medföra någon rimlig grad av synkron semantisk släktskap inom gruppen, och även där släktskap föreligger, behöver den ej vara av den enkla, hierarkiska form som låter sig beskrivas med hjälp av en historisk ursprungsbetydelse. Bättre än en återgång till den etymologiska modellen är en utbyggnad av lemma-lexemmodellen som ger möjlighet att indikera synkron semantisk släktskap (vilken tycks vara en språkpsykologisk realitet som måste beaktas). I en utbyggd modell bör lemmat, i någon form, fortfarande vara den tekniskt överordnade enheten (varigenom de ordklassberoende semantiska faktorerna hålls konstanta), men lexemen grupperas på de sätt som motiveras av de specifika släktskapsförhållandena. I vissa fall kan man tänka sig lexemen inordnade under ett eller flera "superlexem" (≈ ngn vag, gemensam betydelse, som ej nödvändigtvis behöver anges explicit), i andra fall kan det röra sig om ett eller flera "centrallexem" som på olika sätt är förbundna med de mer perifera, utan att vara direkt överordnade. Ett exempel på det senare fallet är prepositionen av, vars "centrallexem" skulle kunna parafraseras med djupkasuset Source (ärva av sin far), vilket har samband med lexemet Agent (boken är skriven av NN), men också, mer indirekt, med lexemet Object (utbyggnaden av kärnkraften). En utbyggnad av lemma-lexemmodellen enligt dessa antydningar fordrar uppenbarligen dels en mycket noggrann indelning i välbeskrivna lexem, såsom underlag för

³ För en detaljkritik av den traditionella etymologiska indelningsprincipen hänvisas till Allén 1967.

analysen av semantiska släktskap, dels en genomtänkt beskrivningsmodell för de olika typerna av släktskap, dvs den kan endast företagas sedan materialet väl etablerats enligt den ursprungliga lemma-lexemmodellen.

2.2 Den självklara förutsättningen för all lexikologi är antagandet att det existerar ett förråd av permanenta, diskreta begrepp som är lexikaliserade, dvs kopplade till likaledes diskreta enheter på uttryckssidan, samt visar så god överensstämmelse mellan olika språkbrukare att det är rimligt att tala om (och söka beskriva) det kollektiva lexikonet.⁴ Vad som närmare skall förstås med ett lexikaliserat begrepp, eller lexem i LDB:s terminologi, är emellertid icke självklart, ej heller hur man skall integrera det teoretiska antagandet om diskreta lexem med de faktiska observationerna av autentiska textmaterial, vilka ofta tycks tyda på ett semantiskt kontinuum. En grundläggande distinktion måste göras mellan det statiska förrådet av lexem och den användning den enskilde språkbrukaren gör av lexemen för att begreppsligt strukturera en viss specifik tanke.⁵ Eftersom lexemförrådet hos varje enskild språkbrukare måste vara ändligt, är det naturligt att antaga att det kompletteras av en uppsättning överlexikala modifieringsregler, som gör det möjligt att använda ett lexem även för besläktade begrepp. Det är också rimligt att tänka sig att även modifieringsreglerna visar god överensstämmelse över kollektivet av språkbrukare. Dylika antaganden ligger till grund för LDB:s allmänna modell för lexemet. På det praktiska planet bygger modellen också på erfarenheter från arbetet med Nusvensk frekvensordbok 4 (NFO 4), i vilken en snarlik modell för morfembetydelser användes med framgång.

I LDB:s modell antages varje lexem ha en väl identifierbar och relativt pregnant kärnbetydelse, kring vilken ett antal besläktade

⁴Det har dock hänt att detta okontroversiella antagande bestridits av allmänlingvister, som hävdar att ord får all sin betydelse av kontexten. Det är ej klart hur tolkning av ett språkligt yttrande egentligen skulle försiggå i en sådan modell.

⁵En djupgående diskussion om lexikonets funktion i språkproduktion och språkperception återfinns i Järborg 1984, som i mycket bygger på personliga erfarenheter av lexikaliskt forskningsarbete.

underordnade betydelser, kallade transfererade betydelser, kan man gruppera sig. De underordnade betydelserna kan härledas ur kärnbetydelsen med hjälp av någon eller några av de generella semantiska transfereringsreglerna. (Dylika härledningarna skall tolkas synkront, som en kartläggning av lexikonets potential, ej diakront-etymologiskt.) Databasen kan givetvis endast förteckna de någorlunda väl etablerade transfererade betydelserna, men genom att många transfereringsregler i princip skulle kunna appliceras på varje kärnbetydelse, ger modellen utrymme för ett visst mått av lexikalisk kreativitet; i detta avseende är modellen alltså ej strikt synkron. De faktiska användningarna av ett lexem kommer således att bilda en semantisk räjong, vars centrum motsvarar kärnbetydelsen och vars periferi bestäms av de längsta tillåtna transfereringskedjorna.

Av både teoretiska och metodologiska skäl är det nödvändigt att antaga att fullständig lexikalisk synonymi ej förekommer (eller, översatt till praktiskt användbara termer, att två olika ord alltid måste antagas ha olika betydelse tills motsatsen bevisats). I LDB:s modell innebär antagandet att alla kärnbetydelser måste vara distinkta; däremot får olika lexems semantiska räjonger överlappa varandra partiellt. Därmed upprättas en skillnad mellan lexikalisk (eller statisk) synonymi, vilken modellen ej tillåter, och synonymi i faktisk användning (dynamisk synonymi), vilken modellen tillåter vid transfererade (vanligen utvidgade eller försvagade) betydelser.

En viktig konsekvens av modellvalet är att en av de mest använda metoderna för lexikalisk semantisk analys, nämligen metoden med s.k. utbytesprov, måste underkännas på teoretiska grunder. I denna metod (som bl.a. använts för SAOB) testas man de olika användningarna av ett givet ord genom att man försöker utbyta ordet mot någon av dess närsynonymer. Om ett utbyte med "oförändrad betydelse" är möjligt i en viss typ av användningar, skall ett separat och distinkt betydelsemoment (lexem) upprättas, som har den testade närsynonymens betydelse. Det klassiska motargumentet mot utbytesprovet är att det är helt beroende av tillgången på närsynonymer, vilken kan vara mycket ojämnt spridd över lexikonet. Givet LDB:s modell faller utbytesprovet på att det sammanblandar statisk och dynamisk synonymi.

I LDB:s modell är alltså lexemet en enhet med välbestämd kärna men vaga gränser. Eftersom de perifera användningarna endast indirekt behöver täckas av kärnbetydelsen kan denna få vara pregnant (dvs fullständigt täcka alla centrala användningar). Därigenom blir det också lättare att satisfiera fordringarna på antisynonymi enligt ovan. (I en modell som ej gör skillnad på kärna och periferi riskerar betydelseerna att urvattnas till vaghet eftersom hela räjongen måste antagas kunna täckas av en enhetlig betydelse.) Genom att modellen å andra sidan medger underbetydelser till en kärnbetydelse blir det enkelt att undvika en osystematisk uppsplittring på separata betydelsemoment. (En dylik atomistisk beskrivning kan också lätt uppkomma i modeller som ej skiljer på kärna och periferi, om tonvikten läggs på de perifera i stället för de centrala användningarna.) Sammanfattningsvis kan om LDB-modellen sägas, att den tenderar att ge relativt få lexem, med pregnanta kärnbetydelser, och relativt många underbetydelser. Därigenom kan modellen ge en viss inblick i lexikonets dynamik. Nackdelen är att modellen, genom att den ställer stora krav på analys och beskrivning, blir mer svårtillämpad än modeller som huvudsakligen är inriktade på ren klassificering av betydelse.

Ett annat självklart antagande vid lexikologiskt arbete är att ordbetydelser (eller lexem) är beskrivbara, oavsett vilken ontologisk status man vill tilldela dem. Inom LDB skärptes detta antagande enligt följande: (i) lexemen kan ej beskrivas exakt men väl approximeras med godtycklig precision, givet tillräckligt långa beskrivningar, (ii) beskrivningar i (kvasilogiskt) symbolspråk kan ej uppnå samma precision som beskrivningar i ord, (iii) det är alltid möjligt att utforma en betydelsebeskrivning i objektspråket (så att beskrivningen kan utbytas mot lexemet). - Beträffande punkt (ii) kan påpekas att ännu intet symbolspråk har konstruerats som är kapabelt att beskriva hela lexikonet eller alla typer av betydelsenysanser.⁶ - Det inses lätt att punkterna (i) - (iii) innehåller ett principiellt krav på att alla lexem i databasen skall beskrivas genom noggranna definitioner. De praktiska konsekvenserna av detta krav diskuteras i kapitel 4 och 5.

⁶Till frågan om definitionssystemets globala cirkularitet återkommer vi i kapitel 5.

Antagandena om beskrivbarhet har emellertid även konsekvenser för analysarbetet, vilket kommer att framgå av nästa kapitel.

Av det ovanstående har framgått vilken betydelsefull roll de semantiska transfereringsreglerna har i LDB:s lexikaliska modell. Egentligen skulle därför en fullständig beskrivning av (det teoretiska) lexikonet bestå dels av en förteckning över alla lexem, dels en explicit redogörelse för transfereringsreglerna. I idealfallet borde därtill preliminära transfereringsregler föreligga innan den egentliga betydelseanalysen påbörjades. I praktiken visar det sig emellertid svårgörligt att uppställa dylika regler utan att kontinuerligt testa dem mot ett framväxande lexemmaterial. Vid LDB-projektets inledning uppställdes därför endast ett antal generella analysprinciper, vilka inkluderade vissa transfereringsprinciper. Under arbetets gång utvecklades sedan dessa principer i riktning mot större konkretion och högre stringens, så att de i nuvarande läge kan betraktas åtminstone som ett partiellt regelsystem. För vissa frekventa typer av transfereringar har och fullständiga detaljregler kunnat formuleras (se vidare avsnitt 3.2). Dock finns de faktiskt använda transfereringsreglerna (med nämnda undantag) ingenstädes dokumenterade utan måste extraheras ur databasens förråd av transfererade betydelser, efter korrigerande av kvarstående inkonsekvenser. Detta är ett stort arbete, som givetvis ej kunde genomföras under projekttiden.⁷ Tills vidare existerar tyvärr LDB:s semantiska transfereringsregler endast implicit, genom de underbetydelser som upprättats till olika lexem och de mer eller mindre utförliga beskrivningar som ges i databasen beträffande sambandet mellan kärnbetydelse och underbetydelse.

Även om det givetvis hade varit önskvärt med explicit angivna transfereringsregler, bör man hålla i minnet att det centrala objektet för den lexikaliska beskrivningen är lexemet, ej regelsystemet. En alltför stark tonvikt på regelsystemet, med åtföljande krav på fullständig konsekvens, skulle kunna leda till betydelse-

⁷ För en första ansats till en dylik analys, se Malmgren 1988. Se även Abelin 1988, för en liknande analys inom ett begränsat fält. I Järberg 1988 föreslås en metod för strukturell analys av transfereringar.

analyser som delvis nonchalerade specifika egenskaper hos det enskilda lexemet, till förmån för en abstrakt systematik.

3. Analys av betydelser följande klassificering av användningarna hos lemmat läge anföras:

LDB:s betydelseanalys hade tre principiella mål: (i) att uppdelade databasens lemmen i distinkta lexem, (ii) att strukturera lexemens betydelseomfång i kärnbetydelser och (perifera) transfererade betydelser, (iii) att utskilja de väsentliga (speciellt de särskiljande) semantiska faktorerna hos kärnbetydelser och transfererade betydelser. I praktiken går det sällan att skilja på de tre delmålen: (i) och (ii) måste ofta lösas i ett sammanhang, medan (iii) normalt är en förutsättning för (i) och (ii) vid sidan av sin huvudfunktion att utgöra underlag för betydelsebeskrivningen. Den faktiska analysproceduren ägde rum i form av en växelverkan mellan fyra moment: semantisk klassificering av belagda användningar, semantisk faktoranalys av de ansatta kärnbetydelserna, tentativ semantisk beskrivning (provdefiniering etc.) av kärnbetydelser och transfererade betydelser, testning av de provisoriskt beskrivna betydelserna. Dessa arbetsmoment beskrivs nedan i detalj.

3.1 Materialet för den semantiska klassificeringen hämtades dels från de textkonkordanser som nämndes i kapitel 1, dels från exempelsamlingarna i tidigare ordböcker (särskilt Illustrerad svensk ordbok och Svensk handordbok), dels ur lexikologernas egen fatabur; det sistnämnda ofta för att fylla uppenbara luckor i det övriga materialet eller för att konstruera problemfall för analysen. (Däremot bortsågs i princip från den semantiska indelningen i tidigare ordböcker.) Enligt resonemanget i kapitel 2 skulle nu exemplen på olika användningar av ett visst lemma ej fördela sig likformigt över ett semantiskt kontinuum utan naturligen gruppera sig kring ett antal "typanvändningar", vilka alltså kan motsvara kärnbetydelser eller (väletablerade) transfererade betydelser. Klassificeringsarbetet kom följaktligen att bestå i att, så fullständigt som möjligt, gruppera materialet kring typanvändningar, att indela dessa i kärnbetydelser och transfererade betydelser samt att söka bestämma de semantiska räjongerna kring kärnbetydelserna. I detta moment kan endast intuitiva skäl ges för de träffade valen. Om t.ex. två starka typanvändningar vardera kan knytas till varsin grupp av mer speciella användningar, medan det finns få eller inga gränfall mellan de två huvudtyperna, finns

det skäl att ansätta två kärnbetydelser med endast nätt och jämnt överlappande räjonger. a ngt

Som en illustration kan följande klassificering av användningarna hos lemmat mage anföras:

De två klara typanvändningarna är givetvis (a) och (1). Till (a) (1) läter sig (b), (c) och (d). Till (f) är (g) nära och (h) nära. (a) tjock mage en, medan skrytningen till (1) och (1) är svagare, rund om magen skiljbar. Idiomat (k) anknäver naturligen till (1). Platt mage och slanka lår de två huvudgrupperna som kan ena en krans av fikonlöv kring magen (a), vilket naturligt skulle sparka ngn i magen om åkomma än en i matsmältningsorganen. Ölmage klassificeringen och grupperingen blev alltså att två kärnbetydelser ansattes, motsvarande (a) - (d), respektive (a) - (k).

(b) ligga på mage

Det centrala momentet i betydelseanalysen är givetvis den (c) en baby i magen de ansatta kärnbetydelserna. Den semantiska faktorn skall icke förstås någon semantisk primitiv, av typ soman-

(d) (cigarrens) maggördel (1), primitive predikat (CAUSE, etc.) eller djuphasus (Agent, etc.), även om dylika primitiver ofta

(e) ont i magen jämförelser på de semantiska faktorerna som urskilja (1) synnerhet djuphasusen har fungerat som ett gott hjälp-

(f) magen full med mat semantisk faktor hos en given kärnbety-

hungrig så det skriker i magen ses som en beståndsdel av

vägen till mannens hjärta går genom magen dess semantiska

en armé marscherar på sin mage följer att faktorerna på mycket

kons fyra magar nivå kan urskiljas, beroende på det ana-

magsäck ste karaktär. Själva kärnbetydelserna förutsätter ha en

magsjukdom struktur och varje urskiljd semantisk faktor skall

alltså ingå som en byggsten i denna struktur. (Faktorerna får så-

(g) krånglande mage (additivt)

orolig i magen

stark mage i sista hand återgående på Aristoteles brukar

nervös mage ses i termer av överbegrepp ("genus proximum")

och underbegrepp (specifika + förhållande djur, etc.), varvid bestäm-

(h) hård/lös mage ses till det minimum som fordras för att

hård/lös i magen begreppet från andra underbegrepp till

överbegreppet i fråga. En dylik analysmetod skulle påvisat mate-

(i) magsur (kritik) förenklat hierarkisk semantisk struktur och

avslöjat riskerna att begreppet blir tillräckligt explicit eller ut-

(j) magstark

ha mage att göra ngt

(k) fjärilar i magen

De två klara typanvändningarna är givetvis (a) och (f). Till (a) ansluter sig (b), (c) och (d). Till (f) är (g) nära och (h) tämligen nära ansluten, medan anknytningen till (i) och (j) är svagare, ehuru klart urskiljbar. Idiomet (k) anknyter naturligen till (g). Det enda gränsfallet mellan de två huvudgrupperna som kan uppställas kring (a) respektive (f) är (e), vilket marginellt skulle kunna avse någon annan åkomma än en i matsmältningsorganen. Resultatet av klassificeringen och grupperingen blev alltså att två kärnbetydelser ansattes, motsvarande (a) - (e), respektive (e) - (k).

3.2 Det centrala momentet i betydelseanalysen är givetvis den s.k. faktoranalysen av de ansatta kärnbetydelserna. Med semantisk faktor skall icke förstås någon semantisk primitiv, av typ semantiskt särdrag (+ Animate, etc.), primitivt predikat (CAUSE, etc.) eller djupkasus (Agent, etc.), även om dylika primitiver ofta fungerar väl som beteckningar på de semantiska faktorer som urskiljs. (I synnerhet djupkasusen har fungerat som ett gott hjälpmedel vid analysen.) En semantisk faktor hos en given kärnbetydelse är ett begrepp som måste ansättas som en beståndsdel av kärnbetydelsen för att dess användning inom hela dess semantiska räjong skall kunna förklaras. Härav följer att faktorer på mycket varierande abstraktionsnivå kan urskiljas, beroende på det analyserade lexemets karaktär. Själva kärnbetydelsen förutsätts ha en inre semantisk struktur och varje urskild semantisk faktor skall alltså ingå som en byggsten i denna struktur. (Faktorerna får således ej betraktas rent additivt.) Traditionellt (och i sista hand återgående på Aristoteles) brukar ett begrepp analyseras i termer av överbegrepp ("genus proximus") och bestämningar (människa = tänkande djur, etc.), varvid bestämningarna gärna reduceras till det minimum som fordras för att särskilja det analyserade begreppet från andra underbegrepp till överbegreppet i fråga. En dylik analysmetod skulle påtvinga materialet en orealistiskt förenklad hierarkisk semantisk struktur och resultatet riskerar att varken bli tillräckligt explicit eller ut-

tömmande. Även i LDB:s analysmodell måste vanligen (dvs. vid substantiv och verb) en av de semantiska faktorerna ges en särställning, nämligen den att vara "högsta nod" i kärnbetydelsens inre struktur, men utan överbegreppets egenskap att i princip omfatta all betydelse hos det analyserade lexemet utom den direkt särskiljande. I LDB:s modell kan de semantiska faktorerna indelas å ena sidan efter strukturell funktion, i vad som lämpligen kan kallas grundkategori (högsta nod) och karakteriseringar (övriga faktorer), å andra sidan efter semantisk vikt i primära, sekundära etc. faktorer, varvid de mindre väsentliga faktorerna ej behöver vara obligatoriska utan må vara vanliga, typiska e.d. Somliga av de lägre värderade faktorerna kommer snarast att motsvara väl-etablerade s.k. konnotationer hos lexemet i fråga; till svaga eller rent individuella konnotationer tas givetvis ingen hänsyn.

Det bör observeras att ingen principiell åtskillnad görs mellan inomspråkliga (rent "lexikala") och utomspråkliga (s.k. encyklopediska) faktorer. I själva verket är det mycket vanligt att ett lexems språkliga användning behöver förklaras med hjälp av encyklopediska faktorer, låt vara av mycket allmän art (så t.ex. kan verb som falla och t.o.m. prepositioner som på, under i sista hand endast förklaras medelst hänvisning till tyngdkraften). Mer specifika encyklopediska faktorer behövs naturligtvis för vissa begränsade kategorier, t.ex. nationalitetsadjektiv, liksom även för vissa typer av konnotationer (punschpatriotism o.d.). Däremot skall sådana begreppsliga distinktioner som grundas på utomspråkliga kriterier (systematik inom ett givet fackområde etc.) ej betraktas som semantiska faktorer, om de icke kan motiveras språkligt, inom den språktyp som skall täckas lexikaliskt. I LDB:s fall rörde det sig närmast om allmänspråk, utökat med elementärt (populärvetenskapligt) fackspråk. Vid gränsfallen urskildes visserligen encyklopediska faktorer, men rankades lågt. För beskrivningen av dylika marginella encyklopediska faktorer utvecklades speciella metoder (se vidare kapitel 5).

Vid den faktiska analysen erhöles de sökta faktorerna framför allt på tre sätt: genom jämförelser med närsynonymer samt med andra ansatta lexem under samma lemma, genom specificeringar av länkarna mellan ansatt kärnbetydelse och ansatta transfereringar inom det givna lexemet samt genom uppställande av följdsatser (av typen: Om ngt är en X så gäller att...). Genom jämförelser erhålls givetvis

de särskiljande faktorerna. Vid undersökningen av en transferering tas transfereringsreglerna och en intuitiv föreställning om den transfererade betydelsen som givna; uppgiften blir då att uppställa en faktor hos kärnbetydelsen som kan tas som utgångspunkt för länken mellan kärnbetydelse och transfererad betydelse. De sålunda erhållna faktorerna tenderar att vara allmänt karakteriserande snarare än direkt särskiljande. Metoden med följsatser kan användas för att finna både särskiljande och karakteriserande faktorer, men är speciellt lämplig för optionella karakteriserande faktorer.

I praktiken finns givetvis ett nära samband mellan momenten faktoranalys och semantisk beskrivning; en faktor måste ju beskrivas på något sätt för att kunna identifieras och användas. Sambandet blir särskilt tydligt vid utväljande av den faktor som skall utgöra grundkategori; denna skall nämligen, i en första, mycket grov approximation, kunna ersätta det analyserade lexemet, åtminstone i typanvändningen. Om K är den föreslagna grundkategorin, kan denna direkt testas i exemplen med hjälp av schablonparafraser som t.ex. ett slags K (vid substantiv) respektive K:a på visst sätt (vid verb). Trots detta är det ej alltid lätt att bestämma grundkategorin; ett vanligt problemfall är sådana verb som kan parafraseras antingen utföra X så att Y blir resultatet eller åstadkomma Y genom att utföra X.

Som exempel på en rimligt långt driven faktoranalys tar vi åter fallet mage.

- (2)
- 1.1 på kroppens framsida (undersida)
- 1.2 undre (bakre) mjuka delen
- 1.3 gärna rundad
- 1.4 gärna sammanfallande med största kroppsomfång
- 1.5 något sårbar
- 1.6 innehåller det mesta av matsmältningen
- 1.7 innehåller andra intima kroppsjorgan
- 2.1 inre organsystem
- 2.2 bearbetar intagen föda
- 2.3 utstöter restprodukter av visst slag
- 2.4 belägen huvudsakligen i nedre (bakre) delen av kroppen
- 2.5 funktionen påverkbar av yttre och inre störningar

2.6 god funktion viktig för välbefinnandet

2.7 kan betraktas som självständigt verkande (Agent)

Här avses 1.1 - 1.7 täcka kärnbetydelsen (a) i exempel (1) och dess räjong, 2.1 - 2.7 kärnbetydelsen (f) och dess räjong. Man kan notera att användningen (b) innebär en fokusering på 1.1, att (d) kan förklaras med utgångspunkt i 1.4, att (g) innebär en fokusering av funktionen (2.2, 2.5, 2.6) och att (h) tar sin utgångspunkt i 2.3. De längre gående transfereringarna (i) och (j) bygger nästan enbart på 2.6 respektive 2.5, vilka tolkas i överförd bemärkelse. Det verkar således som om de angivna faktorerna kan täcka de anförda exemplen på ordets användning. Det återstående problemet gäller huruvida det är korrekt att uppdelas mage på två lexem. Till en början kan det konstateras att de enda faktorer i grupperna 1 och 2 som visar överensstämmelse är paret 1.2 - 2.4. Redan existensen av faktorn 2.7 är ett kraftfullt, kanske avgörande, argument mot en sammanhållning. Viktigare är dock faktorer som är väsentliga för transfereringar: å ena sidan 1.4, å andra sidan 2.5 och 2.6. Dessa faktorer framhäver klart den essentiella skillnaden mellan de tentativa lexemen: det första en kroppsdel, vars (yttre) form kan fokuseras; det andra ett makilda organsystem vars funktion kan fokuseras. Med så många skiljaktiga och så få överensstämmande faktorer kan ej ett enda sammanhållet lexem upprättas utan att alltför stora anspråk skulle ställas på transfereringsreglerna. I föreliggande fall är det därtill nödvändigt att utvälja helt olika grundkategorier för grupperna 1 och 2 (som antytts ovan), 1.2 (+ 1.1) respektive 2.1.

I praktiken är det sällan möjligt eller ens nödvändigt att driva faktoranalysen så långt som i det anförda exemplet, särskilt icke i de många fall där analysens huvudsyfte är att ge argument för hur många lexem som ett givet lemma skall uppdelas i. För avgörande av dylika fall fordras tillgång till de viktigaste faktorerna, jämte några enkla evalueringsregler. De regler som användes inom LDB formuleras nedan för fallet ett eller två lexem (kan lätt generaliseras). Under förutsättning av att faktorerna ej kan härledas ur varandra genom transfereringsregler gäller att två lexem upprättas om minst två faktorer skiljer sig mellan de tentativa kärnbetydelserna. Om en tung faktor skiljer sig, upprättas likaså två lexem (kriteriet på en tung faktor är att den används som utgångspunkt för typiska, frekventa transfereringar). Olika grund-

kategori ger (under förutsättningen ovan) två lexem. För transfereringsreglerna gäller vissa spärrar: från en och samma kärna får ej transfereras på helt motstridiga sätt och en transfererad betydelse får ej utgöra utgångspunkt för nya transfereringar annat än i den ursprungliga riktningen (icke-transitivitet). Observera att fallet mage enligt samtliga regler måste lösas som två lexem. Observera även att användningarna (i) och (j) bygger vidare på (g) "i samma riktning", men att (f) ej får betraktas som transfererad från (a) (eller vice versa); då skulle (f) strida mot både (b) och (c), samtidigt som (f) skulle bli utgångspunkt för nya transfereringar.

Trots att grundprincipen är att varje lemma måste analyseras på sina egna villkor, finns det ett antal standardproblem som kan lösas kategorivis med hjälp av en förenklad faktoranalys. En inventering av de flesta standardproblemen utfördes av medarbetarna S-G Malmgren och K. Norén (se Malmgren-Norén 1982), vilken låg till grund för de schablonregler som sedan användes för att avgöra okontroversiella fall av typen ett eller två lexem. Nedan ges de viktigaste av schablonreglerna. Observera att speciella omständigheter kan ha föranlett avvikande behandling av vissa enskilda fall (allmän konsekvens får aldrig bli viktigare än exakt semantisk analys).

Ett lexem upprättas bl.a. i följande typfall (som alltså kan betraktas som standardtyper av transfereringar):

- o Substantiv som betecknar verksamhet samt Agent för denna (utredning) eller Object för denna (läsning) eller Instrument för denna (betalning).
- o Substantiv som betecknar egenskap samt individ som är bärare av denna (talang, alt) eller enstaka manifestation av denna (cynism, dumhet).
- o Substantiv som betecknar sinnestillstånd samt ngt som är upphov till detta (hopp, besvikelse).
- o Substantiv som betecknar djur eller växt samt produkt därav (mink, ek).
- o Substantiv som betecknar behållare e.d. samt (mängden av) det innehåll som ryms i denna (gryta, tesked).
- o Verb som betecknar enstaka handling samt vanemässig upprepning av denna (dansa, röka).

o Verb som betecknar en sammansatt handling som involverar både en direkt deltagande (mindre) företeelse och en slutligen påverkad (större) företeelse, där endera kan topikaliseras genom val av direkt objekt (muddra en ränna/hamninloppet, torka pannan/svetten ur pannan, knäppa knapparna/jackan).

o Verb som betecknar en handling samt, genom partikeltillägg, dess motsats (knäppa upp, koppla isär, montera ner).

o Adjektiv som betecknar en egenskap mätbar i viss dimension e.d. samt dimensionen i allmänhet (bred/2 meter bred).

o Adjektiv som betecknar själslig egenskap samt ngt som uttrycker denna en ilsken skomakare/ett ilsket brev, en vänlig man/ett vänligt ord).

Två lexem upprättas bl.a. i följande fall:

o Verb som betecknar ett förlopp samt handlingen som består i att orsaka detta förlopp (bollen rullade ut/målvakten rullade ut bollen, inflationen ökar/räntehöjningen ökar inflationen). Vid denna typ, som kan kallas ±Cause, kan andra lexemet schablonparafraseras med "få att..." (få att rulla, etc.).

o Adjektiv som betecknar samband med viss företeelse samt egenskapen att kännetecknas av hög grad av, skicklighet i, etc., denna företeelse (musikaliska akademien/en musikalisk person, en diplomatisk beskickning/en diplomatisk replik). Vid denna typ kallas första lexemet klassificerande och kan parafraseras "som har att göra med..."; andra lexemet kallas karaktiserande och kan parafraseras "som kännetecknas av...".

Hittills har det förutsatts att analysens mål var att finna och upprätta samtliga lexem under ett givet lemma, under förutsättning att lemmat självt var lexikaliskt intressant. Såsom ointressanta räknades främst sammansättningar och avledningar med i princip helt förutsägbara betydelser (angående urvalskriterier se Ralph--Järborg-Allén 1977). Emellertid visade det sig att vissa sammansättningar (inklusive partikelverb) och avledningar skulle kunna uppdelas i såväl helt förutsägbara lexem som lexem med egen, pregnant betydelse. T.ex. kan lemmat hästminne tänkas betyda 'minnesförmåga hos häst', 'enstaka hågkomst hos häst', 'enstaka hågkomst av häst', 'souvenir från (verksamhet med) hästar', förutom den pregnant betydelsen 'mycket gott minne'. I sådana fall uteslöts de triviala lexemen från analysproceduren, sedan deras förekomst

noterats. Endast de pregnanta lexemen upprättades, men försågs med en särskild kod som angav att lemmat även hade flera andra betydelser (och som i ordboken återgavs med förkortningen särsk. i inledningen till lexemdelen).

3.3 Det tredje arbetsmomentet i betydelseanalysen, den tentativa semantiska beskrivningen, har redan berörts i samband med faktoranalysen. Syftet med betydelsebeskrivningen var, som nämnts, att skapa underlag för att testa faktoranalysen mot de preliminärt klassificerade språkliga exemplen. För detta syfte var det i allmänhet tillräckligt att (approximativt) parafrasera kärnbetydelsen hos ett ansatt lexem samt att i princip ange hur de ansatta transfererade betydelseerna skulle kunna härledas ur kärnbetydelsen. (Ungefär denna typ av semantisk beskrivning användes i den sedermera övergivna LEX-filen.)

I motsats till den slutliga definitionen behöver parafrasen ej vara syntaktiskt utbytbar mot definiendum. Däremot bör de viktigaste semantiska faktorerna vara representerade på något sätt och det bör vara möjligt att antyda hur de övriga faktorerna skulle kunna integreras. Givetvis måste grundkategorin vara korrekt. För de transfererade betydelseerna anges vilken faktor i kärnbetydelsen som utgör utgångspunkt ("fokus på faktor X", etc.). För ordvalet gäller ej heller samma restriktioner som för definitionen: så t.ex. är teknisk vokabulär tillåten och undvikande av cirkularitet är ej relevant. Däremot är parafraser i form av synonymer mindre lämpliga för testningsändamål, även i de fall där de fungerar förträffligt som särskiljande etiketter på lexemen.

Det fjärde momentet, testningen, skedde i första hand mot de exempel som redan använts vid den semantiska klassificeringen, i andra hand mot nyfunna eller nykonstruerade exempel, gärna av grännsfallskaraktär. Vid alla tester var kravet att parafrasen (vid behov kompletterad med någon antydd transferering) skulle vara fullständigt semantiskt utbytbar mot definiendum i exemplen (med den självklara reservationen att intet analytiskt uttryck någonsin kan vara språkpsykologiskt ekvivalent med det odelade begrepp det avses beskriva). Förutom direkt jämförelse mellan exemplet och dess substituerade version fanns också alltid möjligheten att jämföra följsatser av de två versionerna (se 3.2 ovan); denna senare metod var särskilt viktig när det gällde utvärdering av de mer sub-

tila distinktionerna. I okontroversiella fall kunde lexikologen fungera som sin egen informant, men vid behov konsulterades givetvis andra personer, vanligen projektmedarbetare men i svåravgörbara fall även utomstående.

Den ovan beskrivna analysproceduren upprepades i princip så många gånger som fordrades för att målsättningarna (i) - (iii), som angavs i inledningen till kapitlet, skulle uppnås. I allmänhet var de centrala delarna av vokabulären (i synnerhet starka verb, "grammatiska" prepositioner samt verbpartiklar) mest arbetskrävande, medan många ord av lexikaliskt mer perifer karaktär endast behövde göras till föremål för en högst rudimentär analys. Det bör observeras att de interna betydelsebeskrivningar som används inom analysproceduren tenderar att successivt förfinas i riktning mot verkliga definitioner, etc., för varje gång proceduren måste upprepas. Därför blev slutresultatet ej sällan att de mest svåranalyserade lexemen fick de mest genomarbetade definitionerna.

4. Betydelsebeskrivningen: formella aspekter

4.1 De genom analysproceduren etablerade lexemens kärnbetydelser beskrevs primärt, och obligatoriskt, med en definition, dessutom optionellt med ett definitionstillägg. Eventuella transfererade betydelse beskrevs med vardera en obligatorisk transfereringsförklaring och en optionell parafras. Därtill kunde betydelsen givetvis karakteriseras och avgränsas på olika indirekta sätt. Bland dessa kan nämnas användningskommentarer (stil, bruklighet, fackområde etc.), exempel, (när)synonymer, motsatsord, jämförelseord (allmänt beslätade), hänvisningar, m.fl. Varje informationstyp lagrades, som nämnts i kapitel 1, som en egen variabel i filen (ursprungligen FDV-filen), vilken sedan kodades typografiskt i ordboken. Det bör observeras att alla exempel, användningskommentarer, etc. registrerades som hörande till antingen kärnbetydelsen eller en särskild, transfererad betydelse och grupperades tillsammans med den relevanta betydelsen i ordboken. Härigenom fick ordboksartiklarna en cyklisk karaktär (med huvudavdelningarna betydelseangivelse - användningskommentar - exempel), vilket ledde till att de transfererade betydelsena även kom att kallas "cykler" (ett lexem kunde t.ex. sägas ha en kärna och tre cykler).

Av de anförda beskrivningskategorierna var definitioner och parafraser skrivna i objektsspråk, dvs. med ord som skulle kunna ersätta definiendum (under vissa förutsättningar). Synonymer, motsatser o.d. var givetvis trivialt objektsspråkliga. Definitionstillägg, transfereringsförklaringar och användningskommentar skrevs i metaspråk, dvs. med ord som beskrev den semantiska användningen av definiendum utifrån, på en mer abstrakt nivå. I definitioner och parafraser fick förkortningar ej användas, då eljest den objektspråkliga karaktären delvis kunde gå förlorad. I övriga beskrivningskategorier användes många både allmänna och speciella förkortningar, för att spara utrymme och i någon mån inmatningsarbete. Som generellt stort skiljetecken användes semikolon, både i databas och ordbok. I definitionstillägg, användningskommentar, transfereringsförklaring etc. skildes alltså information av olika art med semikolon. Semikolon användes även mellan exemplen. I definitioner och parafraser fick givetvis stora skiljetecken ej förekomma; däremot användes tydlighetskommatering vid behov. De parenteser som uppträder i ordboken är i många fall gränsmarkörer för olika informationstyper som genererats av typograferingsprogramet. Parenteser inom en och samma informationstyp (t.ex. definitionstillägg) hör däremot till databasen och användes på samma sätt som i ordinärt skriftspråk, med två undantag. I exempel markerade parenteser optionella ord eller delar av ord. I definitioner (och sekundärt parafraser) gällde speciella regler för användningen av parenteser; se vidare 4.2 samt kapitel 5.

För att (för en mänsklig läsare) minimera definitionernas ambiguitet, vidtogs vissa speciella åtgärder. I de fall där definitionen kunde varieras syntaktiskt valdes givetvis den minst ambigüosa ordföljden. Tydlighetskommatering är redan nämnd; den användes särskilt då en relativsats skildes från sitt huvudord av en tellankommande bestämning. Relativsatser utan utsatt relativpronomen skulle i princip ej förekomma. I de fall där frekventa, homografa definitionsord hade heterografa ekvivalenter, skulle dessa användas: sålunda ersattes konjunktionen som med såsom och adverbet för med alltför. I många fall kom dock dessa tydlighetsåtgärder i konflikt med stilistiska krav, särskilt i vad avsåg databasens praktiska användning. Därför är tydlighetsåtgärderna ej konsekvent genomförda; det gäller framför allt utbytet av som mot såsom.

4.2 På definitionens syntaktiska konstruktion ställdes det grundläggande kravet att den skulle vara fullständigt utbytbar mot definiendum i alla syntaktiska (och även morfologiska) kontexter, vid behov efter applikation av någon naturlig syntaktisk transformation. Detta krav var givetvis i sista hand underordnat de semantiska kraven (se kapitel 5) men kunde dock uppfyllas i det stora flertalet av fall, trots att "definiering" med synonym i princip ej var tillåten (endast sådan definiering kan garantera syntaktisk utbytbarhet). De syntaktiska transformationer som möjliggjorde utbytbarheten fanns ursprungligen ej specificerade utan fick rekonstrueras ur definitionens formulering, ibland med hjälp av parenteser i definitionen eller metaspråkliga anvisningar i definitionstillägget (se vidare nedan). I allmänhet är det trivialt att supplera den behövliga transformationen, som t.ex. vid adjektiv i attributiv ställning:

- (3) äcklig 'som vållar äckel'
äckliga matrester → matrester som vållar äckel
Transformation: Adj N → N Def(Adj)
(Beträffande predikativ ställning, se nedan!)

I svårartade fall, vid ord tillhörande de slutna ordklasserna, kunde syntaktisk utbytbarhet endast uppnås med hjälp av tämligen långsökta och mer eller mindre idiosynkratiska transformationer. Vissa särskilda kategorier av ord har vidare, av andra skäl, systematiskt definierats på sätt som omöjliggör full syntaktisk utbytbarhet. Slutligen har, i enstaka fall, definitionen helt fått utgå, ersatt av en metaspråklig betydelsebeskrivning (något som innebar en avvikelse från projektets ursprungliga målsättning). De olika typerna av problemfall bör dock till allra största delen kunna lösas, särskilt i en databas som ej längre i första hand är tänkt som ett ordboksmanuskript.

En följd av utbytbarhetskravet är att varje ordklass får en särskild typ av definitionsformat. Nedan ges en uppställning av de viktigaste principformaten, ordklass för ordklass.

o Substantiv definieras med en obestämd nominalfras (NP). Undantag utgörs dels av verbalabstrakter, vilka definieras det att + Verb(inf), dels egenskapsbeteckningar, vilka ofta definieras det att vara + Adj. Eftersom dessa typer är bestämda NP blir de ej ut-

bytbara i bestämd form.⁸ Samma problem uppstår vid vissa definitioner inledda av vetenskapen, läran e.d.

Ett liknande problem för den syntaktiska utbytbarheten utgörs av de definitioner (särskilt av artnamn o.d.) som inleds av obestämd artikel, som i

(4) **lunnefågel** 'en svartvit sjöfågel med hög, i sidled hoptryckt näbb'

Uppenbarligen är dylika definitioner ej utbytbara i obestämd form.

Bortsett från de anförda fallen behövs speciella åtgärder vid insättning av substantivdefinitioner endast då ett argument finns markerat, antingen med parenteser, som i

(5) **fiende** 'motståndare (till ngn eller ngt) som är beredd att skada'

eller på annat sätt, som i

(6) **moster** 'syster till viss persons mor'

I kontexter med argumentet utsatt (i de anförda fallen genitiv alternativt till-fras) fordras alltså en "argumentinsättningstransformation" som i och för sig är lätt att specificera. Dylika fall diskuteras utförligare vid verbdefinitionerna, för vilka de givetvis är mer karakteristiska.

Den inre strukturen hos en typisk substantivdefinition är tillfrå-sina huvuddrag tämligen enkel. Definitionen kan inledas med någon eller några adjektiviska bestämningar, därefter följer det substantiviska huvudordet, därefter ofta någon eller några prepositionsfraser och/eller en relativsats. Ibland används en infinitivfras i stället för en relativsats. Den vanligaste strukturen kan sammanfattas med schemat AdjP* N Prep* (RelS/InfP)* (där

⁸ Verbalabstrakterna uppmärksammas i Ohlander 1987, som även innehåller andra intressanta synpunkter på utbytbarhetsproblematiken.

stjärnan står för godtyckligt antal förekomster, från 0 och uppåt, medan snedstreckat anger alternativ). Ett typexempel är

(7) **myrstack** 'stackformig hög av barr, små kvistar och andra växt-
växt- delar som utgör bostad för myrsamhälle'

i vilket alla huvudpositioner i schemat är fyllda. Alternativet med infinitivfras kan exemplifieras med

(8) **uppehållstillstånd** 'tillstånd för utlänning att vistas i
visst land'

Faktiskt förekommande substantivdefinitioner kan vara åtskilligt mer komplexa än exemplen antyder och ej sällan syntaktiskt ambigösa. Två typiska källor för ambiguitet är samordningar (jfr ex. (4)) och närstående prepositionsfraser, vilka senare ju kan utgöra en direkt bestämning till huvudordet eller härröra från en expansion av en annan PrepP. Ett mer systematiskt ambiguitetsproblem, unikt för substantivdefinitioner, utgörs av de många fall där det syntaktiska (eller formella) huvudordet ej är avsett att vara egentligt, semantiskt huvudord i definitionen. Sådana "formella huvudord" står först i definitionen, med det egentliga huvudordet i en direkt följande prepositionsfras (normalt inledd med av). Överlägset vanligast är typ (av), som i

(9) **antilop** 'typ av medelstort, smäckert hovdjur'

Andra konstruktioner med formellt huvudord är sammanfattning(en) (av), system (av), uppsättning (ofta utan preposition), endera (av två), m.fl. Även titel (för) får räknas hit, trots att det är fråga om en delvis annorlunda kategori. Observera att vissa av de anförda orden ibland måste räknas som egentliga huvudord (då med något avvikande konstruktion). Jämför här

(10) (a) **befästningsverk** 'system av befästningar'

(b) **järnväg** 'system för transporter...'

där system i (b) måste betraktas som egentligt huvudord. Motiveringar för distinktionen mellan syntaktiska och semantiska huvudord ges i kapitel 5.

o Adjektiv definieras på två sätt: antingen genom en relativsats eller med en adjektivfras. I det senare fallet råder givetvis full utbytbarhet. Vid definition med relativsats, vilket var huvudalternativet och särskilt begagnades för semantiskt centrala adjektiv, fordras transformationer för att uppnå utbytbarhet. Den enkla flyttningstransformation som används vid insättning av definitionen i attributiv ställning har redan berörts (se exempel (3) ovan). Vid predikativ ställning är direkt insättning omöjlig, varför ett "tomt" huvudord till relativsatsen måste suppleras:

(3') (a) Matresterna är äckliga.

(b) *Matresterna är som vållar äckel.

(c) Matresterna är något som vållar äckel.

Även vid adjektivdefinitioner kan argument finnas markerade, på liknande sätt som vid substantivdefinitioner, vilka alltså får beaktas vid insättning i kontexter med utsatta argument.

Av de två huvudtyperna av adjektivdefinitioner har adjektivfrasen normalt en enkel inre struktur. Ofta består adjektivfrasen antingen av ett adjektiviskt huvudord med framförställd adverbialbestämning eller av två samordnade huvudord. Ibland expanderas adjektivfrasen med en prepositionsfras, en infinitivfras, en attsats e.d. Definitioner i form av relativsatser har givetvis en mycket mer varierande inre struktur. Dock inleds de så gott som alltid av relativpronomen, vilket vanligen är som, ibland vars. Relativsatsen får ej ha formen som är + Adj (avvikelser från denna regel förekommer). De flesta klassificerande adjektiv är definierade i något av schablonformaten som har att göra med + X (vanligast) eller som avser + X, där X står för ett abstrakt substantiv eller ett egennamn (särskilt ortnamn). I övrigt gäller att adjektivdefinitioner i form av relativsatser sällan har så stort syntaktiskt djup som de mera komplexa substantivdefinitionerna; sålunda förekommer bestämningar i form av underordnade relativsatser mycket sparsamt.

o Som substantiv eller adjektiv definieras pronomen och räkneord. De flesta förled som upptagits såsom självständiga lexikonerheter definieras som adjektiv. Pronomen är en heterogen ordklass och sådana pronomen definitioner kan ibland fordra tämligen idiosynkratiska transformationer för att bli syntaktiskt utbytbara. I motsats

härtill är räkneord mycket överskådliga. I princip finns två definitionsformat för räkneord: ett substantiviskt för kardinaltal och ett adjektiviskt för ordningstal:

- (11) (a) tre 'grundtalet som är två plus ett'
(b) tredje 'som är nummer tre i ordningen'

Som synes står den substantiviska definitionen i bestämd form. Den syntaktiska utbytbarheten torde vara acceptabel, om än ej perfekt. Förled definierade som adjektiv får god morfologisk utbytbarhet i kombination med substantiviska slutled (men endast där).

o Verb definieras som infinitivfraser med olika bestämningar, dels sådana av adverbiell art (ofta i form just av adverb), dels sådana med karaktär av argument, i form av rena nominalfraser, prepositionsfraser eller ibland adjektivfraser. Därtill innehåller verbdefinitioner också syntaktiska "luckor", avsedda för insättning av de argument som måste realiseras explicit vid faktisk användning av verbet. S.k. transitiva verb kommer alltså i allmänhet att ha två "luckor" i sin definition: en för det syntaktiska subjektet, en för det syntaktiska direkta objektet. För att en verbdefinition skall vara syntaktiskt utbytbar fordras att dylika luckor finns angivna och kan identifieras, så att de explicita argumenten kan sättas in på korrekt plats (genom vad som ovan kallats argumentinsättningstransformationer).

En naturlig lucka för subjektet uppstår redan därigenom att definitionen är formulerad i infinitiv. Vid s.k. opersonliga verb (regna etc.) gavs en explicit upplysning om att luckan endast fylls med det. Härvid utnyttjades den möjlighet som finns att ge syntaktisk och morfologisk information på lexemnivå, genom en formell kommentar (här lydande iopers. konstruktion). - Den formella kommentaren var tänkt att användas i de fall där ett enskilt lexems syntaktiska och/eller morfologiska uppträdande avviker från vad som kan prediceras från de allmänna lemmauppgifterna om ordklass och böjning. Här kan således anges t.ex. om ett visst lexem knappast förekommer i pluralis eller förekommer mest i sammansättningar, lika väl som syntaktiska upplysningar (som kommer att exemplifieras ytterligare nedan). - För angivande av det syntaktiska objektet (eller annat/andra nödvändiga komplement) har ett antal

olika, mer eller mindre tydliga metoder använts. De viktigaste är följande:

(i) Direkt utsättande av argumentet inom parentes, som i

(12) **slakta** 'avliva och stycka (husdjur) för att utnyttja köttet'

Typfallet torde vara när argumentparentesen innehåller semantisk information som lämpligen meddelas på den syntaktiskt adekvata platsen, såsom i exemplet. Det förekommer dock även att parentesen innehåller en dummy, nämligen ngn eller ngt (vilka är de enda tillåtna förkortningarna i definitionstext). I sådana fall är argumentparentesen endast en platshållare för luckan. Givetvis får parentesen även innehålla en argumentpreposition vid behov, som i

(13) **kalasa** 'äta (på ngt) med välbehag'

Det förekommer också att argumentparentesen föregås av en argumentpreposition (i de fall då definiendum tar ett direkt objekt som i definitionen motsvaras av ett prepositionsobjekt), som i

(14) **uppleva/l** 'vara med om (ngt) som direkt berörd part'

vilket kan sägas vara en övergångsform till nästa typ.

(ii) Utsättande av argumentpreposition, utan tillhörande nominalfras, som i

(15) **kalkylera** 'beräkna det troliga utfallet av'

Formatet med fri argumentpreposition används nästan enbart i slutet av definitionen; om annan argumentplacering är nödvändig används i regel formatet i (14). Undantagsvis står en fri argumentpreposition inne i definitionstexten, som i

(16) **åberopa** 'hänvisa till såsom stöd för sin uppfattning'

I enstaka fall markeras ett argument med två argumentprepositioner.

förutom förse med NP också göra AdjP, få att InfP m.fl. Ehuru den-

(17) **uppfatta/1** 'bli medveten om genom att ta emot och bearbeta (vissa) yttre, fysiska retningar från'

Något vanligare är att argumentet markeras implicit (underförstås) i en position och markeras med argumentpreposition i en annan.

(18) **urskulda** 'försvara genom att anföra förmildrande om ständighet(er) för'

Detta format utgör en övergångsform till typ (iii).

Det bör observeras att den argumentpreposition som diskuterats är utvald på semantiska grunder, för att beskriva förhållandet mellan definitionens (huvud)verb och ett explicit (syntaktiskt) angivet argument. Om detta argument syntaktiskt manifesteras som ett (mer eller mindre obligatoriskt) prepositionsobjekt⁹, behöver givetvis objektets syntaktiskt nödvändiga preposition ej överensstämja med argumentprepositionen. För angivelse av syntaktiskt nödvändig preposition (vilken kan uppträda även vid formaten (i) och (iii)) används den formella kommentaren, på sätt som specificeras nedan.

(iii) Implicit markering av argumentet, via definitionens huvudverbs syntax, som i

(19) **adressera** 'förse med adress'

Eftersom förse fordrar syntaktisk komplettering (eg. snarast med ett s.k. dativobjekt) följer därav att även adressera fordrar ett syntaktiskt realiserat argument, som i detta fall måste sättas in i den underförstådda argumentluckan (efter förse) för att definitionen skall vara syntaktiskt utbytbar. Denna typ av markering av argumentluckan torde vara den vanligaste, bland annat på grund av att den som regel används vid några högfrekventa definitionstyper:

⁹ Kategorin prepositionsobjekt är ej erkänd av alla syntaktiker. Den traditionella beteckningen är givetvis adverbial; man laborerar sedan med "verb som tar obligatoriskt rumsadverbial" och liknande. Projektet Talsyntax införde kategorin objektadverbial (!) för de fall där de traditionella beteckningarna visade sig otillräckliga.

förutom förse med NP också göra AdjP, få att InfP m.fl. Ehuru denna markeringstyp är tydlig nog för en mänsklig läsare (och ofta förtydligas ytterligare med hjälp av definitionstillägget; se vidare nedan) är den givetvis i ännu mindre grad än de tidigare formellt välbestämd, vilket minskar databasens användbarhet i vissa datalingvistiska tillämpningar.

Den inre strukturen hos verbdefinitionerna har redan delvis presenterats och exemplifierats. Det återstår att nämna att huvud verbet ofta föregås av ett (någon gång två) "operators"-verb, av typ råka, låta, försöka, att enklare eller kortare adverbiala bestämningar ofta står framför verbet/verben, samt att strukturen tenderar att vara "bred" snarare än "djup", med många adverbial och/eller argument under huvud verbet. Även om varje argument i princip skulle kunna expanderas till godtyckligt djup är det sällsynt att en verbdefinitions struktur blir mycket djupare än den i ex. (17), vilket innebär att den i djup överträffas av komplexa substantivdefinitioner.

Adverb definieras vanligen som prepositionsfraser, dvs. en preposition följt av en NP som ofta står i bestämd form. Ibland kompletteras definitionen av en "argumentparentes" som snarast har semantiskt förtydligande karaktär och i princip skall strykas vid insättning i en syntaktisk kontext:

(20) över 2/1 'i riktning upp (från ngt)'

Bortsett från denna komplikation är adverbdefinitioner normalt fullständigt syntaktiskt utbytbara. Den inre strukturen är sällan särskilt komplex. Ibland definieras adverb med andra adverb och då oftast med ren s.k. synonym, vilket givetvis helt säkerställer den syntaktiska utbytbarheten.

Prepositioner definieras i principformatet Prep NP Prep, vilket alltid torde vara utbytbart mot definiendum. Ibland används ett ord som tekniskt sett är subjunktion (underordnande konjunktion) i stället för den ena av prepositionerna; särskilt (så) som, som i

(21) mot 1/4 'som ersättning för'

Definition med synonym förekommer ibland, speciellt vid "för-
längda" prepositioner sådana som utav etc.

o Subjunktioner (underordnande konjunktioner) definieras vanligen
med prepositionsfras följd av att,

(22) såvida 'för det fall att'

Dock fordras ej sällan speciallösningar för att något slags utbyt-
barhet skall kunna uppnås. Sålunda definieras det ovan nämnda
såsom i prepositions- och adverbformat (i olika lexem) och det
"allmänt underordnande" att måste definieras i ett eget, unikt
format.

o Konjunktioner (samordnande) är en så heterogen grupp att indi-
viduella definitionsformat blir nödvändiga. Någon enstaka konjunk-
tion (t.ex. ty) definieras i subjunktionsformat.

4.3 Det kan vara lämpligt att avsluta detta kapitel med en över-
syn av den syntaktiska informationen i databasen. Ursprungligen
avsågs databasen även innehålla tämligen detaljerad lexemspecifik
syntaktisk information, som ett komplement till den självklara
morfologiska och semantiska informationen. I FDV-filen reservera-
des ett särskilt fält för syntaktisk konstruktionsangivelse, tänkt
att ges i något slags formell notation. Av tidsbrist kunde dessa
planer aldrig genomföras. Den generella syntaktiska informationen
i databasen är för närvarande av två slag: dels den mycket all-
männa som ges (för varje lemma) i form av obligatorisk ordklass-
angivelse, dels den implicita som ges av definitionens format och
som presenterats i föregående avsnitt. Denna tämligen vaga syn-
taktiska information kompletteras i många fall (dock ej helt sys-
tematiskt) med lexemspecifik information som kan ges (i) som for-
mell kommentar, (ii) i definitionstillägg eller transfererings-
förklaring, (iii) med hjälp av s.k. syntaktiska exempel.

Den formella kommentarens användning har delvis beskrivits ovan;
den är, som sagts, att betrakta som en inskränkning av de möjlig-
heter som lemmats ordklass- och böjningsangivelser antyder. Vissa
adjektiv är sålunda begränsade till antingen endast attributiv
eller endast predikativ användning, som i

(23) **släkt 2** endast predikativt 'som tillhör samma släkt (som)'

(Observera att i ordboksartiklarna den formella kommentaren föregår definitionen, en ordning som kommer att följas i de anförda exemplen.)

Vid verb brukar behovet av syntaktisk information vara störst. Typen iopers. konstruktion har redan nämnts. En annan, mycket frekvent, typ avser prepositionsobjektets preposition (i de fall där den ej är identisk med definitionens fria argumentpreposition), som i

(24) **stå 2/7** med prep. för 'beteckna'

eller, då flera prepositioner kan tänkas

(25) **lita** med prep. 'ha fullt förtroende (för)'

(Här kunde, alternativt, de båda prepositionerna på och till angetts explicit.)

En annan, mycket frekvent typ av formell kommentar avser verb som sällan förekommer i aktiv form, som

(26) **besmitta** vanl. pass. el. perf. part. 'utsätta för smitta'

Många fler typer av formell kommentar med syntaktisk information förekommer i databasen men de torde alla vara tämligen självförklarande. Ibland ges explicit syntaktisk information av samma slag som i den formella kommentaren i ett kommentarfält hörande till en transfereringscykel; i ordboken återges informationen då inom vinkelparenteser, efter betydelsebeskrivningen.

I motsats till den formella kommentaren är definitionstillägg och transfereringsförklaringar avsedda för semantisk information men kan även innehålla indirekt syntaktisk information. Definitionstillägget ger ofta upplysningar om argument e.d. som p.g.a. det syntaktiska utbytbarhetskravet ej får tillhöra den egentliga definitionen. I dylika fall kan man säga att definitionstilläggen fungerar som ett alternativ till argumentparenteser (s.k. syntaktiska parenteser) i definitionerna. Vid en verbdefinition där

argumentluckan för objektet (el. likn.) anges enligt metod (ii) eller (iii) ovan, (dvs. med fri argumentpreposition eller implicit) och där luckan är belägen i slutet av definitionen, får argumentet semantiskt beskrivas först i definitionstillägget, utan särskild markering. Definition och tillägg är alltså avsedda att läsas i följd, som i

(27) **föranleda** 'ge anledning till' ngn handling el. åtgärd

(I ordboksartiklarna följer definitionstillägget direkt efter definitionen och åtskiljs genom att sättas i petit. I exemplen ut-

Ofta är det ej möjligt att utforma definitionen så att argumentluckan kommer sist. I dylika fall markeras i definitionstillägget sådan information som gäller (det syntaktiska) objektet med uttrycket med avs. på, som i

(28) **besiktiga** 'granska på fackmässigt sätt' med avs. på djur, maskin, byggnad etc.

Det gäller alltså att uttrycket med avs. på i definitionstillägget alltid signalerar existensen av ett syntaktiskt objekt.¹⁰ (Däremot gäller givetvis ej omvändningen.) På samma sätt används om i verbdefinitionstillägg för att markera information som gäller det syntaktiska subjektet och i adjektivdefinitionstillägg för att markera information som gäller adjektivets huvudord. Uttrycket betr. används ibland i verbdefinitionstillägg för att markera information som gäller något argumentliknande komplement till verbet, skilt från det syntaktiska (direkta) objektet. Uttrycket i fråga om markerar däremot mer generella samband som ej kan specificeras i syntaktiska termer.

De anförda syntaktiska markörerna (med avs. på, etc.) används, när så fordras, även i transfereringsförklaringarna, som i

¹⁰Härvid ansluter sig LDB och Svensk ordbok till bruket i SAOB. För klarläggningen av SAOB:s principer på denna vitala punkt svarade S.-G. Malmgren.

(29) **uppbära särsk.** '(regelbundet) ta emot' pengar etc.; ofta som rättighet e.d.

/.../ med konstruktion avväxling 'placera väsen först på

äv. med avs. på annat än pengar

(Transfereringsförklaringen följer efter exempel på kärnbetydel-
sen, vilka här utelämnats. I ordboken markeras en transfererings-
cykel med en ofylld kvadrat.)

De s.k. syntaktiska transfereringsförklaringarna ger, som synes.
Här innehåller dock transfereringsförklaringen ingen ny syntaktisk
information, däremot ny semantisk information om det syntaktiska
objektet. Om den semantiska förändringen samtidigt innebär en för-
ändring av konstruktionssätt, händer det ofta att transfererings-
förklaringen formuleras i klart syntaktiska termer. En av de van-
ligaste typerna avser transitiva verb som även konstrueras re-
flexivt, som i fråga i kapitel 5.

(30) **värma 2** 'göra varm(are)' specifik syntaktisk information är
sed /.../ exempel. Exempelen i databas (och ordbok) har primärt
semantiska av. refl. tioner, vilka närmare beskrivs i kapitel 5. Givet-
vis måste dock varje exempel illustrera en faktisk konstruktions-
En liknande typ avser normalt transitiva verb som även konstrueras
utan objekt; här används ofta uttrycken äv. i absolut anv./kon-
struktion. De flesta andra "syntaktiska" transfereringsförklaring-
ar är formulerade på liknande, lättolkat sätt. Ett speciellt for-
mat, som för sin tolkning även fordrar en transfereringsparafra,
används ofta vid sådana verb som kan ta endera av två (varandra
kompletterande) aktanter i handlingen som syntaktiskt objekt.
T.ex. förutsätter verbhandlingen att lasta både något slags gods
och något slags transportmedel; endera av dessa aktanter kan
framhävas genom att göras till direkt objekt.¹¹ Detta förhållande
beskrivs sålunda:

Det och krystat intryck. Därtill kommer att data-
basens användning som ordboksmanuskript fordrade att hänsyn togs
till utrycksproblematiken (ett skäl som numera bortfallit). Sam-
manfattningvis kan sägas att exemplen i allmänhet torde ge en re-
lativt fullig bild av lexemens semantik, inklusive deras syntak-
tiskt markerade argument, men endast illustrationer av deras
mest typiska konstruktionsföretag. De rent syntaktiska exemplen (som

samma igår) bör dock innehålla ngt etc.) är, enligt

¹¹För en närmare diskussion av de semantiska aspekterna hänvisas
till kapitel 5.

(31) lasta/1 'placera last på' visst transportmedel

/.../

ä. med konstruktionsväxling 'placera såsom last' på transportmedel

(Efter transfereringsförklaringen följer transfereringsparafra och "parafrastillägg".)

De s.k. syntaktiska transfereringsförklaringarna ger, som synes, en tämligen explicit information (även om typen konstruktionsväxling ej i sig själv är tillräckligt beskrivande). Det kan dock ifrågasättas om denna metod att beskriva transfererade användningar är teoretiskt tillfredsställande. Egentligen bör ett förändrat konstruktionssätt betraktas som en följd av en betydelseförändring och detta betraktelsesätt bör avspeglas i beskrivningen. Vi återkommer till denna fråga i kapitel 5.

Databasens tredje sätt att ge specifik syntaktisk information är med hjälp av exempel. Exempelen i databas (och ordbok) har primärt semantiska funktioner, vilka närmare beskrivs i kapitel 5. Givetvis måste dock varje exempel illustrera en faktisk konstruktionstyp för ifrågavarande lexem och det vore därför möjligt att utnyttja de semantiskt motiverade exemplen för att uppvisa alla, eller de viktigaste, konstruktionssätten. Tänkbart vore även att använda exempel som enbart är syntaktiskt motiverade. Båda dessa principiella möjligheter har utnyttjats i databasen men ej på något systematiskt sätt, och detta av två skäl. Dels skulle för en komplett syntaktisk exemplifiering krävas ett syntaktiskt förarbete lika omfattande som det som skulle krävs för den ursprungligen planerade formella konstruktionsangivelsen, dels skulle systematiskt varierade konstruktionssätt i exemplen i många fall göra ett onaturligt och krystat intryck. Därtill kommer att databasens användning som ordboksmanuskript fordrade att hänsyn togs till utrymmesproblematiken (ett skäl som numera bortfallit). Sammanfattningsvis kan sägas att exemplen i allmänhet torde ge en relativt fyllig bild av lexemens semantik, inklusive deras syntaktiskt manifesterade argument, men endast illustrationer av deras mest typiska konstruktionssätt. De rent syntaktiska exemplen (som kan igenkännas på att de innehåller ngn, ngt etc.) är, enligt

ovanstående resonemang, de svagast motiverade och förekommer också tämligen sparsamt.

Det finns naturligtvis inga "typiska" exempel i databasen, varför vi nöjer oss med en enda illustration:

(32) ärva 'överta genom arv' med avs. på egendom etc.
ärvda möbler; han ärvde lite aktier och obligationer
efter sin mor; syskonen fick gemensamt ärva sommarstugan
äv. underförstått med avs. på pengar
han har det lite bättre ställt nu sedan han fick ärva

Som synes ger kärnbetydelsens exempel primärt information om typiska argument, särskilt Object. Sekundärt visas hur detta Objekt, normalt manifesterat som direkt objekt, även kan uppträda som huvudord till ett particip i adjektivisk användning. Den transfererade betydelsen innebär i princip att Object-argumentet "flyttas in" i definitionen (med en speciell rollinnehavare) och exemplet på detta blir alltså en absolut (s.k. intransitiv) konstruktion.

5. Betydelsebeskrivningen: innehållsliga aspekter

Som nämnts i avsnitt 4.1 beskrevs kärnbetydelser med definitioner och definitionstillägg, transfererade betydelser med transfereringsförklaringar och parafraser. Därtill fanns ett antal kompletterande möjligheter som ej skall upprepas här. Definitionerna kan i sin tur betraktas dels ur det enskilda lexemets perspektiv ("lokalt"), dels ur det totala definitionssystemets perspektiv ("globalt"). Det följer att detta kapitel lämpligen bör uppdelas i fyra avsnitt: Definitionssystemet, Definiering av kärnbetydelser, Transfererade betydelser, Kompletterande beskrivningssätt.

5.1 Ett principiellt problem med varje system av i ord formulerade definitioner (vilket tenderar att oroa filosofer mer än praktiskt arbetande lingvister) är att systemet növärdigtvis blir cirkulärt, i den meningen att åtminstone en delmängd av de definierade orden direkt eller indirekt definierar varandra. Ett besläktat lexikologiskt problem, som alla, som någon gång bläddrat i en ordbok, torde ha stött på, är när en liten grupp semantiskt närbesläktade ord (i värsta fall två ord) konsekvent används för att definiera varandra. Man kan i det förra fallet tala om global, i

det senare om lokal cirkularitet. Givetvis är lokal cirkularitet det mer elakartade problemet (särskilt som den faktiskt i det överväldigande antalet fall kan undvikas) men man kan även önska att bringa systemets globala cirkularitet under viss kontroll. Det ingick därför i projektets ursprungliga målsättning att databasen skulle byggas upp systematiskt, kring en liten, central vokabulär vars enheter ur filosofisk synpunkt kunde betraktas som axiom, ehuru de självklart skulle förses med definitioner (vilka tillsammans alltså fick utgöra ett globalt cirkulärt system). Den centrala vokabulären (kallad definitionsvokabulären och av en storleksordning uppskattad till kanske 4000 lexem) skulle sedan användas för att, direkt eller indirekt, definiera återstoden av databasens lexem, förutom att den, som nämnts, skulle definiera sig själv. Eftersom varje lexem måste vara definierat i ett ändligt antal "steg", räknat från den centrala vokabulären, skulle det bli principiellt möjligt att urskilja vokabulärskikt med olika grad av semantisk "centralitet". Det skulle sedan ligga nära till hands att de mest perifera skikten (i den anförda meningen) finge en mindre ambitiös och detaljerad semantisk beskrivning. (Se också Ralph-Järborg-Allén 1977.)

En systematisk definitionsvokabulär av ovan antytt slag måste skiljas från en pedagogisk definitionsvokabulär, vilken senare typ använts särskilt i många enspråkiga engelska ordböcker.¹² Sålunda brukar orden i pedagogiska definitionsvokabulärer vara valda på grund av sin vanlighet, snarare än sin lexikaliska centralitet eller precision, och alla definitioner brukar formuleras direkt i definitionsvokabulärens ord (i stället för alternativet med stegvis definiering). När kvantitetsuppgifter ges bör man hålla i minnet att de normalt avser lemman eller motsvarande och att antalet lexem i en pedagogisk definitionsvokabulär ofta är mångfaldigt större, vidare att avledningar till definitionsvokabulärens ord tillåts, utan att de räknas som nya enheter. Det är tydligt att en pedagogisk definitionsvokabulär ej direkt skulle kunnat användas eller ens tas som förebild för LDB:s planerade centrala vokabulär, även om någon sådan hade förelegat på

¹² Den kanske mest kända av dessa under senare år är Longmans Dictionary of Contemporary English (LDOCE) 1973.

svenska.¹³ Det fanns således ingen självklar utgångspunkt för upprättandet av den centrala vokabulären (även om enstaka kandidater för denna tämligen lätt kan uppställas). Valet stod i detta läge mellan, å ena sidan, att satsa resurser på ett delprojekt som upprättade en explicit, systematisk definitionsvokabulär a priori och, å andra sidan, att fastlägga en strategi för systematisk definiering som i slutläget skulle skapa en databas innehållande en implicit definitionsvokabulär som sedan skulle kunna extraheras a posteriori. Det senare alternativet valdes, bl.a. för att det praktiska definieringsarbetet då skulle kunna påbörjas direkt (ehuru med viss risk att behöva modifieras i slutskedet). Tyvärr har den implicita definitionsvokabulär som, trots vissa inadvorter, torde existera inom databasen av tidsskäl ej kunnat extraheras ännu (den planerade men ej utförda disambiguering av definitionsorden som nämndes i kapitel 1 var tänkt som ett led i extraheringsprocessen).

Den valda definieringsstrategin byggde på en växelverkan mellan å ena sidan lexikologernas intuition om semantisk centralitet och semantisk släktskap, å andra sidan systematiska kontroller av relevanta definitioner. Att en "lexikalisk intuition" av det antydda slaget faktiskt existerar torde vara ett tämligen riskfritt antagande; man behöver blott tänka på alla de ordböcker som, med viss framgång, försöker definiera sina enheter i "enklare" (ofta = semantiskt mer centrala) ord. I det systematiska definieringsarbetet utnyttjades denna intuition främst för att utvälja tentativa definitionsord (eg. lexem) för definiering av större eller mindre semantiska klasser av lexem. De tentativa definitionslexemen kunde sedan själva tänkas ingå i mer abstrakta klasser som i sin tur systematiskt skulle kunna definieras med ännu mer centrala definitionslexem etc. Under arbetets gång utbytte projektmedarbetarna ständigt förslag om hur olika klasser skulle definieras; ur denna interna diskussion utkristalliserades småningom en uppsättning mer eller mindre definitiva definitionslexem, med vars hjälp åtskilliga viktiga semantiska klasser kunde definieras. Sålunda definieras de centrala sinnesverben med uppfatta/1, medan måttsadjek-

¹³ Vid projektets början förelåg dock en till svenska översatt och anpassad engelsk definitionsvokabulär. Den visade sig emellertid behöva omfattande bearbetningar för att kunna göras relevant för projektet. Se vidare Ralph 1978.

tiv definieras bl.a. med utsträckning. Det bör påpekas att många definitionslexem är tämligen självklara, speciellt huvudorden i definitioner av konkreta substantiv. T.ex. torde verktyg, ört, rovdjur vara okontroversiella definitionslexem för sina respektive klasser.

De viktigaste kontrollerna av definitionssystemet avsåg givetvis definitionernas semantiska huvudord (där sådana förekom). Grundläggande var kontrollen av de hierarkiskt överordnade lexemen: givet en definition D_n med huvudordet H_n undersöktes först definitionen D_{n-1} av H_n , därefter definitionen $D_{(n-2)}$ av $D_{(n-1)}$:s huvudord $H_{(n-1)}$, etc. Om $H_n = H_{(n-1)}$ hade den snävaste lokala cirkeln undvikits, men kontrollen fortsattes lämpligen fram till något H_i som tillhörde en klart abstraktare kategori än H_n , $H_{(n-1)}$ etc.¹⁴ Som ett trivialt exempel kan nämnas

- (33) **cylinderhatt** '... hatt ...'
- hatt/1** '... huvudbonad ...'
- huvudbonad** 'klädesplagg ...'
- klädesplagg** '... föremål ...'

Redan vid huvudbonad torde man ha nått en tillfredsställande abstraktionsnivå i förhållande till utgångspunkten cylinderhatt; det kan säkert konstateras att inga lokala cirklar uppkommer. Någon gång kunde det vara lämpligt att kontrollera även nästa abstraktionsnivå, trots att denna i princip skulle undersökas i samband med kontrollen av definitionen av huvudbonad (eller dess eventuella systerbegrepp). Överbegreppet klädesplagg ingår för övrigt säkerligen i den implicita definitionsvokabulären. Med föremål/1 är man inne i den mest centrala delen av definitionsvokabulären; först här uppträder egentlig cirkularitet, enligt följande:

- (34) **föremål/1** '... företeelse ...'
- företeelse** '... ngt ...'

¹⁴Det förekommer faktiskt ännu snävare lokala cirkulariteter, som t.ex. under 1/2 'i anslutning till in- eller undersidan av' (i SOB; ursprunglig, korrekt version numera återställd i LDB). Denna typ, där ett ord i viss mån får definiera sig självt, är givetvis ovanlig i SOB men ej sällsynt i t.ex. ISO.

med någon/1 '... [person eller] företeelse ...' bestämmning som offensiv (vilket torde ligga på gränsen för att räknas som ett Det alternativa huvudordet person är givetvis aktuellt endast vid utrumformen av någon. Dessutom bildas en cirkel även av sviten person - varelse - företeelse - någon. Som synes kan den globala cirkulariteten hos i stort sett hela systemet av substantiviska definitioner återföras till denna innersta kärna av definitions- vokabulären, vilken är tillräckligt överskådlig för att kunna definieras och kontrolleras i ett sammanhang.

Kontrollen "uppåt" i definitionssystemet, enligt ovan, var det främsta hjälpmedlet för minimering av cirkulariteten. Det bör dock noteras att det ej alltid var tillräckligt att kontrollera huvudorden, som i fallet

(35) anfall/1 'offensiva (fientliga) åtgärder' angrepp 'direkt anfall', som endast kunde ha fyllits på offensiv 1 'som inriktar sig på angrepp eller anfall'

Med ett på gränsen till betydelseomhet abstrakt huvudord i definitionen hos anfall/1 (vilket i sin tur definierar angrepp) måste givetvis även huvudbestämningen i definitionen kontrolleras, men denna, adjektivet offensiv 1, definieras väsentligen med hjälp av anfall/1. Cirkeln i (35) skulle alltså kunna sammanfattas

(35') anfall/1 '(fientliga) åtgärder som inriktar sig på direkt anfall eller anfall',¹⁵

(Denna tämligen elakartade cirkel är numera upplöst i LDB, ehuru den tills vidare kvarstår i SOB.)

Det var ej alltid självklart när kontrollen måste utvidgas på sätt som i (35) och merparten av ännu kvarstående lokala cirklar i LDB torde vara av denna "blandade" typ, där huvudorden definierar sina viktigaste bestämmningar (snarare än de "rena" cirklarna där huvudorden definierar varandra, som i (34)). En varningssignal just i fallet (35) är att ett tämligen centralt ord som anfall definieras

¹⁵Egentligen skall sedan (35') insättas i sig själv, vilket gör definitionen oändlig. Vi avstår här från denna övning.

med hjälp av en rimligtvis betydligt mer perifer bestämning som offensiv (vilket torde ligga på gränsen för att räknas som ett "lärt" ord).

Det kan i detta sammanhang noteras att "lärdare" eller "tekniska" ord undveks i definitionerna (utom givetvis då de gällde ännu "lärdare" ord), och detta av två skäl: dels är dylika ord ej sällan svårbegripliga i sig själva, dels, och framför allt, måste de betraktas som språkpsykologiskt perifera och bör därför ej få centrala roller i definitionssystemet. På grund av denna princip har definitionssystemet någon gång förlorat goda överbegrepp på lämpliga mellannivåer. Typexemplet är artefakt, som skulle kunnat vara överbegrepp till åtskilliga konkreta substantiv betecknande mänskliga produkter. Denna klass fick nu definieras direkt med föremål/1, ofta kompletterat med en bestämning som (special)-tillverkat e.d. Definitionssystemet kom på detta sätt att innehålla vissa "begreppsliga luckor", som endast kunde ha fyllts på bekostnad av dess trovärdighet som modell av människans interna lexikon.

För de flesta ordklasser gällde, som nämnts, att definitionerna saknade huvudord och kontrollerna "uppåt" fick utföras under hänsynstagande till de specifika problemen. För en relativt stor grupp adjektiv gällde sålunda att de definierades väsentligen med hjälp av andra adjektiv, vilka alltså kom att fungera som ett slags kvasi-huvudord. Nu kan emellertid adjektiv knappast i någon större utsträckning inordnas i semantiska hierarkier liknande dem som är naturliga för substantiv och verb. Följden blir att adjektivdefinitioner kan bilda semantiska kedjor som sträcker sig "sidledes" snarare än "uppåt", vilket kan illustreras med

- (36) **trumpen** 'butter och missnöjd'
butter 'som ofta visar sig tvär och ovänlig'
missnöjd 'som känner missnöje'
tvär/3 'butter och föga sällskaplig'
ovänlig 'som visar negativ inställning'
missnöje 'känsla av att ngt borde vara bättre'
sällskaplig 'som gärna umgås med andra människor'

Som synes definieras trumpen, tvär/3 och, i princip, även butter med andra adjektiv. En partiell cirkel bildas av butter och tvär/3. Det torde vara svårt att hävda att någondera av trumpen, butter och tvär/3 egentligen ligger på en hierarkiskt högre nivå än de andra, varför definitionskedjan här får sägas löpa i sidled. Däremot finns ingen återkoppling från ovänlig, som även intuitivt kanske kan uppfattas som mer grundläggande. De övriga "ändpunkterna" i kedjan är missnöje och sällskaplig, vilka ävenledes kanske tillhör en högre nivå än trumpen etc. Det bör observeras att alla cirkelbildningar mellan adjektivdefinitioner effektivt bryts av definitioner i det reguljära relativsatsformatet (utom vid betydelsetomma satser som i fallet butter). Detta är ett viktigt skäl till att relativsatsor bör vara det enda godtagna formatet för adjektivdefinitioner (för ytterligare skäl hänvisas till 5.2). Via "ändpunkterna" återförs trumpen till grundläggande begreppsstrukturer snarare än grundbegrepp: visa negativ inställning, känna att ngt borde vara bättre, ogärna umgås med andra människor; dylika begreppsstrukturer (här något förenklade) kan naturligtvis aldrig vara cirkulära i någon rimlig mening. Kontrollerna av kedjor av adjektivdefinitioner i sidled kan uppenbarligen bli ganska omständliga, särskilt som de ofta grenar sig med en konjunktion (som i fallen trumpen, tvär/3 i (36)).

De snävaste cirklar som normalt accepteras uppkommer vid prepositionsdefinitioner. Dessa, som nästan alltid är i formatet Prep-NP-Prep (jfr. 4.2), har visserligen sin väsentliga semantiska information i NP:n, vilken alltid bör gå att utforma så, att lokala cirklar bryts. Men eftersom varje prepositionsdefinition fordrar två relaterande prepositioner (vilka givetvis ej får ingå i samma prepositionslemma som definiendum) och dessa ofta måste tillhöra kategorin "grammatiska" prepositioner kan cirklar inom denna kategori ej undvikas. Många dylika cirklar står att finna bland prepositioner som av, för, i, med, på, till m.fl. Detta cirkularitetsproblem beror naturligtvis djupast sett på att prepositioner bildar ett eget, välbestämt lexikaliskt subsystem som i princip ligger på en meta-nivå i förhållande till huvudordklasserna. (Jfr även Järborg 1988.)

Förutom den lokala cirkulariteten undersöktes vid kontrollen "uppåt" av definitionsorden även databasens interna semantiska koherens, varmed förstås att definitionsorden, i den användning de

har i en viss, specifik definition, själva skall vara fullgott semantiskt beskrivna inom databasen. Kravet innebar alltså att databasen skulle kunna "beskriva sig själv" (dock med två principiella undantag, dels s.k. självklara sammansättningar och avledningar, dels egennamn; se vidare nedan). Kravet på intern semantisk koherens kan specificeras enligt följande: (i) varje definitionsord skall klart kunna hänföras till ett och endast ett i databasen förefintligt lexem, (ii) definitionslexemen skall i princip användas endast i sina kärnbetydelser, (iii) om, på grund av speciella omständigheter, ett definitionslexem används i en transfererad betydelse måste det röra sig om en väletablerad, välbeskriven och frekvent sådan.

Eftersom kontrollen av intern koherens ställer större krav i fråga om lexikalisk intuition, semantisk analysförmåga och allmän noggrannhet än kontrollen av cirkulariteten är det naturligt att den ej hunnit utföras i tillfredsställande utsträckning och att åtskilliga inadvartenser kvarstår i databasen. De ovan berörda prepositionerna har även här visat sig särskilt problematiska; det är många gånger svårt eller omöjligt att entydigt hänföra en definitionspreposition till ett bestämt lexem. Sålunda har det visat sig att beskrivningen av av 1/1 'med utgångspunkt eller ursprung i' och av 1/6 'i hänseende till' måste göras mer distinkt för att underlätta gränsdragningen, att ett nytt, mer pregnant lexem med ungefärlig betydelse av Experier måste avskiljas från för 1/6 samt att bl.a. på 1 och till 1 (till 2 i SOB) troligen delvis måste analyseras om, åter på grund av problem med många gränsfall.

Något liknande gäller vid åtskilliga grundläggande verb (ofta starka eller eljest oregelbundna) i definitionerna: det går ej alltid att entydigt hänföra dem till bestämda lexem och det händer påfallande ofta att de används i en transfererad betydelse. Det mest flagranta fallet gäller det frekventaste huvudordet i verbdefinitioner, nämligen göra 2 (borde vara göra 1), som praktiskt taget alltid är göra 2/2 i sin 2:a (visserligen relativt välbeskrivna) transfererade betydelse. Vid det i definitioner av klassificerande adjektiv vanliga avse kan man, på grund av beskrivningens vaghet, tveka mellan avse/2 och avse/3 (skall vara /3, medan det kan diskuteras huruvida ej /1 och /2 borde sammanföras).

Ett mera dolt problem uppträder i samband med en ofta förbisedd sammansättningsstyp, nämligen partikelverb. I LDB behandlades partikelverben så långt möjligt som tillfälliga sammansättningar med förutsägbar betydelse, något som fordrade, och har lett till, en mycket noggrann analys och beskrivning av de flesta frekventa verbpartiklar. Endast de partikelverb vars betydelser ej kan betraktas som kombinationer av delarnas betydelser etablerades som egna lemman (dvs. samma princip som gällde för ordinära sammansättningar och avledningar). Nu är emellertid analysen av partikelverbs betydelse en relativt krävande uppgift som ytterligare komplicerar den besvärliga kontrollen av databasens interna koherens eftersom både verblexemen, partikellexemen och deras kombinationer måste undersökas. Icke sällan visar det sig t.ex. att ett synbarligen helt tillfälligt partikelverb innehåller en extra semantisk faktor som ej ingår i delarnas betydelser och denna faktor kan mycket väl vara väsentlig då partikelverbet används i en definition. Partikelverb i definitioner tenderar alltså att skapa problem för databasens interna koherens; om möjligt bör de undvikas och där så ej kan ske bör de analyseras mycket noga, vilket ibland kommer att framtvinga etablerandet av nya partikelverbslemman.¹⁶

Medan partikelverb sålunda blev problematiska för den inre koherensen, på ett från början ej förutsägbart sätt, innebar bruket av övriga "självklara" sammansättningar och avledningar, liksom av egennamn, ett medvetet undantag från principen att databasen skulle kunna beskriva sig själv. Vad egennamnen beträffar var undantaget givetvis nödvändigt; som konstaterades i kapitel 3 måste åtskilliga lexem analyseras med hjälp av encyklopediska faktorer och en stor del av dessa motsvaras av egennamn och egennamnsliknande beteckningar (såsom t.ex. beteckningar på tidsavsnitt). En lexikalisk databas kan, av både teoretiska och praktiska skäl, icke förteckna egennamn eller andra encyklopediska ord. Det intressanta är alltså ej att självbeskrivningskravet ej går att uppfylla utan hur och var definitionssystemet måste kompletteras med rent encyklopediska enheter. Detta är ett av de

¹⁶En grundlig semantisk analys, delvis byggande på material från LDB, står att finna i Norén (under arbete), vars resultat tycks stödja de här framförda rekommendationerna.

delproblem som kommer att undersökas i ovan nämnda fortsättningsprojekt.

Förutsägbara sammansättningar och avledningar som ej ingick i eller semantiskt behandlades i databasen fick användas i definitionerna av två skäl: dels (och i praktiken viktigast) förenklades därigenom formulerandet av definitionerna, dels kunde med hjälp av sammansättningar ofta lämpliga hierarkiska mellannivåer introduceras. - I det praktiska arbetet kom även sammansättningar någon gång att användas i disambiguerande syfte, som i

(37) hårknut 'huvudhår som virats ihop i en knut'

Här specificerar huvudhår en pregnant transfererad betydelse av hår men är självt ej definierat i databasen. - Det står klart att om de i definitionerna använda sammansättningarna och avledningarna verkligen har förutsägbara betydelser kan databasen, på ett mer abstrakt plan, fortfarande anses vara internt koherent (eller självbeskrivande). Frågan är nu givetvis vad som skall förstås med "förutsägbar betydelse". Till en början måste den reservationen göras att någon mekanisk förutsägbarhet ej kunde krävas, eller, med andra ord, att en mänsklig språkbrukare förutsattes. Givet detta, fordrades av en förutsägbar avledning dels att grundordets relevanta lexem var definierat i databasen, dels att avledningstypen var semantiskt genomskinlig och icke-ambiguös. Av en förutsägbar sammansättning fordrades dels att både efterledens och förledens relevanta lexem var definierade i databasen, dels att den semantiska relationen mellan leden var icke-ambiguöst tolkbar.

Tydliggen bör det vara möjligt att kontrollera de använda avledningarna, även om vissa avledningstyper kan vålla problem; så t.ex. betecknar ofta en verbavledning på -ning förutom verbhandlingen i sig också någon konkret företeelse (resultat etc.) som har samband med denna. Det är svårare att hålla de använda sammansättningarna under kontroll, givetvis på grund av att så många olika typer av relationer mellan leden kan förekomma. I början av projekttiden gjordes en liten undersökning av dylika relations typer på ett material bestående av de sammansättningarna som hade använts spontant i parafraserna i LEX-filen och som preliminärt kunde bedömas som ej kvalificerade till att själva ingå i databasen. Undersökningen visade att inom detta material (där ling-

visterna alltså vinnlagt sig om att söka använda förutsägbara sammansättningar) fem typer av relationer klart dominerade, nämligen (med djupkasusetiketter), Material (silversked), Purpose (dessertsked), Location (skrivbordslåda), Agent (polisbrutalitet) och Object (barnmisshandel).¹⁷ Som en första approximation kunde kanske dessa fem typer ha ansetts "normala" och därmed tolkbara; däremot är de, som redan exemplen antyder, ibland ambigösa (jfr skrivbordslampa - Purpose/Location, polismisshandel - Agent/-Object). I praktiken kom undersökningsresultatet snarast att tas som ett slags bekräftelse på att sammansättningar, bildade med omdöme, kunde få användas i definitionerna, oavsett om de var odefinierade i databasen.

Någon ytterligare systematisk undersökning av sammansättningar har ej företagits. I stort sett kan nog hävdas att sammansättningarna i definitionerna är bildade med omdöme och i många fall torde även sammansättningslederna, i synnerhet slutlederna, ha kontrollerats. Däremot kan intet säkert sägas om vilka sammansättningstyper som förekommer, långt mindre huruvida de är "förutsägbara" i strikt mening (enligt ovannämnda kriterier). Även bortsett från de tidigare anförda inadvertenserna kan databasen alltså ej vara internt semantiskt koherent, ens i den utvidgade, abstrakta bemärkelsen, på grund av att användningen av sammansättningar (inklusive partikelverb) i definitionerna ej har kunnat regleras på ett tillräckligt stringent sätt. I fortsättningen bör därför odefinierade sammansättningar (liksom partikelverb och troligen även avledning) ej tillåtas i definitionerna: dessa bör antingen omformuleras eller måste de oundgängliga sammansättningarna registreras och definieras i databasen, hur "förutsägbara" de än må verka. För att bedömas som oundgänglig bör en hittills odefinierad sammansättning fylla en nödvändig hierarkisk mellannivå i definitionssystemet; sammansättningar som endast införts som ett formuleringshjälpmedel bör däremot utrensas. Disambiguerande sammansättningar behövs ej i databasen men möjligen i ett ordboksmanuskript (eventuellt skulle dylika sammansättningar alltså kunna förekomma i SOB-filen, varvid förlederna lämpligen placeras inom parentes).

¹⁷ De anförda exemplen är valda för tydlighets skull, ej hämtade ur undersökningen.

I det praktiska kontrollarbetet lades den ojämförligt största insatsen på kontrollerna "uppåt" i definitionssystemet, dels på grund av att de avsåg de grundläggande systemegenskaperna icke-cirkularitet och koherens, dels också på att sådana kontroller i viss mening är lätta att utföra, eftersom man naturligtvis förfogar över de primära undersökningsobjekten (= definitionslexemen). Vid kontroller i de övriga hierarkiska riktningarna, nämligen "sidledes" (med avseende på s.k. systerbegrepp) respektive "nedåt" (med avseende på s.k. dotterbegrepp eller underordnade begrepp) kan undersökningsobjekten endast vara intuitivt och partiellt kända, tills hela definitionssystemet dels har fullbordats i en första version, dels blivit disambiguerat, dels underkastats en åtminstone rudimentär strukturanalys. (För att man skall kunna återfinna alla "systrar" eller "döttrar" till ett givet lexem fordras uppenbarligen att alla huvudord i definitionerna har blivit taggade och disambiguerade.) Det är således föga märkligt att kontrollerna sidledes och nedåt kom att utföras i mycket begränsad utsträckning under projekttiden. Efter slutförandet av ett smärre uppföljningsprojekt förelåg bättre förutsättningar för dylika kontroller i form av en preliminär strukturmärkning av c:a 30 % av databasens definitioner.¹⁸ Flera av de nedan anförda exemplen emanerar från detta strukturmärkta material.

Kontrollerna "sidledes" och "nedåt" kan användas för att undersöka definitionssystemets allmänna konsekvens och särskilt dess strukturella konsekvens. Med det senare menas i princip att definitionssystemet skall nå ett gemensamt överbegrepp till flera underbegrepp (på samma hierarkiska nivå) i lika många hierarkiska "steg", oavsett vilket underbegrepp man utgår från. Ett enkelt exempel är

- (38) (a) **urmakare** 'person som yrkesmässigt tillverkar och lagar ur'
 (b1) **gördelmakare** 'hantverkare som tillverkar bälten'
 (b2) **hantverkare** 'person som yrkesmässigt utövar ett hantverk'

¹⁸ Projektets titel var Semantisk strukturering av Lexikalisk databas; det beskrivs kortfattat i Allén-Järborg 1988. Projektet är att betrakta som en (visserligen extensiv) förundersökning till det större projekt som skisseras i Järborg 1988.

(33') cylinderhatt < hatt < huvudbonad < beklädnad < föremål
Uppenbarligen ligger urmakare och gördelmakare på samma hierar-
kiska nivå och lika uppenbart är att person måste vara gemensamt
överbegrepp. Som synes av (38) (a) definieras emellertid urmakare
direkt med person (ett hierarkiskt steg), medan gördelmakare de-
finieras indirekt, via (38) (b1) och (b2) (två hierarkiska steg).
På denna punkt är databasen alltså ej strukturellt konsekvent
definierad.

Ett fall av typen (38) är relativt enkelt att kontrollera "sidle-
des" eftersom det ej vållar några svårigheter att uppställa sys-
terbegrepp inom vissa välstrukturerade lexikala subsystem. Ett
likartat fall togs upp i exempel (33), vars definitionskedja kan
sammanfattas som

(33') cylinderhatt < hatt < huvudbonad < klädesplagg < föremål

Ett självklart "system" till huvudbonader är givetvis sko-
don, men här ser kedjan annorlunda ut:

(39) pumps < sko(r) < fotbeklädnad

Fotbeklädnad finns ej i databasen men grundordet är definierat:

(40) beklädnad/1 'klädesplagg' inklusive skodon och huvudbo-
nader
<sammanfattande beteckning>

Som framgår är beklädnad beskrivet som om det vore ett lämpligt
överbegrepp (däremot är definitionen troligen felaktig, eftersom
beklädnad tycks vara mer abstrakt än klädesplagg). Med stöd i (40)
kan man modifiera (39) till antal sammansättningar med prägnant
(och i systematisk-biologisk mening svärförutsägbar) betydelse.

(39') pumps < sko(r) < fotbeklädnad < beklädnad < klädes-
plagg < föremål

Vid jämförelse mellan (33') och (39') finner man åter skillnader i
antalet hierarkiska steg, dvs strukturell inkonsekvens. I detta
fall kompliceras dessutom bilden av den felaktigt angivna relatio-
nen mellan beklädnad och klädesplagg. Mer konsekvent vore

(33'') cylinderhatt < hatt < huvudbonad < beklädnad < föremål

(39'') pumps < sko(r) < fotbeklädnad < beklädnad < föremål

medan klädesplagg reserverades för sitt naturliga användningsområde (= annat än huvudbonader och skor). Huruvida sedan klädesplagg skulle definieras med beklädnad som överbegrepp eller definieras direkt med föremål finge avgöras genom särskild utredning. Det har f.ö. visat sig att ordinära klädesplagg har definierats ömsom med klädesplagg, ömsom med plagg som huvudord, vidare att definitionerna av dessa ord är skiljaktiga. Detta är ett (av många) exempel på allmän inkonsekvens i definitionssystemet som kan upptäckas vid kontroll sidledes.

Ehuru konsekvens i definitionssystemet givetvis är önskvärd är det (som tidigare påpekats) viktigt att kunna avväga det "globala" kravet på konsekvens mot det grundläggande lexikologiska kravet på adekvat semantisk beskrivning av enskilda enheter. Betrakta

- (40) (a) katt < rovdjur < däggdjur
(b) häst < hovdjur < (dägg)djur
(c) myrslok < däggdjur
(d) bältdjur < däggdjur

Uppenbarligen visar exemplet på en strukturell inkonsekvens, i det att (a) och (b) definieras med en mellannivå som saknas i (c) och (d). Emellertid kan denna inkonsekvens lätt försvaras med åtminstone två goda argument. För det första är olika djurarter olika förtrogna, vilket har till följd att motsvarande språkliga beteckningar kan ha mycket olika vikt inom språksystemet. Ett ord som katt i (a) ingår i ett stort antal idiomatiska uttryck och talesätt, liksom i ett stort antal sammansättningar med pregnant (och i systematisk-biologisk mening svårförutsägbar) betydelse. Det torde vara svårt att finna något enda uttryck av liknande slag med myrslok. Givetvis bör det språkligt viktigare artnamnet erhålla en fylligare betydelsebeskrivning. För det andra existerar i fallen (a) och (b) goda mellanbegrepp, som här dessutom råkar vara både språkligt-funktionellt och systematiskt adekvata. I synnerhet mellanbegreppet rovdjur i (a) är ytterst relevant för beskrivningen av hur ordet katt fungerar i språket. I motsats härtill existerar i (c) och (d) inga lämpliga språkligt-funktionella mellanbe-

grepp och det systematiska mellanbegreppet (tandfattiga däggdjur) är givetvis helt oanvändbart, även bortsett ifrån att det rör sig om en tvåordsbeteckning.¹⁹ Konklusionen är alltså att man måste skilja mellan systematisk-klassifikatorisk konsekvens (som bygger på utomspråkliga kriterier) och språkligt-funktionell konsekvens (som bygger på inomspråkliga kriterier); endast den senare kan och skall upprätthållas vid uppbyggnad eller korrektion av definitionssystemet. (Det kan tilläggas att den här anförda distinktionen, ehuru tämligen självklar i teorin, har visat sig mycket svår att i praktiken fatta och applicera, något som icke minst har gällt för kritikerna av Svensk ordbok.)

(b) Isgräs > ballongboll, bardy, basketboll, fotboll.
Några egentliga kontroller "nedåt" utfördes aldrig under själva projektiden men har, som sagts, i viss utsträckning kunnat utföras på det strukturmärkta materialet. Ur detta material kunde lätt erhållas de lexem som var direkta underbegrepp till ett givet ord (vilket alltså var huvudord i dessas definitioner). Om definitionssystemet är strukturellt konsekvent skall givetvis samtliga underbegrepp till ett givet ord ligga på samma hierarkiska nivå. Sålunda skulle fall av typen (38) kunna kontrolleras enligt följande:

(38') (a) person > urmakare

(b) person > hantverkare

(etc.)

och det är uppenbart att underbegreppen ligger på olika nivå. I själva verket visade det sig att person (som är det överlägset frekventaste huvudordet i det strukturmärkta materialet och säkerligen även totalt i LDB) har fått definiera en synnerligen heterogen samling underbegrepp, där man i många fler fall än hantverkare-typen kunde ifrågasätta om icke ett (eller rentav två) mellanbegrepp borde användas.

¹⁹Troligen klassas djurnamn sådana som bältdjur och myrslok spontant som "konstiga, exotiska djur" e.d. av den ordinäre språkbrukaren, oavsett om han kan återkalla den systematiska klassifikationen. En dylik naiv, men på många sätt adekvat, beskrivning kan tyvärr ej komma i fråga för användning i definitionstext.

Kontrollen nedåt kan också användas för att utvärdera sådana definitionsord (i synnerhet sammansättningar) som ej själva är be- lagda eller definierade i databasen. Om ett dylikt ord definierar en enhetlig och naturlig grupp underbegrepp, av ej alltför litet omfång, finns det skäl att betrakta det som ett strukturellt nöd- vändigt mellanbegrepp och införa det som egen enhet i databasen. I det strukturmärkta materialet återfinns bl.a.

(41) (a) halvädalsten > granat, heliotrop, jade, jaspis, kamé, kattöga

(b) lagspel > ballongboll, bandy, basketboll, fotboll, handboll, ishockey
(lexemnumren har här utelämnats)

Som synes definieras i (41) mycket enhetliga grupper av underbe- grepp, varför både halvädalsten och lagspel bör införas i data- basen. (Dessutom borde halvädalsten ha införts redan för sin egen skull, eftersom ordet har karaktär av term inom mineralogin.)

I det föregående har upprepade gånger frågan om definiering i ett större eller mindre antal steg berörts, dvs om ett visst lexem skall definieras direkt med ett tämligen abstrakt ord som huvudord (= ett steg), eller om det skall definieras indirekt, med hjälp av en eller flera hierarkiska mellannivåer (= flera steg). Som provisorisk beteckning på fenomenet införs termen definitionens hierarkiska räckvidd, varvid alltså stor räckvidd = definiering med ett hierarkiskt högt överordnat huvudord. Det torde ha fram- gått att LDB:s definitionsstrategi i princip varit inriktad på liten hierarkisk räckvidd (och många hierarkiska mellannivåer). Från definitionssystemets synpunkt blir definiering med liten hierarkisk räckvidd lika med stegvis uppbyggnad av systemet, vil- ken term även fortsättningsvis kommer att användas där så är lämp- ligt. Detta är motsatt strategi mot den som brukar väljas i ord- böcker uppbyggda kring en s.k. pedagogisk definitionsvokabulär, där målsättningen är att så direkt som möjligt nå tillbaka till definitionsvokabulärens enheter (vilka betraktas som bekanta för läsaren). För definitioner med stor hierarkisk räckvidd talar alltså främst pedagogiska skäl. För definitioner med liten hier- arkisk räckvidd kan goda teoretiska argument åberopas, avseende

då det globala definitionssystemet som den enskilda definitionen; därtill kommer även några praktiska fördelar.

På det globala planet innebär en liten hierarkisk räckvidd att mellanbegreppen får sin korrekta plats i definitionssystemet. Vi hänvisar här än en gång till exempel (38). Om alla hantverkarbeteckningar som urmakare, gördelmakare etc. definierades direkt med person (= den större hierarkiska räckvidden) skulle det erhållna definitionssystemet ingenstades innehålla informationen att hantverkare är det närmaste överbegreppet; i bästa fall skulle denna typ av information (vital för allt semantiskt nätverkstänkande) kunna deduceras ur underbegreppens och mellanbegreppets definitioner (om dessa är tillräckligt explicita). Om däremot hantverkarbeteckningarna definieras med hantverkare (= den minsta hierarkiska räckvidden) får detta ord automatiskt sin riktiga position, som överbegrepp till urmakare etc. och som underbegrepp till person.

För den enskilda definitionen innebär användande av närmast högre överbegrepp (= liten hierarkisk räckvidd) ett enkelt sätt att förmedla information om direkt begreppslig samhörighet. Betrakta (42) A-språk 'första främmande språk i grundskolan'

Det är uppenbart att de adjektiviska bestämningarna i (42) ej har samma status (de kan t.ex. ej kastas om med bibehållen betydelse). I själva verket utgör frasen främmande språk en tvåordsbeteckning på ett enhetligt begrepp (vilken beteckning dock knappast är tillräckligt lexikaliserad för att kunna tilldelas rangen av "tekniskt" idiom, långt mindre av tvåordslemma). Eftersom det tyvärr saknas ett ettordslemma synonymt med främmande språk finns knappast någon praktisk möjlighet (förutom långsökta arrangemang med parenteser e.d.) att markera denna begreppsliga samhörighet i en ordinär definitionstext.²⁰ I fall av typen (38) kan däremot samhörigheten i uttrycket person som yrkesmässigt... enkelt förmedlas genom användande av det närmaste överbegreppet hantverkare. Men exemplet visar också på vad man skulle kunna kalla en "prak-

²⁰Det finns dock tekniska möjligheter för detta i det fortsättningsprojekt som beskrivs i Järborg 1988. Ett förslag till markeringsystem ges i Järborg (under arbete).

tivisk-teoretisk" fördel med den s.k. stegvisa definieringen, nämligen möjligheten att enkelt täcka någon eljest lätt förbisedd semantisk faktor. Om man "sätter in" (38) (b2) i (38) (b1) erhålls (ungefär):

(38) (a) **urmakare** 'person som yrkesmässigt tillverkar och lagar ur'

(b') **gördelmakare** 'person som yrkesmässigt utövar hantverket att tillverka bälten'

Av (b') fås tilläggsinformationen att det att tillverka bälten är ett hantverk (vilket det att tillverka och laga ur tydligen icke vore). Om man därtill betraktar

(43) **hantverk** 'arbetsform vid produktion där arbetet utförs i liten skala med tekniskt enkla hjälpmedel och kräver god yrkesskicklighet'

ser man att mellanbegreppet hantverk innehåller tre väsentliga semantiska faktorer (småskalighet, låg teknisk nivå, god yrkesskicklighet) som alltså ej täcks av den "direkta" definitionen (38) (a). Risken att missa dylika faktorer i beskrivningen torde vara ett i praktiken större problem än svårigheten att markera begreppslik samhörighet.

På det lexikografiska planet balanseras den eventuella pedagogiska nackdelen med liten hierarkisk räckvidd, nämligen att brukaren av ordboken ibland måste slå upp flera mellanbegrepp innan han nått fram till en definition uteslutande bestående av förtrogna ord, av fördelen att mellanbegreppen presenteras på sin logiska plats i systemet. Därigenom torde många mellanbegrepp snabbt komma att läras in, varefter de troligen kommer att förenkla ett mer extensivt bruk av ordboken. När det gäller konkreta ordböcker kan man heller ej bortse från den högst påtagliga fördelen att stegvis definiering helt enkelt är avsevärt mindre utrymmeskrävande (jämför här (38) (b1) med (38) (b') och gör tankeexperimentet att (43) dessutom skulle sättas in i (38) (b')). Sammanfattningsvis synes strategin med liten hierarkisk räckvidd i definitionerna och stegvis uppbyggnad av definitionssystemet vara avgjort överlägsen. Det verkar också vara i detta avseende som LDB:s allmänna definitions-

strategi tillämpats mest konsekvent; fallen med onödigt stor hierarkisk räckvidd (sådana som urmakare och flera andra direkta underbegrepp till person) tycks vara relativt sett glesare än t.ex. fallen med lokal cirkularitet eller dålig intern konsistens.

5.2 Det övergripande formella villkoret på definitionen var att den skulle vara syntaktiskt (och morfologiskt) utbytbar mot definiendum. På det innehållsliga planet uppställdes två övergripande krav: dels att definitionen skulle vara semantiskt utbytbar mot definiendum i alla centrala användningar (se 3.1), dels att definitionen skulle vara analytisk, dvs. formulerad så, att varje semantisk faktor som urskildes vid analysen beskrevs med ett eget ord eller en egen fras i definitionstexten (se 3.2). Av dessa krav är definitionens analyticitet i praktiken det viktigaste eftersom den semantiska utbytbarheten endast låter sig testas med hjälp av en tillräckligt utförlig definition. Dessutom skulle i princip de semantiska faktorerna redan ha testats mot lexemets typanvändning i samband med betydelseanalysen, varav följer att en definition som är analytisk därigenom också bör vara semantiskt utbytbar. En analytisk definition (i den här använda meningen) har två klara fördelar framför andra alternativ: dels ger den fylligast möjliga semantiska beskrivning av kärnbetydelsen, dels erbjuder den utgångspunkter för beskrivning av eventuella transfererade betydelser. I en ordbok har analytiska definitioner den pedagogiska fördelen att så långt möjligt göra den lexikala analysen genomskinlig för läsaren. Den främsta nackdelen är att det kan vara tidsödande (och ibland näst intill omöjligt) att formulera definitionen så, att den uppfyller kravet på analyticitet och ändå förblir läsbar.

Det kan synas som om det förelåge en motsättning mellan kravet på analyticitet och den i föregående avsnitt presenterade principen om liten hierarkisk räckvidd i definitionerna. Genom den sistnämnda principen kommer många definitioner att byggas upp kring s.k. mellanbegrepp, vilka i sig kan innehålla flera, separat utskiljbara, semantiska faktorer (jämför diskussionen kring exemplen (38) och (43) i avsnitt 5.1). Emellertid är de faktorer som innehålls i ett mellanbegrepp ej unika eller särskiljande för ifrågasvarande definiendum utan för en hel klass av med definiendum besläktade begrepp. Det är därför inget egentligt brott mot kravet på analyticitet att dessa icke-specifika semantiska faktorer en-

endast är indirekt (via mellanbegreppet) representerade i definitionen. Svårigheterna att formulera en heltäckande definition insågs på ett tidigt stadium och ledde till införandet av en kompletterande beskrivningsmöjlighet för kärnbetydelsen, det s.k. definitionstillägget. Detta användes i huvudsak för mer eller mindre metaspråklig beskrivning av de lägre rankade semantiska faktorerna, ned till vissa typer av konnotationer. Därtill användes definitionstillägget i vissa fall för beskrivning av (ej nödvändigtvis lågrankade) faktorer som av syntaktiska skäl ej kunde beskrivas i definitionstexten (se om detta avsnitt 4.3) samt ibland även för uppgifter om stilart, typiskt ämnesområde o.d.; denna typ av uppgifter hör eljest närmast hemma i användningskommentaren (se 5.4). En viktig, syntaktiskt betingad, användning av definitionstilläggen som är lätt att förbise är den att ge semantisk information om huvudorden till adjektivdefinitioner. För denna funktion kan definitionstillägg knappast undvaras. Det är naturligt att behandla definitionstilläggets huvudanvändningar i detta avsnitt, eftersom det fungerar just som en kompletterande semantisk beskrivning.

För mycket lågrankade semantiska faktorer, särskilt sådana av konnotationstyp eller sådan som avser mycket specifik encyklopedisk information, fanns en ytterligare beskrivningsmöjlighet, nämligen inbäddning i exempel (se vidare avsnitt 5.4). Denna typ av semantisk beskrivning är egentligen teoretiskt otillfredsställande, från databasens synpunkt, även om den torde kunna fungera i en ordbok. Det är emellertid komplicerat och tidskrävande att utarbeta en mer systematisk och stringent beskrivningsmetod för de lågrankade faktorerna. I avvaktan på att en sådan metod kan utvecklas har åtminstone dessa marginella faktorer blivit registrerade i databasen, visserligen på ett indirekt sätt som kräver viss uppmärksamhet för att kunna utnyttjas. På längre sikt bör givetvis all semantisk information, oavsett rankning, göras helt explicit och placeras i en utbyggd definition eller i ett förbättrat, mer stringent, definitionstillägg. - Idealet torde egentligen vara att klara sig även utan det vanligen metaspråkliga definitionstillägget, men detta ideal är troligen ej uppfyllbart i praktiken; därtill bleve en så långt utbyggd definition ofta helt

oläslig.²¹ I det planerade fortsättningsprojektet kommer dessutom möjligheter att finnas att integrera de objektspråkliga och de metaspråkliga semantiska beskrivningarna utan att hänsyn behöver tas till läsbarhetsaspekter (se Järborg 1988).

Genom användandet av definitionstillägg och (i förekommande fall) indirekt semantisk beskrivning i exempel kunde definitionen i princip koncentreras på de viktigaste semantiska faktorerna hos kärnbetydelsen. Ett generellt undantag måste därvid göras för ord eller delar därav som placerats inom parentes. Det finns fyra huvudtyper av parenteser i databasens definitioner: (i) syntaktiska parenteser (eller argumentparenteser; se 4.2), (ii) semantiska parenteser, (iii) portmanteau - parenteser, (iv) illustrativa parenteser. Av dessa var de två första typerna förutsedda från början, medan de två senare kom att utvecklas under arbetets gång. De olika typerna har helt skilda semantiska funktioner, vilka närmare skall presenteras nedan; tyvärr kan dessa funktioner ej bestämmas med hjälp av några formella kriterier. I stora drag torde kunna hävdas att parenteser överanvänts i definitionerna och att databasen på denna punkt, liksom på andra, skulle behöva stramas upp.

De syntaktiska parenteserna har redan exemplifierats tämligen utfylligt i 4.2. Det är viktigt att notera att information som ges inom syntaktiska parenteser normalt är semantiskt högrankad. En syntaktisk parentes markerar redan genom sin existens att ifrågasvarande begrepp har en relationell karaktär, något som uppenbarligen måste anses vara semantiskt centralt. Det är många begrepp som måste analyseras såsom innehållande relationella semantiska faktorer, särskilt givetvis verbala begrepp, men de som markeras med och beskrivs inom syntaktiska parenteser och alltså måste manifesteras explicit i den syntaktiska kontexten har en särställning, i det att de kan betraktas som lexikaliskt topikaliserade.²² I gruppen av s.k. kommersiella verb (köpa, sälja, betala

²¹ Den som tvivlar på detta ombedes undersöka de längre definitionerna i SAOB.

²² I Norén (under arbete) görs en distinktion mellan referentiella och relationella betydelsekomponenter; den här använda beteckningen relationell semantisk faktor är inspirerad härav. Termen topikalisering skall här förstås i ungefär den mening den har i Järborg 1984, där en distinktion görs mellan topikalisering och fokusering.

o.d.) föreligger samma uppsättning av "primitiva förlopp" (ngn överlämnar ngt, etc.) och samma uppsättning aktanter (Agent, Co-Agent, Source, Goal, Object, Co-Object), vilket lett somliga forskare till att hävda att verben står för "samma" begrepp. Uppenbarligen skiljer sig dock verben begreppslikt genom att olika aktanter topikaliserar, och, parallellt härmed, genom att de "primitiva förloppen" organiseras i annan ordning. Så t.ex. topikaliserar (primärt resp. sekundärt) vid köpa Agent-Goal och Object, vid sälja Agent-Source och Object, vid betala Agent-Goal (?) och Co-Object, andra skillnader att förtiga. (Observera dock att medan Co-Object alltid är ett värde, lämpligen i pengar, är Object normalt en konkret vara men ej så vid betala! De intressanta förhållandena vid motsvarande passiva verb kan vi här ej gå in på.) - Vid ett och samma lemma motiverar topikaliseringskillnader normalt upprättandet av en transfererad betydelse, snarare än ett nytt lexem (se här 4.3, exempel (31) lasta). - Termen semantisk fokusering har hittills avsiktligt lämnats vag men av ovanstående framgår att den avser relativ rankning mellan de högrankade faktorer som i det givna sammanhanget kan betraktas som jämförbara (som alltså verbens aktanter i diskussionen ovan). Som framgår av avsnitt 4.2 markerades endast en mindre del av de explicit syntaktiskt manifesterade faktorerna med parenteser och även där så skett innehåller parentesen ofta endast en platshållare (ngn, ngt). Övrig information om relationella faktorer (där sådan är erforderlig) får sökas i definitionstillägget, där den ofta är speciellt kodad (se 4.3). Härav följer att även definitionstillägget kan innehålla högrankad semantisk information, om det rör sig om faktorer som måste manifesteras syntaktiskt.

Syntaktiska parenteser är normalt lätt igenkännliga för en mänsklig läsare. I tveksamma fall kan man testa genom att sätta in definitionen i en kontext (lämpligen ett av databasens exempel), varvid omedelbart bör framgå att parentesens innehåll avser ett argument eller motsvarande till definiendum, ej definiendum självt.

torde gälla att semantiska parenteser föredragits för optionella (snarare än påfallande lågrankade) faktorer som naturligen låter sig beskrivas i objektspråk och vars beskrivningar lätt kan integreras i den övriga definitionstexten. Ett enkelt exempel är

(44) harakiri '(högtidligt) självmord genom uppristning av magen'

där den näst intill obligatoriska bestämningen högtidligt, som uppenbarligen ej kan anses vara lågrankad, naturligen kan integreras i definitionstexten (medan det vore långsökt att placera den i definitionstillägget). Det enda problemet med denna typ av användning av semantiska parenteser är att det ej finns något sätt att ange graden av optionalitet hos ifrågavarande faktor. Jämför

(45) ängest 'starkt negativ känsla av att vara utsatt för press eller (obestämda) hot' ofta långvarig och ingripande i hela livsföringen

Här anges tydligt i definitionstillägget att de optionella (och ej allra högst rankade) faktorerna långvarig och ingripande i livsföringen uppträder ofta (en av flera standardiserade grader på en relativ skala; se nedan). Däremot finns ingen uppgift om hur karakteristiskt för lexemet det är att hoten är obestämda.

Genom att den semantiska parentesens sålunda kännetecknas av en viss vaghet, både vad gäller rang och optionalitetsgrad hos den information den innesluter, låg det i praktiken nära till hands för lexikologen att använda den som en allmän "reservationsmarkör" i de fall där han kände viss osäkerhet över giltigheten av sin semantiska analys. En del av de semantiska parenteserna i databasen torde kunna karakteriseras som dylika reservationsmarkörer (eller "ängslighetsparenteser"), som egentligen bör strykas. Trots detta föreligger ett sådant fall vid

(46) harm '(moraliskt betingad) vrede' vanl. över ngt som uppfattas som ett övergrepp, en oförrätt e.d.

Om parentesens innehåll borträknas, tycks definitionen påstå att harm = vrede. Eftersom synonymi ej föreligger uppstår frågan hur parentesens egentligen bör utläsas. Skall den t.ex. tolkas som

(46') harm 'vrede som mest/vanligen/ofta är moraliskt betingad'

Parafraaser av typ (46') är uppenbarligen nonsens och en snabb kontroll av de typiska användningarna av harm ger vid handen att något slags "moralisk" semantisk faktor behövs i beskrivningen (vilket även framgår av definitionstillägget i (46)). Slutsatsen måste bli att parenteserna i (46) förmodligen är överflödiga. (Det kan tilläggas att åtskillig semantisk information synes ha placerats i definitionstillägget i stället för i definitionen av liknande skäl. Även i användningen som reservationsmarkör är alltså definitionstillägg och semantisk parentes ofta utbytbara.)

Från dylika reservationsmarkörer är steget kort till en annan typ av överanvändning av semantiska parenteser, vilken kom i bruk då det ordbokmanus som genererats ur databasen undergick (projektinterna eller projektexterna) kontroller. En av dessa kontroller utgjordes av insättning av definitionen i de anförda exemplen, vilken kontroll alltså i viss mån testade definitionens semantiska (och givetvis även syntaktiska) utbytbarhet, låt vara på ett ytligt sätt och på ett material som i princip endast var ett urval av det som underlåg den ursprungliga semantiska analysen. Vid dylika insättningstest konstaterades ibland att kombinationen av definition och språklig omgivning verkade tautolog, som i följande (fingerade) exempel:

- (47) (a) sjöhävning svag sjögång
(b) de kände en svag sjöhävning

Som framgår ger en insättning av (a) i (b) en dubbel förekomst av bestämningen svag, vilket onekligen ser ut som en tautologi. Emellertid måste konstateras dels att de två svag har olika funktioner som gör att de ej är jämförbara (det ena är en beskrivning av en väsentlig semantisk faktor, det andra är en fritt tillagd bestämning), dels att konstruktioner där en bestämning eller motsvarande framhäver eller understryker en redan förefintlig semantisk faktor hos sitt huvudord är mycket vanliga i naturligt språk (vad än logikerna må önska). Egentligen borde därför fall av typ (47) (som rätteligen skulle betecknas kvasitautologi) ej föranleda någon åtgärd. Eftersom dylika fall dock bedömdes "se illa ut" i en ordbok, ändrades antingen definition, exempel eller båda. Den bästa, men mest arbetskrävande, ändringen var en omarbetning av definitionen

så, att den kritiska faktorn beskrevs på ett helt annat sätt som icke rimligtvis kunde komma i konflikt med definiendums naturliga språkliga kontexter. Ändring, inklusive strykning, av exempel in-
nebar endast att problemet gjordes mindre synligt. En enkel och
föga förpliktande ändring av definitionen var att sätta den "tau-
tologa" bestämningen inom parentes; alltså

(47) (a') **sjöhävning** '(svag) sjögång'

Med denna parentes (som kanske kan anses vara en semantisk-syn-
taktisk parentes) blir insättning i (b) möjlig utan att tautologi
uppkommer. Denna praktiska fördel upphävs dock helt av det fel-
aktiga budskap som parentesens förmedlar, nämligen att bestämningen
vore optionell och/eller lågrankad. Ej heller denna typ av paren-
teser har något berättigande i en lexikalisk databas.

Semantiska parenteser har även använts för att avgränsa delar av
ord, som i

(48) **abborrfisk** 'typ av (sötvattens)fisk besläktad med abborren'

Här blir den semantiska parentesens snarast mer svårtolkad än när
den innesluter hela ord. Förmodligen skulle (48) kunna tolkas
ungefär som:

(48') **abborrfisk** 'typ av fisk eller vanligen sötvattenfisk be-
släktad med abborren'

Semantisk parentes kring delar av ord är ej alltid lätta att
skilja från nästa grupp, de s.k. portmanteau-parenteserna.

Som framgått av exemplen och diskussionen kan semantiska parente-
ser normalt igenkännas på att de i definitionstexten kan ersättas
med en gradangivelse (t.ex. vanligen), utan väsentlig förändring
av definitionens innehåll. Alternativt kan parentesens innehåll
tänkas förflyttad till definitionstillägget, ävenledes med en
gradangivelse. Här kan vara platsen att ange den relativa skala
(s.k. ordinalskala) som i databasen använts för gradangivelse: om
tecknet ">" har sin normala matematiska betydelse gäller nästan
alltid/enbart>mest>vanligen>ofta>ibland>någon gång. Med en kvali-

kvalitativ dimension tillagd har också särskilt>gärna använts (där kanske kvantitativt särskilt ≈ vanligen och gärna ≈ ofta).

Med portmanteau-parenteser förstås parenteser som använts för att komprimera definitionstexten genom att sammanföra flera beskrivningsled. Ofta rör det sig om en trivial sammanföring av olika böjningsformer av ett och samma ord, som i

(49) **åtgärda** 'göra till föremål för åtgärd(er)'

(50) **värma** 2/1 'göra varm(are)'

Denna typ av parenteser kan alltid framgångsrikt ersättas med en disjunktion:

(49') **åtgärda** 'göra till föremål för åtgärd eller åtgärder'

(50') **värma** 2/1 'göra varm eller varmare'

De mer korrekta omskrivningarna i (49') och (50') är onekligen stilistiskt något klumpiga, varför det kunde vara rimligt att använda portmanteau-parenteserna i ordbokens definitioner. I nuvarande läge förekommer definitioner av typ (49) och (50) även i databasen, där de i princip ej borde finnas (även om denna typ av portmanteau-parentes är tämligen oskyldig). - Man kan här notera att åtskilliga definitionsord av neutralt genus är avsedda att vara funktionellt dubbeltydiga i fråga om numerus.

En mer problematisk typ är den där parenteserna innehåller en hel strukturell nivå i definitionen, som i

(51) **upphovsman** 'person som skapat (grundtankarna i) viss (tankemässig) företeelse'

(52) **adhesion** 'attraktionskraft mellan (molekylerna i) två kroppar i beröring med varandra'

Här är det mer tveksamt om portmanteau-parenteserna verkligen kan ersättas med disjunktioner, eftersom de alternativa leden ej har samma status. I fallet (51) tyder förekomsten av exempel som den ursprunglige/egentlige upphovsmannen på en distinktion mellan en

"ideell" och en "teknisk-juridisk" betydelsenyans (där den förra uppenbarligen måste fokusera grundtankarna). - Observera att den andra parentesen i (51) är en s.k. semantisk parentes! - I fallet (52) är det tydligt att inkluderandet av parentesinnehållet innebär en "vetenskaplig" precisering av definitionen, snarare än ett alternativ. Av dessa skäl måste

(51') **upphovsman** 'person som skapat viss (tankemässig) företeelse eller grundtankarna i viss (tankemässig) företeelse' eller flera bokstäver uttalsnests'

bedömas som föga övertygande, medan

(52') **adhesion** 'attraktionskraft mellan två kroppar, eller mellan molekyler i två kroppar, i beröring med varandra'

helt enkelt är felaktig och vilseledande. Troligen bör den information som dylika parenteser avses förmedla snarare redovisas i form av en transfererad betydelse.

Den mest elakartade typen av portmanteau-parentes är den som innehåller ett alternativt huvudord (vilket inleder definitionstexten), som i

(53) **adjunkt /1** '(titel för) ämneslärare på högre skolstadier'

(54) **skepsis** '(benägenhet till) tvivel eller misstro'

Typen i (53) är relativt frekvent i ordbok och databas. Både för (53) och (54) vore en omskrivning i form av en disjunktion nonsens (i ännu högre grad än vid (51)-(52)). Det är uppenbart att parenteserna i (53) och (54) tillgripits i syfte att spara in en transfereringscykel som bedömts som trivial och/eller alltför utrymmeskrävande. Denna sparsamhet går dock ej att rättfärdiga med några teoretiska argument. Sammanfattningsvis står det klart att den från början ej förutsedda typ av parenteser som här kallats portmanteau-parenteser bör utrensas ur databas och ordbok, möjligen med undantag för fallen med alternativa böjningsformer (som i (49), (50)), vilka av stilistiska skäl kan tolereras i en ordbok. - Som framgått borde en portmanteau-parentes kunna omskrivas med disjunktion; ett väl så viktigt yttre kännetecken (i substantivdefinitioner) är att portmanteau-parenteser normalt innehåller

N+Prep, vilket klart särskiljer dem från syntaktiska och semantiska parenteser.

Illustrativa parenteser finns av två typer, för vilka det gemensamma är att parentesinnehållet i princip skulle kunna ersätta definitionen eller en viss del därav. Vid den första typen är parentesinnehållet icke-språkligt, som i

(55) apostrof 'ett litet tecken (') som bland annat anger att en

Här är parentesinnehållet uppenbarligen en bild av definiendum som (i viss utsträckning) skulle kunna ersätta den semantiska beskrivning som ges i definitionen men på intet vis kan anses ingå i denna. I exemplet ges bilden visserligen i direkt anslutning till definitionens huvudord men intet hindrar att den placeras fristående. Denna typ av illustrativa parenteser tillhör alltså ej den egentliga databasen (och har införts i avsikt att göra ordboken mer läsartillvänd). - Frågan om hur semantiska beskrivningar i ord skulle kunna kompletteras med bildbeskrivningar (och hur de, eventuellt, skulle kunna integreras till en kombinerad "lexiko-visuell" databas) är föga undersökt på det teoretiska planet. På det praktiska, lexikografiska planet har visserligen s.k. bildordböcker existerat under lång tid men det kan diskuteras huruvida bilderna har tillfört något väsentligt, utom möjligen inom vissa begränsade områden.²³ De "bilder" som i SOB (och LDB) förekommer inom parenteser är givetvis av en ytterst speciell typ, eftersom det enbart rör sig om typografiska tecken. Varför just denna kategori skulle behöva illustreras är ej klart (särskilt som illustrationen lika väl kan utföras i ett språkligt exempel - vilket faktiskt också gjorts i fallet apostrof).

Den andra typen av illustrativa parenteser karakteriseras av att innehållet utgörs av ett mer eller mindre specialiserat begrepp

²³ De ordböcker som producerats i det s.k. LEXIN-projektet (se t.ex. Svenskt-serbokroatiskt lexikon) innehåller ett "bildappendix" som illustrerar ett strategiskt urval av dylika begränsade områden, vilket förefaller vara en bättre lösning än de som använts i traditionella bildordböcker.

som kan ersätta den närmaste föregående frasen (eller ordet). Som exempel kan anföras

(56) tillhållarlås 'dörrlås med (visst antal) rörliga spärrar
(tillhållare)'

Här är uppenbarligen tillhållare ett alternativ till frasen rörliga spärrar. Anledningen till att tillhållare ej använts direkt i definitionen (trots principen om liten hierarkisk räckvidd, se 5.1), torde vara att ordet bedömts vara alltför tekniskt eller specialiserat. (Därvid är att märka att tillhållare självt ej definieras som en rörlig spärr.) Denna typ av illustrativ parentes har alltså använts för att, inom ramen för en analytisk, allmänspråklig definition, i förbigående ge en upplysning om en speciell fackterm, som i allmänhet torde ha bedömts vara mindre förtrogen för språkbrukaren än definiendum självt. Det kan diskuteras huruvida en inskjuten parentes i definitionstexten är det bästa sättet att förmedla denna typ av information ens i en ordbok; i en databas bör en mer explicit beskrivningsform väljas. Om verkligen facktermen/mellanbegreppet är klart mindre språkligt central än definiendum (och således ej bör användas direkt i definitionen) finns åtminstone två goda varianter: dels att integrera facktermen i definitionen med hjälp av en förklarande fras, som i

(56') tillhållare 'dörrlås med (visst antal) rörliga spärrar som kallas tillhållare',

dels att ange facktermen i definitionstillägget, som i

(56'') tillhållare 'dörrlås med (visst antal) rörliga spärrar'
vilka kallas tillhållare och /.../

(En tredje möjlighet vore att meddela informationen i ett s.k. encyklopediskt exempel, men denna utväg är knappast mer explicit än en illustrativ parentes.)

Sammanfattningsvis gäller att de illustrativa parenteserna bryter mot definitionssystemets principer, antingen genom att införa en icke-lexikalisk dimension i beskrivningen eller genom att tillåta

strukturellt föga välbestämda alternativa definitioner.²⁴ Följaktligen bör illustrativa parenteser lika väl som portmanteau-parenteser, utmönstras ur databasen. - Parentestypen igenkänns på att den kan parafraseras antingen med hjälp av "som liknar vidstående bild" e.d. (jfr (55)) eller med hjälp av "som kallas ..." e.d. (som i (56)). uppriktig ett karaktärsdrag; öppet en definitionerna i (58) och (59) är riktiga. Är det ävenledes uppenbart

Med de modifikationerna som bruket av de olika parentestyperna föranleder, borde alltså gälla att varje tillräckligt högt rankad semantisk faktor som urskilts vid betydelseanalysen av en given kärnbetydelse skulle beskrivas av ett eget ord eller en egen fras i definitionstexten (medan lägre rankade faktorer kunde beskrivas i definitionstillägget eller eventuellt på annat sätt).²⁵ - Ett undantag har redan nämnts ovan, nämligen det (tämmligen frekventa) fall där ett delkomplex av faktorer kan beskrivas med ett i definitionssystemet väl användbart mellanbegrepp. - De ej sällsynta klart icke-analytiska definitioner som uppträder i databas och ordbok kan följaktligen ha uppkommit antingen genom bristfällig faktoranalys eller på grund av formuleringsproblem (i den mån dessas orsaker går att särskilja; jfr 3.2). Särskilt ofta synes adjektivdefinitioner vara icke-analytiska. Man kan här jämföra

(57) **frimodig** 'öppen och orädd'

med

(58) **frank 1** 'som säger sin mening utan omskrivningar'

(59) **uppriktig** 'som låter sitt handlande svara mot sina egentliga (innersta) känslor och övertygelser'

²⁴ Det är också möjligt att betrakta de illustrativa parenteserna av den andra typen som ett slags inbäddade och indirekta definitioner av ifrågavarande mellanbegrepp ((56) vore alltså sekundärt även en definition av tillhållare, i en specialbetydelse). Också med detta betraktelsesätt innebär dylika parenteser ett brott mot definitionssystemets principer.

²⁵ Ett tilläggskrav på den analytiska definitionen, som följer av de allmänna principerna i 3.2, är att den syntaktiska strukturen i definitionstexten i möjligaste mån skall avspegla den semantiska struktur som råder mellan de semantiska faktorerna. Detta krav är av större intresse för fortsättningsprojektet (se Järborg 1988), varför det här endast omnämns.

Det är uppenbart att (57) ej kan betraktas som analytisk medan (58) och (59) klart visar att en tämligen hög grad av analyticitet är möjlig för adjektiv tillhörande samma "semantiska fält" (låt vara att frimodig primärt tycks beskriva en attityd, frank ett sätt att uppträda, uppriktig ett karaktärsdrag). Oavsett om definitionerna i (58) och (59) är riktiga, är det ävenledes uppenbart att de, med sin explicita formulering, är mycket lättare att kontrollera (och falsifiera) än den vaga definitionen i (57). De tämligen frekventa adjektivdefinitionerna av typen (57) torde i allmänhet kunna skyllas på otillräcklig betydelseanalys (snarare än på formuleringssvårigheter). Som exemplen visar bör det vara möjligt, om än ofta tidsödande, att ersätta dylika oanalytiska definitioner i databasen med analytiska.

Här är icke platsen för någon mera ingående exemplifiering och diskussion av olika typer av oanalytiska definitioner, eftersom detta egentligen skulle kräva detaljerade betydelseanalyser i varje enskilt fall. Dock bör nämnas att ett förbluffande stort antal lexem har "definierats" med definitioner bestående av endast ett ord (s.k. synonymer); i ordboken c:a 6000 lexem = 1/11 av det totala antalet. Betydelsebeskrivning enbart med synonym uteslöts visserligen ej i de ursprungliga projektplanerna, men det förut-sattes att detta alternativ skulle användas enbart för lexem tillhörande de mest perifera delarna av det lexikaliska systemet (t.ex. obsoleta, högtidliga eller skönlitterära ord som rent innehållsligt svårligen kunde skiljas från mer kuranta närsynonymer. Se även Ralph-Järborg-Allén 1977). Emellertid har ettorsdefinitioner även kommit att användas vid åtskilliga icke-marginella lexem under lexikaliskt centrala och högfrekventa lemman, som i fallet

Här ska vi ge ett idiomatiskt uttryck, som i databasen själv beskrivs som ett sådant, dels under ta/1, dels under ta/2, vilket senare lemman enbart förekommer i idiom. I princip kan ett idiom

- (60) (a) ta/8 'utföra' faktorer och därmed byta mot analytisk
 (b) ta/10 'bedöma' dock ta vara på betydelsen av en enda faktor, men
 (c) ta/12 'åstadkomma' gånghastigt. Definitioner med idiom
 (d) ta/15 'vika av' som kräver på definitionssystemets in-

(Här bör även (d) räknas som en ettordsdefinition, eftersom partikelverbet kan vara fast sammansatt.) Verbet ta är det lemma i databasen som är uppdelat på flest lexem (nämligen 16) och det är lite beklagligt att dessa är beskrivna med så varierande grad av analyticitet, från mycket detaljerade definitioner (som ta/1 m.fl.) till de totalt oanalytiska ettordsdefinitionerna. (Även om man ej skulle kunna analysera fram en distinktion mellan t.ex. ta/10 och bedöma vore det att föredraga att definitionen av bedöma användes även för ta/10, vilket åtminstone skulle skapa balans mellan lexemen under ta och göra dem lättare jämförbara inbördes.) Eftersom en så långt driven lexemuppdelning som i fallet ta rimligtvis måste grundas på en omfattande betydelseanalys, borde det ej saknas underlag (i form av utskilda semantiska faktorer) för mer analytiska definitioner än de i (60) (a) - (d); troligen bör avsteget från principerna i detta och liknande fall snarast skyllas på formuleringsproblem. - Det bör påpekas att de (när)synonymer som använts såsom ettordsdefinitioner mycket lätt kan beredas plats i den totala semantiska beskrivningen av ifrågavarande lexem, i de fall de bedöms kunna på ett väsentligt sätt bidra. Möjligheten finns nämligen att föra en god närsynonym till den särskilda synonymkategorin i FDV-formatet (vilken i ordboken återges med förkortningen SYN. samt synonymen, allt inom vinkelparenteser). Se vidare avsnitt 5.4

En icke önskvärd definitionstyp som lämpligen kan tas upp i detta sammanhang är den där definitionstexten, eller en central del därav, består av ett idiom. Denna typ torde vara relativt vanligare vid parafraseringar av andra idiom än vid normala lexemdefinitioner, men uppträder dock någon gång även vid dessa, som i

(61) hjälplos 'som inte kan ta vara på sig själv'

Här är ta vara ett idiomatiskt uttryck, som i databasen själv beskrivs som ett sådant, dels under ta/1, dels under vara 2, vilket senare lemma enbart förekommer i idiom. I princip kan ett idiom omfatta flera semantiska faktorer och därmed bryta mot analyticitetskravet. I (61) torde dock ta vara beteckna en enda faktor, men analyticiteten blir knappast genomskinlig. Definitioner med idiom riskerar vidare att bryta mot kravet på definitionssystemets interna koherens (se 5.1) genom att idiomens betydelse sällan är tillräckligt väl beskriven i databasen. I fallet (61) råkar idio-

met vara beskrivet på olika sätt på de båda beläggsställena, vilket givetvis är ett elakartat brott mot den interna konsistensen. Slutligen kan idiom i princip ej anses tillhöra de mest centrala delarna av vokabulären, varför idiom i definitioner i allmänhet måste komma att motverka den systematiska uppbyggnaden av definitionssystemet (jfr 5.1). På det praktiska planet gäller, parallellt härmed, att idiom (under i övrigt lika omständigheter) måste försvåra uttolkningen av definitionen för ordbokens läsare.

Återstoden av detta avsnitt skall ägnas åt några speciella principer för formulering av definitioner (= objektspråklig beskrivning av primära semantiska faktorer) som tillämpats i LDB. En första punkt att observera är att det ej råder någon entydig motsvarighet mellan faktorer och definitionsord: många semantiska faktorer (låt återges med fraser. Detta gäller icke minst den (vid verb och substantiv förekommande) s.k. huvudkategorin, som kan motsvaras av definitionens syntaktiska huvudord men ej behöver göra det. Ett tämligen vanligt fall kan representeras av

(62) **oxidation** 'kemisk reaktion mellan syre och ett annat grundämne'

Här motsvaras huvudkategorin ej av det syntaktiska huvudordet reaktion utan av frasen kemisk reaktion (eftersom reaktion/2 är alldeles för generellt). De flesta "huvudkategorifraser" torde ha denna struktur, nämligen Adj-N. Men även strukturen N-PrepP uppträder som huvudkategorifras, som i

(63) **slalom** 'utförsåkning på skidor längs en slingrande bana som markeras med uppställda stavar'

där frasen alltså är utförsåkning på skidor.²⁶

För de tydligaste och frekventaste huvudkategorifraserna svarar ett relativt litet antal typer med strukturen N-PrepP, som alla (i

²⁶ Vid vissa uttryck, särskilt Adj-N, som skenbart är huvudkategorifraser, är huvudordet i och för sig tillräckligt specifikt men ambiguöst; eftersom lexemnummer o.d. ej tolereras i definitioner är bestämningen att betrakta som en disambiguering i stället för en specificering. Se vidare Järborg 1988; jfr även exempel (37) i 5.1.

(i motsats till vad fallet var i (62) och (63)) lätt kan utbytas mot sammansättningar. Hit hör typer som

(64) **bröst/1** 'främre övre delen av bålen'

och det i kapitel 4 anförda

(10) (b) **järnväg** 'system för (längre) transporter av gods eller passagerare, som använder sig av dragfordon med sammankopplade vagnar vilka framförs på och är bundna till speciella banor av utlagda stålskenor'

Som synes kunde (64) i princip omformuleras med det enkla huvudordet båldelen och (10) (b) på samma sätt med transportsystem (låt vara att det förra ordet är en något ovanlig sammansättning och det senare ej lätt kan integrera den fakultativa bestämningen längre). Det är givet att ord som del och system fordrar komplettering med argument för att bli tillräckligt specifika för användning i definitioner.

I de anförda fallen har det syntaktiska huvudordet ej ensamt varit semantiskt huvudord men dock varit kärnan i huvudkategorifrasen. (Med en provisorisk term skulle denna senare typ av huvudord kunna kallas "semantiska huvudord av andra graden". När det i fortsättningen talas om semantiska huvudord görs ingen åtskillnad mellan dessa typer, vilka ändå i många fall kan vara svåra att särskilja.) Som noterats i kapitel 4 är emellertid många substantiviska definitioner medvetet uppbyggda så, att det syntaktiska huvudordet följs av det egentliga semantiska huvudordet i en prepositionsfras (prepositionsattribut). De "formella" huvudord det gäller tillhör en liten begränsad grupp (typ, system, uppsättning, endera m.fl.). Något behov av denna konstruktionstyp hade ej förutsetts vid projektets början utan kom att framväxa under arbets gång; det är därför förståeligt att konstruktionstypen aldrig blev tillfredsställande analyserad och bruket av den ej helt konsekvent.

Vid de flesta definitioner innehållande formellt huvudord visar sig den semantiska beskrivningen av definiendums huvudkategoriska vara funktionsuppdelad i en substantiell del och en del som endast klassificerar den typ av entitet i vilken huvudkategorin uppträ-

der. Det formella huvudordet används därvid för att beskriva vissa avvikande entitetstyper; det är alltså icke enbart ett syntaktiskt huvudord. I det överväldigande antalet substantiviska definitioner är entitetstypen självklar: det rör sig om enstaka objekt (konkreta eller abstrakta) eller också ämnen e.d. i massform. Då något av dessa normalfall ej föreligger blir däremot frågan om entitetstyp intressant, som vid

(65) **skänkel/1** 'endera av de två rörliga armarna på sax eller tång'

Här rör det sig som synes om objekt som primärt uppträder två och två. Substantiellt är alltså en skänkel en rörlig arm men den tillhör en entitetstyp som naturligen grupperas i par. På motsvarande sätt beskrivs vad man skulle kunna kalla "spontana" respektive "organiserade" kollektiv i (66) respektive (67):

(66) **lövverk** 'ansamling av lövbärande grenar och kvistar'

(67) **bjälklag** 'system av bjälkar som bär upp en byggnads golv och innertak och håller samman väggarna'

(Man kan notera pluralformen hos de egentliga huvudorden.)

Det vanligaste formella huvudordet, typ (av), beskriver i sin normalanvändning en något abstraktare entitetstyp, nämligen en som knappast uppträder annat än som taxonomisk klass. Betrakta åter

(48) **abborrfisk** 'typ av (sötvattens)fisk besläktad med abborren'

Det är föga troligt att abborrfisk någonsin används konkret om enstaka objekt (man må testa med t.ex. Vi fångade tre abborrfiskar i morse, Till middag blir det stekt abborrfisk); normalanvändningen torde vara i generiska satser som Abborrfiskar kännetecknas i allmänhet av..., Abborrfiskarna är närmast besläktade med... o.d.

Abborrfisk uppträder alltså praktiskt taget enbart som en taxonomisk klass (något som sekundärt givetvis alltid är möjligt för de flesta substantiv) och "entitetstypsbeskrivningen" typ (av) tycks vara fullt adekvat. Emellertid visar det sig att detta formella

huvudord har överanvänts i rätt stor utsträckning. Man kan jämföra (48) med

(68) falk 'typ av snabbflygande rovfågel, med nedåtböjd övernäbb, som kan slå bytet i luften'

Visserligen behöver falk specificeras, normalt med en förled, för att det skall bli fråga om egentliga artnamn (jfr pilgrimsfalk, aftonfalk, stenfalk, tornfalk) men det enkla falk används mycket ofta konkret (och troligen relativt sällan som taxonomisk klassbeteckning), detta i stark motsats till abborrhisk. Här föreligger således en konflikt mellan ett språkligt kriterium (falk som konkret beteckning) och ett encyklopediskt kriterium (falk som en principiell beteckning för en taxonomisk klass, i detta fall en s.k. underfamilj). Enligt principerna i kapitel 3 borde de encyklopediska semantiska faktorerna underordnas de allmänspråkliga och eftersom beskrivningen skulle avspegla analysen borde det formella huvudordet typ (av) ej använts i detta fall. Detta blir ännu tydligare om man lämnar de biologiska taxonomierna och i stället betraktar artefakter som

(69) falkonett 'typ av lättare kanon'

I detta fall (och delvis även i (68)) skall typ (av) ej i första hand ses som en beskrivning av en entitetstyp utan som en dummy för en möjlig encyklopedisk specificering. Detta är i princip samma funktion som den som fylls av den egendomliga, men lexikografiskt hävdvunna, användningen av obestämda artikeln, som i det tidigare anförda

(4) lunnefågel 'en svartvit sjöfågel med hög, i sidled hoptryckt näbb'

Enligt principerna i avsnitt 3.2 skall encyklopediska semantiska faktorer (sådana som de mer tekniska karakteriseringarna) ej rankas högt och följaktligen ej normalt vara representerade i definitionen. Det är däremot både rimligt och önskvärt att markera existensen av dylika specificerande faktorer, särskilt i sådana fall som (69) och (4) där beskrivningen redan antyder möjligheter till ytterligare specificering. Den naturliga markören vore frasen ett slags, som tydligt utsäger vad som fordras: nämligen att ytter-

ligare specificering skulle kunna anföras men ej befunnits lämplig i själva definitionstexten. Slutsatsen blir alltså att typ (av) bör renodlas som beskrivning av en viss entitetstyp (taxonomisk klass) och att både typ (av) och en/ett i funktionen av "encyklopedisk dummy" bör ersättas av ett slags. (Genom en sådan förändring skulle också, såsom biprodukt, en bättre syntaktisk utbytbarhet erhållas.)

I detta sammanhang bör även bruket av det (att) i definitioner av verbalsubstantiv och egenskapsbeteckningar kommenteras, trots att det ej rör sig om något formellt huvudord i den ovan anförda meningen. Det (att) infördes på ett tidigt stadium som en lexikografisk anpassning i den s.k. läsarvänlighetens intresse; det förenklar också otvivelaktigt formuleringsarbetet. Nackdelarna torde dock överväga: dels är uttrycket onödigt oprecist, dels utgör det en dålig utgångspunkt för beskrivning av eventuella transfererade betydelser, dels är det (som konstaterats i 4.2) ej fullt syntaktiskt utbytbart. Uttrycket det (att) bör därför ersättas, förslagsvis med någon av typerna handling/förlopp/tillstånd/... (bestående i att) vid verbalsubstantiv samt med egenskap (bestående i att) vid egenskapsbeteckningar. Det kan också ifrågasättas om icke definitionerna av verbalsubstantiv skall utbyggas ytterligare, genom att den följande infinitiven alltid förses med argument, där så är möjligt, eller eventuellt genom att infinitivfrasen ersätts med en att-sats, som i

(70) **borrning**

- (a) 'det att borra'
- (b) 'det att borra i eller genom ngt'
- (c) 'handling bestående i att borra'
- (d) 'handling bestående i att borra i eller genom ngt'
- (e) 'handling bestående i att ngn borrar i eller genom ngt'

De nu existerande typerna är (a) (som är den som föreligger vid borra) och (b); rekommenderad typ är (d) eller kanske (e). Det faktum att definitionerna blir längre med den föreslagna omformuleringen kan tas med jämnmod, eftersom ifrågavarande lemman (normalt avledningar med -ning respektive -het) ej har registrerats (och ej bör registreras) i databasen om de ej på något sätt är semantiskt anmärkningsvärda, t.ex. genom att behöva uppdelas i

flera lexem eller genom att det enda lexemet även har transfere-
rad(e) betydelse(r).

I diskussionen kring formella huvudord ovan påpekades behovet av en s.k. encyklopedisk dummy, dvs. en semantisk "platshållare" för en beskrivning av sådana encyklopediska faktorer som ej kan rankas såsom primära. Här skall nu de encyklopediska beskrivningsproblemen tas upp till en mer generell behandling. Till en början erinras åter om grundprinciperna från 3.2, nämligen att de faktorer som upprättades vid betydelseanalysen skulle kunna motiveras språkligt med avseende på ifrågavarande lexems användning i allmänspråk och elementärt (mer eller mindre populärvetenskapligt) fackspråk. Eftersom beskrivningen skall följa analysen blir följden ofta att ord betecknande likartade typer av företeelser i världen kan komma att beskrivas med högst varierande grad av precision och detaljrikedom. Man kan här jämföra

(71) myra 'en liten, normalt vinglös insekt som lever i stora välorganiserade samhällen'

(72) myrbagge 'en skalbagge' som till utseendet påminner om en myra

Den kraftiga obalansen mellan (71) och (72) kan verka inkonsekvent för en entomolog men är mycket lätt att motivera språkligt. Myra torde vara en av de frekventare djurbeteckningarna i allmänspråket och uppträder ofta i liknelser o.d. (en transfererad betydelse och ett idiom har mycket riktigt upprättats under myra i databasen), medan det verkar osannolikt att man alls skulle påträffa myrbagge i någon än så stor allmänspråklig korpus. Det kan vidare noteras att två encyklopediska (och systematiskt väsentliga) faktorer faktiskt beskrivs i (71), beroende på att dessa (vinglös, samhällsbildande) är språkligt motiverade, vilket i sin tur givetvis ursprungligen beror på att motsvarande egenskaper hos företeelsen myra är lätta att observera för lekmannen. Allmänt sett finns två huvudkällor för de språkligt motiverade karakteriserande faktorerna hos artnamn o.d.: dels lätt observerbara yttre egenskaper (som i fallet myra), dels funktion i den mänskliga kulturen (i alla betydelser). Den senare typen kan illustreras med följande exempel, där undantagsvis också en jämförelse med en annan ordbok verkar befogad:

- (där t.ex. myra skulle höra till den förra, brödfrukt till den
- (73) ISO:
- (a) **brödfrukt** 'frukt av brödfruktträdet' och klassbeteckningar:
 - (b) **brödfruktträd** 'trädet Artocar'pus inci'sa'
- (74) LDB:
- (a) **brödfrukt** 'en stor stärkelserik frukt varav mjöl till bakning kan beredas'
 - (b) (**brödfruktträd** endast redovisat som exempel under **brödfrukt**)

Den strikt botaniska beskrivningen i (73) är säkerligen systematiskt korrekt, även om det latinska artnamnet torde äga utomordentligt ringa förklaringsvärde för en icke-specialist (se dock om exakta encyklopediska uppgifter nedan). Däremot är (74) den språkligt mycket bättre beskrivningen, och detta redan av det skälet att det primära ordet i (74) är brödfrukt medan det i (73) är brödfruktträd, helt i strid med den morfologiska strukturen. Beskrivningen i (74) (a) ger en förklaring redan till de grundläggande sammansättningsleden och förklaringen avser just en kulturellt (närmare bestämt ekonomiskt) betydelsefull egenskap. Intetdera av orden brödfrukt eller brödfruktträd torde vara lätt att påträffa i en textkorpus, men ingen tvekan råder om att det är frukten snarare än trädet som den ordinäre språkbrukaren kan tänkas vara förtrogen med. Denna (relativa) förtrogenhet innehåller troligen även en föreställning om att brödfrukt är ett slags baslivsmedel på söderhavsöar (en föreställning härrörande närmast från enklare äventyrslitteratur). Beskrivningen i (74) (a) kunde för övrigt gärna kompletteras med någon geografisk bestämning, i definition eller definitionstillägg. För fullständighetens skull skall tillfogas att (73) har illustrerats med en liten teckning av en brödfrukt (på kvist), som verkar tillföra föga information.

Kulturstorlek

Det bör kanske poängteras att den typ av beskrivning som i LDB används för artnamn o.d. visserligen ej avspeglar någon biologisk systematik men ej på grund av detta är mindre "vetenskaplig". Primärt är beskrivningarna alltså grundade på den språkliga verkligheten (som analyserats enligt språkvetenskapliga principer) men språket reflekterar, som visats ovan, väsentliga iakttagelser av sådant som form, färg, levnadssätt, ekonomisk-kulturell betydelse etc. Därför blir ofta sekundärt de språkligt grundade beskrivningarna även vad man skulle kunna kalla ekologiska eller etnologiska

(där t.ex. myra skulle höra till den förra, brödfrukt till den senare typen). I detta sammanhang bör man erinra sig existensen av "rent" ekologiska och etnologiska klasser och klassbeteckningar: sålunda betecknar ett ord som gräsätare en ekologisk gruppering och ord som skaldjur, ädelfisk etnologiska grupperingar (från något skiljaktiga perspektiv), utan något direkt samband med systematiskt bestämda grupper. (Det faktum att dylika klassbeteckningar är icke-systematiska kan ibland behöva påpekas, vilket i så fall vanligen har utförts genom en användningskommentar; se vidare 5.4.)

Vid artnamn och liknande beteckningar på naturföreteelser tenderar konflikten mellan språklig och encyklopedisk analys och beskrivning att vara störst, vilket motiverar den relativt ingående behandlingen ovan. För hanteringen av encyklopediska faktorer i allmänhet utvecklades vissa generella principer som nu närmast skall presenteras. Åtskilliga encyklopediska faktorer avser egenskaper som skulle kunna uttryckas med exakta måttuppgifter, med exakta uppgifter om belägenhet i tid och/eller rum etc. För beskrivningen av dylika faktorer infördes principen att s.k. exakta koordinater så långt möjligt skulle undvikas i definitionerna; de skulle uttryckas i en relativ skala i en mer eller mindre funktionell dimension. (Det är givetvis omöjligt att undvika exakta koordinater vid t.ex. nationalitetsadjektiv.) Jämför här

- (75) sand 'material bestående av små mineralkorn'
(76) grus 'material bestående av grovkorniga stenpartiklar'

(77) singel 2/4 'naturligt förekommande stenmaterial med 7-70 mm kornstorlek'

Den kritiska faktorn är här materialkornens grovlek som uttrycks med relativa mått i (75) och (76), låt vara att de kunde gjorts bättre inbördes jämförbara. I (77), däremot, ges en bestämning i millimeter, vilket strider mot den generella principen. Till saken hör att både sand och grus i tekniska sammanhang brukar definieras med ett exakt grovleksintervall; visserligen torde singel i sig vara ett mer "tekniskt" lexem än de andra men det är knappast skäl

nog att frångå grundprincipen. Intet hindrar att det exakta måttet ges i ett definitionstillägg, som vid

(78) **starkvin** 'vin med hög alkoholhalt' enl. svensk lagstiftning 16-22 volymprocent

(Här borde nog också funnits en uppgift om att sådant vin ej kan framställas enbart genom naturliga jäsningsprocesser, vilket nog är mer grundläggande än vad som råkar stipuleras i ett visst lands lagar. Jfr även nedan.)

En annan möjlighet är att använda människans egna mått i beskrivningen, som i

(79) **klappersten** vanl. koll. 'knytnävsstor rund sten' i större mängd, vid stränder och i åkerjord

(Vi kan i detta sammanhang notera att det artificiella metersystemet - eller SI-systemet - är svårdefinierat i ord och i sista hand måste återföras just till människans egna mått. De äldre måttssystemen är från denna synpunkt mycket bättre!)

Ett mer subtilt exempel på tillvägagångssättet vid exakta koordinater är

(80) **medeltid** vanl. best.f. sing. 'tiden mellan Roms fall och Amerikas upptäckt' kännetecknad av stark kyrkomakt, feodalväsen och relativt primitiv ekonomi

Eftersom definiendum är en tidsperiod vore det enklast att direkt ange ändpunkterna (säg, i runda tal, 500 - 1500 e. kr.). En sådan beskrivning har dock, förutom den ordinära nackdelen att vara onödigt teknisk och "exakt", knappast något förklaringsvärde i förhållande till det urval av karakteriserande andrarangsfaktorer som beskrivs i definitionstillägget. De "relativa ändpunkter" som använts i (80) är visserligen historiska händelser som sekundärt kan ges en exakt förankring i tiden men poängen är att de har ett funktionellt samband med de karakteriserande faktorerna: med Roms fall förloras en statlig centralmakt m.m., med Amerikas upptäckt följer expansion av handeln m.m.

samhälleliga bestämmelser och konventioner. Principen är här att dylika s.k. accidentella faktorer skall undvikas i definitionerna och i stället omskrivas i funktionella termer. Ett enkelt exempel är

(81) (a) rattfylleri '(brott som består i) framförande av motorfordon i berusat tillstånd' i Sverige med en alkoholhalt i blodet av över 1,5 promille

(b) rattonykterhet '(brott som består i) framförande av motorfordon i lindrigt berusat tillstånd' i Sverige med en alkoholhalt i blodet av mellan 0,5 och 1,5 promille

Hur t.ex. lindrig berusning exakt skall avgränsas är naturligtvis en tillfällighet (något som understryks av det faktum att gränsen i skrivande stund skall sänkas). Uppläggningsen i (81), med en funktionell beskrivning i definitionen och en exakt promillegräns i definitionstillägget följer helt grundprinciperna. (Jfr även exempel (78)).

I följande exempel är den accidentella faktorn ej något exakt mått utan ett konventionellt villkor:

(82) (a) ISO **socionom** 'person som är diplomerad från socialinstitut'

(b) LDB **socionom** 'person som har yrkesutbildning för sociala yrken' och som har examen från socialhögskola e. d.

Definitionen (a) är, som synes, formulerad uteslutande som en beskrivning av en accidentell faktor (som numera råkar vara föråldrad), nämligen en ytterst specifik utbildning. I (b) däremot är själva definitionen formulerad i funktionella termer och den närmare beskrivningen av utbildningen (vilken ej är så bindande som i (a)) återfinns i definitionstillägget. Härigenom blir (b) ej så tidsbunden som (a) och ej heller lika geografiskt bunden. Det förtjänar nämligen att påpekas att beteckningar på samhälleliga företeelser mycket väl kan användas för sina närmaste utländska motsvarigheter, varvid en "exakt" beskrivning av accidentella faktorer snarast blir ett handikapp. Jämför

svarigheter, varvid en "exakt" beskrivning av accidentella faktorer snarast blir ett handikapp. Jämför

(83) (a) ISO borgmästare 'ordförande i rådstugurätt och magistrat (el. i någondera av dem)'

(b) LDB borgmästare 'högste ämbetsman i stad' med varierande ställning och befogenheter under olika tider och i skilda länder

Det torde vara enkelt att visa, att ordet borgmästare ofta används om funktionärer i andra tider och/eller i andra länder än dem för vilka den mycket speciella beskrivningen i (a) är giltig. Den funktionella (ehuru givetvis vagare) beskrivningen i (b) är därför att föredraga. Till yttermera visso kompletteras (83) (b) med två transfereringscykler med specialiserade betydelser, varav den ena i stort sett överensstämmer med (a); den andra är den nästan konträra 'vald ledare i stad'. Båda dessa underbetydelser täcks givetvis av huvudbeskrivningen i (83) (b). Det kan tyckas att denna vaga beskrivning strider mot en av de teoretiska grundprinciper som uppställdes i 2.2, nämligen den om de pregnanta kärnbetydelserna. Motsägelsen är dock skenbar: borgmästare är, i det allmänna språkbruket, helt enkelt en mycket vidsträckt beteckning, ej en exakt term.

En annan typ av systematisk vaghet i beskrivningen må här beröras i korthet, nämligen den som använts vid s.k. klassificerande adjektiv. Som exempel kan tagas ett nationalitetsadjektiv som tysk:

(84) (a) ISO tysk 2 'som tillhör eller utmärker eller stammar från Tyskland'

(b) LDB tysk 1 'som har att göra med Tyskland'

Problemet med en beskrivning som den i (a) är att den verkar uttömmande utan att egentligen vara det. (Jfr ett motexempel som kungens tyska resa 'kungens resa till Tyskland', där resan uppenbarligen varken tillhör, utmärker eller stammar från Tyskland.) I dylika fall är en funktionellt vag beskrivning som den i (84) (b) lämpligare, eftersom den ej är vilseledande. - Som nämnts i kapitel 4 har för klassificerande adjektiv även formuleringen som av

ser X använts; trots sin något naiva stil är dock som har att göra med X tydligare och därför lämpligare.

För utpräglade facktermer gäller att de praktiskt taget uteslutande bestäms av encyklopediska faktorer, vilka därtill ofta är svåra att beskriva på ett någorlunda kortfattat sätt utan användning av andra facktermer, ofta lika lexikalt perifera som definiendum och alltså, enligt principerna i 5.2, ej tillåtna i definitionen. I de fall där det ej rörde sig om de mest centrala eller högrankade faktorerna visade det sig lämpligt att antydast existensen av dylika encyklopediska faktorer i stället för att försöka beskriva dem.

Som nämnts ovan har formuleringen typ av ofta (felaktigt) använts som dummybeskrivning i sådana fall, liksom också obestämd artikel. Bland övriga dummybeskrivningar kan nämnas särskild och viss, det sistnämnda dock även ofta använt för att markera argument (se exempel (6)). Givetvis kunde denna typ av antydningar ej användas för de mest centrala faktorerna. För dessa utvecklades en beskrivningsteknik som skulle kunna kallas renodling och innebar en fokusering på den mest typiska och karakteristiska aspekten. Betrakta

(85) derivata 'matematisk funktion som beskriver (viss) annan funktions variation' bl.a. var den växer resp. avtar

Beskrivningen i (85) är på intet vis uttömmande; sålunda utelämnas t.ex. derivatans samband med den ursprungliga funktionen och även det som utsägs är någorlunda giltigt endast för funktioner av en (reell) obekant. Däremot torde (85) beskriva den mest bekanta och kanske den vanligaste användningen av derivatan. I viss mening skulle man kunna hävda att (85) försöker förklara "vitsen med begreppet": varför matematiker finner det meningsfullt att införa och studera denna hjälpfunktion. Det kan noteras att den s.k. renodlingen åter innebär en fokusering av de funktionella aspekterna (oundviklig ordlek), på liknande sätt som vid beskrivningen av artnamn (jfr exempel (74) ovan). Påpekas bör även att i jämförelse med en "renodlad fackterm" som (85) är den fulla fackbetydelsen ofta ett slags generalisering (och därför ibland också en försvagning).²⁷

²⁷ Här finns en utvecklingshistorisk parallell, i det att vissa facktermer ursprungligen haft en relativt pregnant, "renodlad" betydelse som sedan generaliserats av inomvetenskapliga skäl (sådant som

Till det sistnämnda kan fogas att det finns åtskilliga ord som kan betraktas såsom allmänspråkliga i sin mer inskränkta (eller renodlade) betydelse men såsom facktermer i sin generellare betydelse; detta i motsats till flera av de ovan diskuterade fallen. Ett typiskt sådant ord är

(86) **export** 'organiserad utförsel av varor från det egna landet' till utlandet

ibbl. även betr. annat än varor

Från ett fackspråkligt perspektiv vore det möjligen naturligt att betrakta export som en "exakt" term och definiera den 'utförsel av varor och tjänster'.²⁸ Nu är dock "utförsel av tjänster" en mycket abstrakt föreställning utan gott stöd i allmänspråket medan ordet i sig ej gärna kan bedömas som ett rent fackord. Man kan t.ex. notera sammansättningar som exporthamn (som knappast torde hantera tjänster) eller säregna uttryck som osynlig export som poängterar egendomligheten i en "utförsel" som ej avser varor (och som snarast används a posteriori för att "förklara" en oväntat positiv betalningsbalans). Den beskrivning som valts i (86), givetvis grundad på en långt djupare betydelseanalys än den här antydda, utgår, helt korrekt, från en pregnant allmänspråklig kärnbetydelse, som sedan i vissa fackspråkliga sammanhang kan utvidgas till större generalitet men också större vaghet.

De redovisade principerna ger, som framgått, en relativiserad och funktionellt inriktad beskrivning av fackord och andra encyklopediskt avgränsade ord. Beskrivningen kan kompletteras på flera sätt. Möjligheten att ge "exakta koordinater" i definitionstillägget har redan berörts. (Som framgår av exempel (81) borde kanske dylika definitionstillägg vara explicit daterade, för att framhäva att "exaktheten" kan ha begränsad giltighet.) Ibland ges liknande, mer eller mindre exakta, upplysningar i stället i s.k.

"generalisering till n dimensioner" inom matematiken).

²⁸ En dylik lösning föreslogs av journalisten Albinsson-Bruhner i SvD 1987. Se vidare Ohlander 1987.

encyklopediska exempel, vilka diskuteras utförligare i avsnitt 5.4. Som påpekades i inledningen till detta avsnitt bör en semantisk beskrivning i exemplets form endast betraktas som en temporär åtgärd (och möjligen som en bekväm redovisningsform i en tryckt ordbok); detta gäller i högre grad för encyklopediska exempel än för s.k. konnotationsexempel eftersom en explicit beskrivning givetvis är svårare i det senare fallet.

Den viktigaste kompletteringen utgörs emellertid av de fackområdesbeteckningar och liknande som anges i användningskommentarerna för kärnbetydelserna och/eller de transfererade betydelserna. Dessa beteckningar diskuteras utförligare i avsnitt 5.4, men en särskild typ förtjänar att kommenteras här, nämligen den som markerar att definitionen avser en förenklad (ofta "renodlad") fackterm och som kan exemplifieras med

(85') **derivata** 'matematisk funktion som beskriver (viss) annan funktions variation' bl.a. var den växer resp. avtar <noggrannare def. i mat. sammanhang>

Användningskommentaren i (85') signalerar alltså dels att derivata är en matematisk fackterm, dels att den givna definitionen ej är en uttömmande beskrivning av denna fackterm.

Formuleringen i (85'), liksom tidigare de s.k. dummybeskrivningarna (ett slags ... o.d.), väcker frågan huruvida icke en "noggrannare definition" borde ges i databasen i stället för att bara antydas. Till en början måste konstateras att eftersom de föreliggande semantiska beskrivningarna i princip skall vara adekvata för dylika fackords allmänspråkliga kärnbetydelse, skulle de "noggrannare definitionerna" ej ersätta utan endast komplettera dessa beskrivningar. Man kan här jämföra

(87) (a) LDB **meter 1** 'ett grundläggande längdmått' ung. motsvarande en vuxen mans steglängd

(b) ISO **meter 1** 'längdmått motsvarande 1/10 000 000 av avståndet mellan polen o. ekvatorn, enheten i metersys-

temet

- (c) meter 'längdmått som motsvarar längden av en viss måttstav i Paris'
- (d) meter 'ett visst antal våglängder av ljus utsänt från en viss isotop i ett visst excitationstillstånd'
- (e) meter 'en viss bråkdel av den sträcka som ljuset förflyttar sig under en sekund'

Här är (a) en allmänspråklig beskrivning som också ger en praktiskt användbar indikation av den absoluta storleken av enheten (stämmer dock bättre för längdmåttet yard), medan (b) - (e) är "noggranna definitioner" som är praktiskt föga användbara. Definitionen (b), som f.ö. kan misstolkas (avståndet måste mätas längs jordytan, ej i rät linje), representerar det ursprungliga uppfyllda idealet, medan (c) - (e), vilka i läsbarhetens intresse förenklats genom att de exakta uppgifterna endast antytts, representerar olika årgångar av "vetenskapliga" definitioner. Ingen av (b) - (e) skulle kunna göra anspråk på rangen av huvuddefinition, ehuru (e), kanske i den givna formuleringen, samt möjligen (b), i en förbättrad version, skulle kunna vara användbara som något slags vetenskaplig respektive historisk komplettering.

I den mån databasen ej längre betraktas enbart som ett ordboksmanuskript blir utrymmesfrågor mer eller mindre irrelevanta; därför skulle en möjlig lösning kunna vara att de eventuella "noggranna definitionerna" inkorporeras som extra transfererade betydelser (specialiserade eller generaliserade) under de existerande, allmänspråkligt beskrivna kärnbetydelserna. I ett fall som (87) kunde man alltså få en kärna och två extra cykler (en kurant vetenskaplig, en historisk), i ett fall som (85) en kärna och en extra cykel (en fylligare och mer generell matematisk fackbetydelse). Dock måste två reservationer göras. Dels är det icke självklart att varje exakt encyklopedisk detalj verkligen hör hemma ens i en dylik kompletterande "noggrann definition". (Man kan här betrakta (87) (d).) Om beskrivningen av kärnbetydelsen är grundad på en analys av ordets allmänspråkliga användning bör den "noggranna definitionen", i analogi därmed, grundas på en analys av ordets fackspråkliga användning och det är långt ifrån säkert att t.ex.

gärna återfinns i en relativsats, en infinitivbestämning e.d. mot slutet av definitionen. Vid verbdefinitioner gavs den viktigaste informationen vanligen som en svit av strukturellt jämbördiga argument, ofta i form av prepositionsobjekt (det hänvisas här även till resonemanget kring s.k. topikaliserade argument ovan). De mest specifika karakteriseringarna beskrevs vanligen genom att de relevanta argumenten utbyggdes till infinitivkonstruktioner (jfr exempel (17)). Ett alternativ som användes då den specifika karakteriseringen avsåg någon särskild funktion e.d. hos det topikaliserade argumentet (= det syntaktiska objektet) var att lösgöra denna funktion som ett extra argument, inlett av (så)som (jfr exempel (16)). Praxis för åstadkommande av entydighet har redan berörts (se 4.1); här skall endast påpekas att motsättningar ibland kunde uppkomma mellan å ena sidan läsbarhetskravet, skärpt till ett krav på stilistisk smidighet eller elegans, och å andra sidan entydighetskravet, i kombination med det övergripande kravet på analyticitet. Vissa definitioner är därför alltför kortfattat och elliptiskt formulerade, något som intressant nog ofta visade sig vid taggningen av deras semantiska struktur i fortsättningsprojektet.

- Jönberg, J. Arbetsrapport för Lexikalisk databas. Arbetsrapport. Göteborg, Språkdata 1976
- Jönberg, J. A Semantic Model for a Performance Framework. Studies in the Possibilities of Formalization. Diss. Göteborg 1984
- Jönberg, J. Towards a Formalized Lexicon of Swedish. I Studies in Computer-Aided Lexicology. Stockholm 1988
- Jönberg, J. Kortel för semantisk märkning av definitioner. Under arbete
- Kangroo Dictionary of Contemporary English. London 1978
- Malgören, E.-G. On Regular Polynomy in Swedish. I Studies in Computer-Aided Lexicology. Stockholm 1988
- Malgören, E.-G. & Norén, K. Ambiguitetstyper i svenskan. Göteborg, Språkdata 1981
- Norén, K. Svenska partikelverb. Diss. under arbete
- Ohlander, S. "Ord, ord, hemskt mycke ord." Kring en svensk läro- och läroordbok. I Moderna språk LXXXI, nr 3. Stockholm 1982
- Ordlexikon över svenska språket. Utg. av Svenska Akademin. Lund 1896
- Ralph, B. Slutsrapport över projektet Svensk datoriserad lexikon. Göteborg, Språkdata 1978

Preliminär bibliografi

- Abelin, Å. Patterns of Synaesthesia in the Swedish Vocabulary. I Studies in Computer-Aided Lexicology. Stockholm 1988
- Allén, S. Förhållandet mellan skrift och tal. I Språk, språkvård och kommunikation. Red. K.-H. Dahlstedt & B. Molde. Stockholm-Uppsala 1967
- Allén, S. Nusvensk frekvensordbok baserad på tidningstext. 1. Graford, homografkomponenter (NFO 1). Stockholm 1970
- Allén, S. Nusvensk frekvensordbok baserad på tidningstext. 2. Lemman (NFO 2). Stockholm 1971
- Allén, S. et al. Nusvensk frekvensordbok baserad på tidningstext. 4. Ordled, betydelser (NFO 4). Stockholm 1980
- Allén, S. & Järborg, J. Semantisk strukturering av Lexikalisk databas. Slutredogörelse. Göteborg, Språkdata 1988
- Bonniers svenska ordbok. Red. S. Malmström & I. Györki. Första uppl. Stockholm 1980
- Illustrerad svensk ordbok. Red. B. Molde. Tredje uppl. Stockholm 1977
- Järborg, J. Arbetsfiler för Lexikalisk databas. Arbetsrapport. Göteborg, Språkdata 1978
- Järborg, J. A Semantic Model for a Performance Framework. Studies in the Possibilities of Formalization. Diss. Göteborg 1984
- Järborg, J. Towards a Formalized Lexicon of Swedish. I Studies in Computer-Aided Lexicology. Stockholm 1988
- Järborg, J. Manual för semantisk märkning av definitioner. Under arbete
- Longman Dictionary of Contemporary English. London 1978
- Malmgren, S.-G. On Regular Polysemy in Swedish. I Studies in Computer-Aided Lexicology. Stockholm 1988
- Malmgren, S.-G. & Norén, K. Ambiguitetstyper i svenskan. Göteborg, Språkdata 1982
- Norén, K. Svenska partikelverb. Diss. under arbete
- Ohlander, S. "Ord, ord, hemskt mycke ord." Kring en svensk definitionsordbok. I Moderna språk LXXXI, nr 3. Stockholm 1987
- Ordbok över svenska språket. Utg. av Svenska Akademien. (SAOB). Lund 1896 -
- Ralph, B. Slutrapport över projektet Svensk definitionsvokabulär. Göteborg, Språkdata 1978

- Ralph, B., Järborg, J. & Allén, S. Svensk ordbok och Lexikalisk databas. Förstudierapport. Göteborg, Språkdata 1977
- Sjögreen, C. Creating a Dictionary from a Lexical Database. I Studies in Computer-Aided Lexicology. Stockholm 1988
- Svensk handordbok. Konstruktioner och fraseologi. Red. T. Johannisson & K. G. Ljunggren. Stockholm 1966
- Svensk ordbok. Stockholm 1986
- Svensk-serbokroatiskt lexikon. Stockholm 1985
- Svenska Akademiens ordlista över svenska språket. Tionde upplagan. (SAOL₁₀). Stockholm 1973
- Östergren, O. Nusvensk ordbok. Stockholm 1919 - 1971

