

Lennart Lönngren, Uppsala

#### LEXIKA, BASERADE PÅ SEMANTISKA RELATIONER

Den första fråga man måste ta ställning till om man vill bygga upp en tesaurus, alltså ett lexikon baserat på semantiska relationer, är om man skall tillämpa någon form av hierarkisering och hur i så fall denna skall se ut. I princip vill jag förkasta tanken på att begreppen som sådana kan ordnas hierarkiskt; jag tror alltså inte på några universella eller särspråkliga semantiska primitiver à la Wierzbicka (1972). Normalt är det nog att konstatera att det föreligger ett associativt samband mellan två begrepp, t.ex. *tand* och *bita*, samt att fastställa styrkan hos och arten av detta samband utan att postulera något riktningsförhållande.

Det är emellertid praktiskt att organisera ett tesauruslexikon hierarkiskt. Det innebär en förenkling så till vida att man ersätter en mångfald av relationer med i princip en enda, dependens. Jag tänker mig här en mer djup- och genomgående hierarkisering än den vi finner i Rogets lexikon (1962, första gången utgivet 1852) och dess svenska efterbildning, Bring (1930), där man definierat ett begränsat antal "begreppsklasser" och hänfört varje ord till en sådan. Frågan är bara om detta är möjligt, alltså om en sådan hierarkisering står i samklang med ords kattens inre natur. För att citera Kassabov (1987, 51) gäller det här att undvika det allmänna misstaget att "attempt to prove the systematic character of vocabulary not by establishing the inherent principles of its inner organization, but by forcing upon the lexical items the networks of pre-formulated systems".

Om vi börjar med den del av ordförrådet som är sammanbunden morfologiskt-derivationellt, alltså det subsystem som omfattas av ordbildningsrelationer, så är det ju helt uppenbart att språktecknens uttryckssida kännetecknas av skiftande komplexitet och att man kan se ett slags vertikal relation, alltså riktnings- eller dependensförhållande, mellan lexem som *sand* och *sandig*, där det ena är avlett av det andra. Att organisera ett språks hela ordförråd i hierarkiska ordbildningsnästen är dock ingen lätt sak. En orsak till detta är att lexemens betydelser långt ifrån alltid återspeglas av ordbildningsstrukturen, en annan att det finns så många fall av identisk komplexitet, där leden trots detta måste ordnas i vertikal led, därför att det inte finns något mindre komplext lexem i ordbildningsnästet i fråga, t.ex. *såg* - *såga*, *ljus* - *lyså*.

I de fall uttryckssidans komplexitetsförhållanden förefaller klara har man i allmänhet antagit att det finns en direkt komplexitetsmotsvarighet på innehållssidan, alltså att i vårt exempel *sandig* är även semantiskt mer komplext än *sand* och innehåller komponenten 'sand' på samma sätt som strängen *sandig* innehåller strängen *sand*. Detta resultat kommer man ju också till om man förutsätter att vi har att göra med tre språktecken (förutom flexionsmorfemet) - vardera med en uttrycks- och en innehållssida - nämligen de elementära *sand*- och *-ig* samt det komplexa *sandig*. Men då får man problem i sådana fall där den semantiska relationen förefaller identisk men där den formella komplexiteten pekar åt olika håll, t.ex. i paren *undersöka* - *undersökning* och *analys* - *analysera*, *fysik* - *fysiker* och *biolog* - *biologi*. Vill man fastställa språkoberoende semantiska hierarkiska förhållanden får man ytterligare problem i fall där nära semantiska ekvivalenter i två språk uppvisar olika riktningsförhållanden med avseende på formell komplexitet, jfr sv. *seger* - *segra* -

segrare men tj. vítěz ('segrare') - vítězit ('segra') - vítězství ('seger'). Det förefaller ohållbart att försöka lösa alla dessa problem genom att laborera med t.ex. nollsuffix, tomma suffix och tilläggskriterier av mer eller mindre semantisk natur. En språkoberoende hierarkisering försvåras naturligtvis också av att det ofta förekommer att två begrepp är sammankopplade ordbildningsmässigt i ett språk men inte i ett annat, t.ex. nyckel, som är (synkroniskt) oderiverat i svenska och engelska, men i tyska och finska deriverat av 'låsa' (Schlüssel) resp. 'öppna' (avain); vidare är det vanligt att en och samma företeelse har olika derivationell anknytning, jfr t.ex. 'näsduk', som i svenska och ryska är förknippat med 'näsa', i tjeckiska, tyska och danska med 'ficka', i engelska med 'hand' etc.

Det är ändå viktigt att framhålla, att problemen visserligen är många nog att förbittra tillvaron för den som vill sammanställa ett hierarkiskt derivationslexikon, men ändå inte så många att de fördunklar det faktum att det finns en tendens, särskilt inom ramen för ett språkssystem men även universellt: formell komplexitet tycks harmoniera med vissa andra egenskaper, framför allt med allmän frekvens (dvs hur ofta använt och därmed hur väl känt ett ord är för den genomsnittlige språkbäraren). Det är inte bara så att i varje frekvensordlista ordlängd och frekvens är i stort sett omvänt proportionella utan det förhåller sig också så att av två ord som är direkt "vertikalt" förbundna med varandra i en ordbildningsrelation är det enklare, alltså dominerande, lexemet oftast mer frekvent än det deriverade.

Mot denna bakgrund förefaller det ovannämnda företaget inte helt utsiktslöst. Det är emellertid, som alltid, viktigt att börja med tydliga fall och med dessa som mönster fortskrida till de mindre tydliga. En lämplig utgångspunkt är reguljära sammansättningar av typen metallindustri där hela ordet utgör en hyponym till huvudledet industri; sammansättningen är också mindre frekvent än huvudledet. Man kan alltså säga att dessa egenskaper klart harmonierar med den derivationella strukturen.

Sådana idealiska förhållanden råder långt ifrån alltid i sammansättningar. Dessutom måste vi kunna hantera andra typer av morfologiska samband samt fall där derivationellt samband saknas, t.ex. i par av typen tand : bita. Vi måste också ofta acceptera att ett dominerande lexem inte alltid kan vara en hyperonym. För det första är det svårt att hitta hyperonymer till många lexem, t.ex. vanlig, skratta, dimma, för det andra är hyperonymen inte alltid den starkaste, mest markanta relationspartnern. Det ger t.ex. inte så mycket att ge del som dominerande lexem till kroppsdel och världsdel; här är det i stället bättre att ge 'helheten', dvs kropp resp värld. Likaså är det troligen bättre att koppla samman järnväg med tåg än att låta det domineras av lexemet väg; och kopplingen till metallen järn är klart perifer, vilket visar att man inte alltid kan låta sig vägledas av ordbildningsstrukturen; ett annat tydligt exempel på detta är skärgård, som inte bör analyseras i sina sammansättningsled utan kopplas ihop med lexemet ö, jfr här t.ex. finskans saaristo. Genom att på detta sätt frigöra oss från ordens inre form tar vi ett steg på vägen mot en universell begreppshierarkisering.

Vad vi främst behöver är ett oberoende semantiskt kriterium som är mer generellt än "hyponymi-hyperonymi"-relationen, en asymmetri som kan läggas till grund för ett dependensförhållande. Det finns faktiskt en sådan asymmetri, och jag skulle vilja kalla den "lexikalisk determinism". Den består i att man givet en semantisk relation och en relationspartner A med större entydighet kan bestämma relationspartnern B än omvänt, dvs än vad som blir fallet om man

utgår från B och söker A. De tydligaste exemplen återfinns vi inom hyponymi-hyperonymi-relationen samt inom del-helhetsrelationen. Sålunda kan vi gå entydigt från bil till fordon, från röd till färg, från styre till cykel, från nav till hjul osv., men inte omvänt. Från dessa enkla fall kan vi gå vidare till andra semantiska relationer. Det finns också någonting som vi kan kalla "situationell determinism", nämligen i fall som snyta ← näsa (det semantiska objektet för 'snyta' är alltid 'näsa', medan däremot 'näsa' kan förekomma i många andra situationer än 'snyta'; ett likartat förhållande råder mellan regna och vatten. Även inom ett ordbildningspar som sten → stena finner vi sådan assymmetri. Naturligtvis finns det många nära relaterade ordpar mellan vilka determinism saknas, t.ex. cykel och hjul: hjul förekommer ju inte endast på cyklar och cyklar har ju andra delar vid sidan av hjulen (även om hjulen är mycket väsentliga, vilket reflekteras av att benämningarna på cykel i många språk är deriverade av hjul). Ett sådant faktum som att alla hjul inte har ekrar behöver dock inte sänka determinismen mellan hjul och ekrar, eftersom ekrar ändå bara förekommer på hjul.

Jag har alltså accepterat ett dependensbegrepp som utgör en sammanvägning av flera olika faktorer, vilka på det hela taget samverkar, nämligen: a) formell komplexitet, t.ex. sand → sandig; b) frekvens, t.ex. annars → eljest; c) stilistisk markering (dvs ett stilistiskt neutralt lexem betraktas som dominerande), t.ex. rolig → kul, båda → bägge; d) determinism, t.ex. näsa → snyta, hjul → nav. Den bästa sammanfattande benämningen på detta dependensbegrepp torde vara "enkelhet", inbegripande såväl uttrycks- som innehållssida: av två lexem, mellan vilka man vill etablera ett direkt vertikalt samband, väljs det enklare som det dominerande.

Nu kan man invända: att återföra ett ord på enklare begrepp är inte något nytt. Det är vad man alltid har gjort i ordboksdefinitioner. Jag skulle vilja säga, att det är vad man tror sig ha gjort. I själva verket vimlar det av exempel på motsatsen. Särskilt vanligt är att mycket enkla begrepp definieras med hjälp av mer komplexa (mindre kända). Ett belysande exempel är Ralph (1979), där enkla verb av typen jubla jämförs med parafraser av typen uttrycka jubel. Andra exempel är förvärva fångst (= fånga), konsumera dryck (= dricka). Det är svårt att hålla med författaren: "Varje [i parafrasen] ingående ord blir följaktligen semantiskt mindre komplext än utgångsverbet." (s. 164). I jakten efter mer generella begrepp tvingas man ofta söka sig till språkets periferi, vilket f.ö. också Brings tesaurus är ett gott exempel på; här finner vi begreppsklasser som "Omängdhet" och "Inrymningsplats". Det är under sådana förhållanden inte underligt att de ofta påtalade cirkeldefinitionerna i enspråkiga lexika uppstår. Det är enligt min mening helt omöjligt att definiera varje lexem, särskilt lexemen i de övre frekvensområdena, med hjälp av (genuint) enklare ord. I den hierarkiska struktur jag här beskriver är det ju inte heller fråga om att definiera lexemen.

Med utgångspunkten i deterministiska sammansättningar bör man vid sidan av ett huvudled kunna acceptera även ett modifierande led som dominerande lexem, även om detta naturligtvis komplicerar lexikonets struktur. Vid ett lexem som metallindustri ansätts alltså två dominerande lexem, nämligen "föräldraparet" industri och metall; härvid kan vi kalla huvudkomponenten industri "moder" och den modifierande komponenten metall "fader" (komponenterna ges också alltid i denna ordning). Det är viktigt att observera att de ovan nämnda "harmonierande" egenskaperna vad gäller frekvens och determinism huvudsakligen endast gäller för modersdominanten, ej för fadersdomi-

nanten; denna utgör ju inte heller en hyperonym till derivatet.

För att fortsätta analogin med strukturen hos en familj är det endast vissa lexem som får ett komplett föräldrapar, alltså två domineranter; vid bildningar medelst affix är det mestadels tillräckligt att ange endast en förälder, t ex *sandig* → *sand*. När den morfologiska kopplingen saknas kan vi mera fritt välja mellan att ge en eller två domineranter; sålunda kan man tveka om man skall låta öga definieras enbart genom *se* eller genom *se plus ansikte*. Liksom det inom ordbildningen förekommer kopulativa sammansättningar, av typen *röd-gul*, måste man även utanför den morfologiska sfären ibland ge föräldrapar som är kopulativt sammanbundna, t.ex. *ranglig* ← *lång* + *ma-ger*.

Ett dominerande lexem kan naturligtvis ha många barn, i förhållande till vilket det kan uppträda som moder eller fader eller ensamförälder. Mera ovanligt är att ett och samma föräldrapar har flera barn, t ex *veta* + *vilja* → *intresse*, *undersöka*, *nyfiken*...

För att pröva om dessa tankar är realiserbara har jag hittills dels bearbetat en mindre korpus på 17.000 löpord och ur den skapat ett litet lexikon, dels kompletterat detta med ord ur allmänna frekvenslistor samt lagt till många ord för att knyta ihop dependenskedjorna, så att jag f.n. har en fil med drygt 6.000 komplexa "lemman" av följande utseende:

```
metallindustri ← [industri + metall]
plastindustri ← [industri + plast]
industriarbetare ← [arbetare + industri]
industri ← tillverka
tillverka ← göra
göra ← PRIM
```

Dvs de har alla strukturen  $A \leftarrow [B (+ C)]$ .

Den typ av lemman som jag skapar skiljer sig från de traditionella inte bara i det avseendet att de är komplexa, alltså ett slags "superlemman", om man så vill, utan också därigenom att polysemi beaktas. Ett morfologiskt lexem klyvs i två semantiska, om de har helt eller partiellt olika domineranter, t.ex. *byte* → *byta*, *bytel* - *fånga*; *bruka* → *vanlig*, *bruka1* → *använda*, *bruka2* → *odla*. I undantagsfall kan två lexem med identisk enda förälder klyvas, under förutsättning att "barnaskarorna" är klart åtskilda, t.ex. *växt* (→ *blad*, *blomma*) och *växt1* (→ *missväxt*), vilka båda återförs på *växa*.

Alla dependenskedjor avslutas i toppen med ett fiktivt lexem, PRIM. Jag har hittills definierat ca 50 lexem, som domineras enbart av PRIM (se Appendix). I och med att jag har arbetat igenom ett grundlexikon har till stor del lemmatiseringsprinciperna utkristalliserats. Eftersom även de mest frekventa orden, där ordbildningsrelationerna inte i så hög grad bestämmer lexemens betydelser, finns medtagna, kan man säga att den svåraste delen av arbetet är gjord.

Ur denna fil kan man - lämpligen automatiskt, alltså medelst ett dataprogram - generera ett ganska stort antal lexempar. Härvid tillkommer, förutom den direkt givna vertikala föräldrar-barn-relationen, två horisontella relationer, äkta-make-relation och syskon-relation, alltså:

- a)  $A \leftarrow B$ ; dvs A domineras av B (föräldrar-barn-relation)
- b)  $B + C$ ; dvs B kan sättas samman med C ("äkt-make-relation")
- c)  $A \# D$ ; dvs A och D har (minst) en gemensam dominant ("syskonrelation")

Några exempel på lexempar som genereras ur de ovan anförda komplexa lemmarna:

- a) metallindustri ← industri, metallindustri ← metall;  
industri ← tillverka
- b) industri + metall; arbetare + industri
- d) metallindustri # plastindustri

Det genererade materialet kan listas på olika sätt, men ett överskådligt sätt är att ge samtliga ingående lexem i bokstavsordning och till varje lexem ge varje tänkbar direkt relationspartner, med relationen specificerad enligt nyssnämnda klassifikation. Den i sammanhanget centrala a-relationen, som definierar själva dependensen, har jag för listningsändamål hittills valt att klyva i två varianter, en med dominerat led på första plats, såsom i utgångslemmat (←), och en omvändning (→). Jag får därmed ordboksentryn av följande utseende:

industri      ← tillverka  
                 → industriell, plastindustri, industriarbetare...  
                 + metall, plast, verkstad, arbetare...  
                 # fabrik, producera...[← tillverka]

Det är emellertid klart att det behövs ytterligare specificeringar. Inom den vertikala relationen visar redan pilen dependensens riktning, men dessutom behöver anges om det är fråga om dominans från moder, fader eller ensamförälder. Inom äkta-make-relationen behöver anges riktningen i det fall då vi har att göra med huvudled och determinand (vi behöver t.ex. kunna skilja på *avfallsindustri* och *industriavfall*). Inom syskonrelationen, slutligen, behöver vi skilja på om lexemen har gemensam moder, gemensam fader, båda gemensamma, eller om det föreligger en mer avlägsen relation, som i paret *metallindustri # industriarbetare*, där dominanten är moder i ena fallet och fader i det andra (könsbyte!). Särskilt nära relaterade är fall med gemensam moder. Man bör observera att syskonrelationen, till skillnad från de övriga, är potentiellt mångställig, dvs att det här är mindre relevant att tala om lexempar. Vi får snarare serier av typen *kaffe, te, vin, öl...* (sammanhållna av hyperonymen *dryck*). Det visar sig för övrigt att just den intuitivt kända graden av samhörighet inom syskongrupperna utgör en god kontroll på att kopplingarna blivit riktiga gjorda vid lemmatiseringen.

Här ges ytterligare ett exempel, med dels flera icke-derivationella samband, dels mer specificerade relationer:

ved      ←1 bränsle; ←2 träd  
            →1 björkved; →2 vedbod  
            +1 björk; +2 bod  
            #1 bensin, kol, olja [gemensam "mor"]; #2 bark, dunge,  
            trädgren [gemensam "far"]; #3 al, asp, björk; bränsle-  
            tank, bränslecell [gemensam "mor/far"]

Om man listar materialet på så sätt att varje lexem får utgöra ett eget entry, som ovan föreslogs, får man problem framför allt med redundans inom syskonrelationen. Om något lexem dominerar ett stort antal andra lexem (t.ex. *inte*, som är fader till alla lexem med en negativ komponent, såsom adjektiv av typen *ledsen, omöjlig*), ger detta faktum upphov till ett mycket stort antal symmetriska lexempar av typen *ledsen # omöjlig*. Här bör man alltså lägga in något slags

begränsning i programmet, t.ex. att man avbryter listningen efter den tionde partnern och hänvisar till det gemensamma dominerande lexemets entry.

Jag skall nu gå vidare i arbetet och enligt dessa principer lemmatisera en ordformsfil som genererats ur en textkorpus som består av artiklar ur tidskriften *Forskning och Framsteg* och som omfattar nära 100.000 löpord. Det är enligt min åsikt flera fördelar med att arbeta utifrån en korpus, snarare än utifrån ett traditionellt lexikon. Ett av de avgörande skälen är att man lätt kan inkorporera även encyklopedisk kunskap, något som kan vara värdefullt bl.a. vid informationssökning. Denna kunskap läggs huvudsakligen in via egennamnen, framför allt givetvis sådana egennamn som inherent har en specifik referens, t.ex. geografiska namn. Följande exempel visar vilka lemmatiseringsprinciper jag här har tillämpat: Olof ← förnamn, Fido ← namn + hund, Olof ← Palme, Palme ← politiker, Einstein ← fysiker, Malmö ← stad, Madrid ← huvudstad + Spanien, Vänern ← sjö, Ladoga ← sjö + Sovjetunionen etc. Exemplet *Olof Palme* visar att dependensrelationen även kan användas för att binda samman delar av fasta syntagmer, i de fall då man fortfarande vill ha de ingående komponenterna åtkomliga.

\* \* \*

En annan typ av lexikon skall jag här skissera mycket kort; jag befinner mig ännu endast i början av materialinsamlingen. Detta lexikon bygger på hypotesen, alternativt postulatet, att varje lexem omger sig med ett kognitivt nätverk, där det självt upptar en nod. Uppgiften består i att inventera dess "nära nodgrannar", alltså de lexem som det har en direkt relation till, och att specificera denna relation i semantiska, kvalitativa termer. Detta förutsätter att vi har bestämt oss för vissa principer för att representera kunskap. Mitt förslag är att representera kunskap med naturliga lexem i kombination med en uppsättning logisk-semantiska valensramar. Jag arbetar sedan några år tillbaka med följande ramar:

A		KO		
B	KS	KO		
C		KO	KO'	
D	KS	KO	KO'	
E		KO	KO <sup>-</sup>	KO <sup>+</sup>
F	KS	KO	KO <sup>-</sup>	KO <sup>+</sup>
G		KO	KL	
H	KS	KO	KL	
I		KO	KL <sup>-</sup>	KL <sup>+</sup>
J	KS	KO	KL <sup>-</sup>	KL <sup>+</sup>

Systemet gör först och främst en distinktion mellan objekt (T) och situationer (P), härövan generaliserat till K ("kategori"). Som synes bestäms konfigurationerna av dels antalet aktanter, dels dessa aktanters semantiska roller. Denna klassifikation avspeglar grundläggande logiskt-semantiska egenskaper som terminativ, agentiv, lokalistisk m.m. Jag använder mig av tre grundläggande semantiska roller, S, O och L; i terminativa situationer, alltså sådana som inne-

bär egenskapsförändring, förvandling eller förflyttning klyvs dock O resp L i två faser, markerade med + resp -. Roller kan vidare sammanlösas och redupliceras, men jag skall inte gå in på detta här.

Det nätverk som sålunda kan byggas upp runt varje lexem har en nära motsvarighet i de s.k. "lexikaliska funktioner" som har använts av Mel'čuk i hans språkmodell "meaning-text" och som har en framträdande plats i det lexikon - "Explanatory Combinatorial Dictionary" (1984) - som han och hans medarbetare har sammanställt. Exempel på dessa lexikaliska funktioner är: *Magn*(sjuk) = svårt; *Operl*(inflytande) = utöva; *Caus<sub>loc</sub>*(fartyg) = varv. Tanken var ursprungligen att man med hjälp av dessa funktioner skulle reducera antalet enheter på en mer generell, "djupare" representationsnivå, t.ex. ett interlingua för automatisk översättning.

Funktionerna är meningsfulla att använda endast när värdet på dem vid ett givet argument är åtminstone någorlunda bestämt, entydigt. Man kan därmed säga att de definierar den lexikaliska determinismen i ett givet språk, inbegripande dels icke förutsebar sarspråklig information, dels kunskap om världen, eller i varje fall kunskap om hur det typiskt ser ut i världen, det som i olika sammanhang kallats "frames", "scripts", etc.

Att representera dessa funktioner genom hänvisning till olika positioner i ett kognitivt nätverk erbjuder, enligt min uppfattning, ett enklare och mer generellt system än det mel'čukianska. Gentemot det hierarkiskt uppbyggda system, som jag nyss har presenterat, har det den fördelen att man kan redovisa alla relationspartner till ett givet lexem; man behöver med andra ord inte välja mellan t.ex. 'näsa' och 'ficka' som dominant vid lexemet *näsduk* utan kan redovisa båda, med angivande av dels arten av semantisk relation, dels determinism och eventuellt relevans:

B/J:	$TO_b \leftarrow TL_j^-$	<i>näsduk</i> - <i>näsa</i> (B = använda, J = snyta)
H:	$TO \leftarrow TL$	<i>näsduk</i> - <i>ficka</i> (H = förvara)
C:	$TO \leftarrow TO'$	<i>näsduk</i> - <i>duk</i> ; (C = "kind-of")

etc., där / anger att två situationer är involverade och pilen markerar determinismens riktning; i bokstavskombinationen TO, TL etc. betecknar den första kategori, den andra roll.

Om man i stället för ett ensamt lexem låter ett lexempar utgöra argument i funktionen ökar determinismen i nätverket avsevärt. Här kan vi också få in associationer av typen *elefant* : *grå*, *brandpost* : *röd* (se Charniak 1983, 442 ff), jfr även *häst* : *havre*, *kanin* : *morot*, alltså sådana för vilka man ofta använt en "attribute-value-struktur", se bl.a. Woods (1975, 50 ff). Dessa missar man ofta i den enklare hierarkiska strukturen. Å andra sidan har attribute-value-strukturen en begränsad tillämpning: vilket "attribut" skall man t.ex. ansätta vid relationen *damm* (= eng. *dust*) : *torr*? Endast relativt sällan är leden i ett givet lexempar förenade på mer än ett sätt, dvs så att det i en tredje nod tillåts mer än ett lexikaliskt värde, jfr:

B:	$TS \leftrightarrow TO$	<i>bonde</i> - <i>åker</i> (B = bruka)
J:	$TS \leftrightarrow TL^+$	<i>bonde</i> - <i>åker</i> (J = så, gödsla)

Många rollkonfigurationer tillåter olika kategoribeteckningar, jfr:

G:	TO→TL	blomma - vas
G:	PO↔TL	sova - säng
G:	PO←PL	svettas - arbeta

Syftet med båda typerna av ordböcker är att lämna ett bidrag till det nödvändiga arbetet med att inventera de otaliga intersubjektiva associationer, vilka utgör en förutsättning för språkförståelsen. Dessutom finns mer konkreta tillämpningar, såsom en semantiskt baserad informationssökning och automatisk textindexering, där en inventering av detta slag kan visa sig mycket användbar.

### Referenser

- Bring, S. C., *Svenskt ordförråd ordnat i begreppsklasser*, Uppsala 1930.
- Charniak, E., *Context Recognition in Language Comprehension, Strategies for Natural Language Processing*, London etc. 1982, s. 435-454.
- Kassabov, I., *On the Problem of Defining the Core of the Vocabulary of the Bulgarian Language*, *Linguistique Balkanique* 30:1 (1987), 51-55.
- Mel'čuk, I. A. & Žolkovskij, A. K., *Tolkovo-kombinatornyj slovar' sovremennogo russkogo jazyka*, Wien 1984. (= Wiener Slavistischer Almanach, Sonderband 14).
- Ralph, B., *Primitiva verb i svenskan, Förhandlingar vid sammankomst för att dryfta svenskans beskrivning* 11, Stockholm 1979, 163-174.
- Roget's Thesaurus of English Words and Phrases*, London 1962 (Longmans).
- Wierzbicka, A., *Semantic Primitives*, Frankfurt/M 1972.
- Woods, W. A., *What's in a Link: Foundations for Semantic Networks, Language, Thought, and Culture*, New York etc. 1975, s. 35-82.

### Appendix: Lexem dominerade av PRIM

all	före	ljud	och	vad
annan	göra	ljus (subst)	om	var
använda	ha	med	på	vara
bara	hur	men	rak	varm
bra	hända	mot	röra	vem
den (art)	i	mycken	så	veta
fort	inte	måste	säga	vid
framme	just	namn	tal	vilja
färg	kant	natur	till	än
för	kunna	när	tänka	