

SALDO

Beskrivningen av SALDO finner du på adressen

<http://spraakbanken.gu.se/saldo>

För att söka i SALDO går du härifrån till "Testa SALDO" och klickar därefter på "Sök i SALDO", eller också kan du gå direkt till den här adressen:

<http://spraakbanken.gu.se/ws/saldo-ws/fl/html>

En tom ruta kommer upp och du uppmanas att skriva in ett ord – inte nödvändigtvis en lexikonform, det kan också vara en böjld form.

Hur det går till att söka i SALDO kan enklast visas med några konkreta exempel.

EXEMPEL

Exempel 1

1. Skriv in ordet *stenyxa*.

2. Du får upp det inskrivna nyckelordet i kraftig stil, en andra rad med ordet "sammansättningsanalys", samt därunder en rad, som vi kallar presentationsraden. Denna är uppdelad i fyra fält. I det första fältet finner du lexemet *stenyxa*, dvs ordet sett som en betydelseenhet. I det tredje fältet finner du lemmat *stenyxa*, dvs ordet sett som en formenhet. Här är den viktigaste egenskapen ordklassstillhörighet, vilken också anges inom parentes direkt efter ordet; *mn* betyder substantiv.

Fältet mellan dessa båda innehåller lexemets semantiska deskriptorer. Den första, som vi kallar primär, är obligatorisk. Det är det ord i språket som bäst svarar mot två krav: att vara enklare än nyckelordet och att vara en nära semantisk "granne" till nyckelordet. Enkelhetskriteriet är sammansatt av flera kriterier: bland annat är ett ord A enklare än ett ord B om A är mera frekvent (och därmed mer känt, mer använt) än B, om A är mer stilistiskt neutralt än B (detta gäller framför allt synonymer) och om A har en morfologiskt enklare struktur än B (detta gäller framför allt ordbildningsrelationen). Det finns också rent semantiska relationer, nämligen de hierarkiska, där detta kriterium kan tillämpas. Sålunda är ett överordnat begrepp enklare än ett underordnat och benämning på en helhet enklare än benämningen på en del.

I detta fall är den primära deskriptorn *yxa*, vilket också är hyperonymen till *stenyxa*, men detta är förstås bara en av möjliga semantiska relationer mellan ett lexem och dess primära deskriptor.

Efter den primära deskriptorn och åtskilt med ett plustecken följer i detta exempel ytterligare en deskriptor, vilken vi kallar sekundär. Denna deskriptor är det ord som, när väl huvuddeskriptorn är utsedd, på bästa sätt ytterligare preciserar (explicerar) nyckelordets betydelse. Den sekundära deskriptorn är inte obligatorisk; den saknas hos drygt hälften av alla lexem i SALDO. Nedan, under exempel 5, 6 och 7, kommer du att se att vi nu har öppnat för möjligheten att ange mer än en sekundär deskriptor.

Det fjärde fältet i presentationsraden, "korpus", ger möjlighet att se efter hur ordet (egentligen lemmat) är representerat i de korpuser språkbanken förfogar över.

3. Klicka nu på lemmat, dvs ordet i det tredje fältet.
4. Du får upp morfologisk information: grundform, böjningsmönster, ordklassstillhörighet samt i vissa fall, såsom här, inherent drag. Det viktigaste inherent draget är substantivets genus ("u" betyder utrum). Dessutom ges en böjningstabell med alla de former substantivet *stenyxa* kan anta, inklusive den form ordet antar som första led i en sammansättning.
5. Klicka på ordet "sammansättningsanalys".
6. Du får en automatisk analys av ordets beståndsdelar. Denna analys är inte nödvändigtvis "korrekt". I detta fall hittar analysprogrammet som förled inte bara substantivet *sten* utan också verbet *stena*.
7. Klicka på lexemet *stenyxa*, dvs ordet i det första fältet.
8. Du får upp lexemet tryckt i kraftig stil. Därunder ges inom en ram lexemets deskriptorer i var sitt fält, nu markerade med pilar: den primära deskriptorn *yx*a med dubbel uppåtpil, den sekundära *sten* med enkel uppåtpil. (Utformningen av pilarna växlar beroende på webbläsare; i Firefox kommer de bra fram.) Pilarnas riktning anger att deskriptorerna är överordnade i förhållande till nyckelordet (lexemet). Samtliga i lexikonet inlagda lexem utgör nämligen ett sammanhängande, hierarkiskt ordnat nätverk.

Intill den dubbla uppåtpilen står en siffra inom klammer. Denna anger på vilket "djup" ordet *stenyxa* befinner sig: genom att gå via första deskriptorn kommer man nämligen till nätverkets topp med sex steg.

På raden under deskriptorramen återfinns du lemmat, vilket även här, liksom i presentationsraden, är klickbart. I vissa fall ges på denna rad mer utförlig morfologisk information, bl.a. stavningsvarianter. Om du t.ex. har skrivit in *symtom* får du i presentationsradens lemmafält fram bara den stavningsvarianten, medan du här, under deskriptorramen, får fram även en alternativ stavning: *symptom*.

Under lemmaraden finns en "ordros". Klicka på den (fungerar tyvärr f.n. inte på Explorer). När den öppnar sig visar sig i centrum ordet *stenyxa*. Intill ligger en liten blå ring. Från denna går linjer till mer perifert placerade ringar, intill vilka man ser orden *båtyxa*, *flintyxa* och *trindyxa*. Klicka nu på ordet *stenyxa* eller på dess ring (effekten blir densamma). En ny ordros kommer fram; i centrum av denna står första deskriptorn till *stenyxa*, alltså *yx*a. Från tillhörande ring går ett ganska stort antal linjer ut till ord i periferin, bland vilka du även finner *stenyxa*. Om du fortsätter på samma sätt kommer du successivt till orden *hugga*, *slå*, *mot*, *med* och *ha*. När du klickar på *ha* kommer du till SALDO:s topp, det konstgjorda ordet PRIM. Avståndet från *stenyxa* till PRIM är sju steg, dvs *stenyxa* har djupet sju. Medeldjupet för alla lexem i SALDO ligger på sex (nästan exakt). Ytterst få lexem har ett djup överstigande 10. Maximidjupet är f.n. 15.

Under ordrosen står återigen ordet "korpus", dvs det finns även här möjlighet att kontrollera beläggen för lemmat i fråga.

Resten av sidan upptar de lexem som har nyckelordet som deskriptor. Eftersom det alltså är frågan om underordnade lexem anges de med nedåtriktade pilar. Dubbel nedåtpil anger de ord som har nyckelordet som primär deskriptor, enkel nedåtpil de lexem som har nyckelordet som sekundär deskriptor. Båda dessa klasser av lexem kan

vara tomma eller ha ett i princip godtyckligt antal medlemmar. Därför anges här antalet under pilen. Om antalet är noll markeras detta enbart med en asterisk i lexemrutan.

I vårt exempel finns i den klass som har *stenyxa* som primär deskriptor lexemen *båtyxa* med den sekundära deskriptorn *båtliknande*, *flintyxa* med den sekundära deskriptorn *flinta* och *trindyxa* med den sekundära deskriptorn *rundad*. (Just dessa tre ord, *båtyxa*, *flintyxa* och *trindyxa*, fanns ju också med i ordrosen för ordet *stenyxa*.) Observera att den sekundära deskriptorn föregår det underordnade lexemet och ges i svart, dvs oklickbar, form. Nästa fält, markerat med enkel nedåtpil, skall innehålla den klass av underordnade lexem som har *stenyxa* som sekundär deskriptor. Denna klass är som synes tills vidare tom. Detta är inte nödvändigtvis ett slutgiltigt förhållande. SALDO fylls kontinuerligt på med nya ord, och man kan tänka sig att här sätta in t.ex. ordet *stenyxkultur*.

9. Gå nu upp till deskriptorramen och klicka på den primära deskriptorn *yxa* (markerad med dubbel uppåtpil).

10. Du får upp sidan för detta lexem. Som sekundär deskriptor anges PRIM. Detta är ett konstgjort noll-lexem, som sätts in i de fall när det inte finns något riktigt lexem som fyller platsen för en deskriptor. Det är helt enkelt en markör för att nätverket inte har någon fortsättning uppåt. Ordet *yxa* har alltså egentligen bara en primär deskriptor. Ungefär 40 procent av alla ord i SALDO har PRIM som sekundär deskriptor. Ett betydligt mindre antal, 43, har PRIM som både primär och sekundär deskriptor. Dessa lexem är inbördes semantiskt obesläktade. Om du vill veta vilka dessa "ultimata" ord är kan du någonstans klicka på deskriptorn PRIM.

Bland underordnade lexem med *yxa* som primär deskriptor finner du först sådana som formellt har PRIM som sekundär deskriptor och därunder sådana som har ett riktigt lexem som sekundär deskriptor. Dessa senare är sorterade efter vänsterkolumnen, dvs efter sekundärdeskriptorn. Här återfinner du bl.a. lexemet *stenyxa*.

I tabellens högra fält finner du ord som har *yxa* som sekundär deskriptor. Även dessa rader är sorterade efter den vänstra (svarta) kolumnen.

11. För jämförelsens skull, skriv nu i stället in ordet *yxa* i den tomma ramen. Du får upp två presentationsrader. På den första står, som väntat, substantivet *yxa* med en deskriptor. Den artificiella sekundära deskriptorn PRIM finns här ej utsatt; för att presentationsraden skall bli så enkel som möjligt utelämnas den "tomma" deskriptorn PRIM i de fall då den är sekundär deskriptor. På den andra raden finns en verbal flerordning (markerad vbm), nämligen *yxa till*. Man kan i SALDO alltså söka inte bara på hela lexem utan också på delar av flerordsuttryck.

Alla blåfärgade lexem är klickbara, och du kan därmed ta dig vidare till en mängd andra lexemsidor. Du kan också återvända till sidan för *stenyxa* och klicka på den sekundära deskriptorn *sten*.

Exempel 2

1. Skriv in ordet *damm*.

2. Du får upp två presentationsrader. I andra raden är såväl lexem som lemma försett med indexsiffran 2. I första raden saknas indexsiffran, men vi skall förstå det så att vi har att göra med *damm*¹ (såväl lexem som lemma). Detta betyder att *damm* är uppdelat på två lexem, vardera med sitt speciella böjningsmönster.

3. Klicka på lemmat *damm* (nn). Du får upp en formtabell, där du bland annat finner *dammar*. Skriv in detta ord i sökfältet (du måste först lämna formtabellen).

4. Presentationsraden för *damm*² är nu borta (*damm* i betydelsen eng. ”dust” har ingen pluralform), men i stället har det tillkommit hela åtta verbala lexem, tre ”enordsverb”, markerade (vb), och fem ”flerordsverb”, markerade (vbm). ”Flerordningar” finns det gott om i detta lexikon, och vi kallar dem också för lexem. I detta fall rör det sig om verb med partikel, *av*, *ner* och *till*. *Damma av* är uppdelat på två lexem, eftersom det kan ha en egentlig och en metaforisk betydelse. Enordsverben är uppdelade på två lemman, eftersom det finns ett intransitivt och ett transitivt *damma*. Det intransitiva *damma* saknar ju t.ex. formen *dammad*. *Damm*² dyker här upp som huvuddescriptor till *damma*¹ och *damma ner* och som sekundär descriptor till *damma*² och *damma av*¹. *Damma ner* representerar bara ett lexem men har fått två rader, vilket beror på att lemmat kan ha två former, med *ned* resp. med *ner*. (Detta avsteg från principen presentationsrad = lexem är inte särskilt tillfredsställande och kommer troligen att korrigeras så småningom.)

Exempel 3

1. Skriv in ordet *kurs*.

2. Du får upp hela sex presentationsrader. *Kurs* är dels grundform, dels genitiv av *kur*. Grundformen *kurs* representerar tre lexem, men bara ett lemma. Grundformen *kur* representerar ävenledes tre lexem, men uppdelade på två lemman, *kur*¹ (nn) och *kur*² (nn).

3. Skriv in formen *kurer*. Du får upp de två presentationsrader vars lemman innehåller denna form, alltså med lexemen *kur*¹ och *kur*².

4. Skriv in formen *kurar*. Som väntat får du upp presentationsraden för lexemet *kur*³, men – hoppsan – också för verbet *kura*!

Exempel 4

1. Skriv in ordet *rå*.

2. Du får upp inte mindre än elva presentationsrader. Två av dessa är verbala flerordningar, nämligen *rå för* och *rå om*. Såsom enordslexem förekommer *rå* som synes i nio varianter. Dessa är fördelade på tre ordklasser, nn (substantiv), av (adjektiv) och vb (verb). Adjektivet *rå* har bara ett paradigm, dvs motsvaras av bara ett lemma. Verbet *rå* kan, beroende på betydelsen, böjas på två olika sätt, därav fördelningen på två lemman. Substantivet *rå*, slutligen är fördelat på tre lemman.

3. Skriv in formen *rår*. Denna form saknas hos adjektivet *rå*.
4. Skriv in formen *rån*. Denna form saknas hos adjektivet och verbet *rå*. Här tillkommer dock två nya lexem, vilka har *rån* som grundform.
5. Skriv in formen *rått*. Denna form saknas hos substantivet *rå*.
6. Skriv in formen *råt*. Denna form saknas hos adjektivet och verbet *rå*, samt hos substantivet *rå*¹. Hos substantivlemmat *rå*² förekommer den som alternativ form till *rået*. Hos substantivlemmat *rå*³ förekommer den som alternativ form till *rån*; här råder således vacklan mellan genus utrum och neutrum. – Alternativa former utgör givetvis inte grund för uppdelning i olika lemman.

Exempel 5

1. Skriv in ordet *pulsar*.
2. Du får nu upp tre presentationsrader, vilket innebär att *pulsar* återfinns inom tre olika lexem, nämnda i första fältet: *puls*², *pulsa* och *pulsar*. De tre formerna *pulsar* är homografa – inte homonyma, eftersom ju betoningen är olika. På samma sätt: om du skriver in *kör* får du med två olika uttal av den inledande konsonanten. SALDO är ett lexikon för skriftspåket och återspeglar därmed inte homonymi.
 Indexsiffran 2 på lexemet *puls*² betyder att det också finns ett lexem *puls*¹.
2. För att undersöka den saken närmare går du tillbaka till den tomma SALDO-rutan överst och skriver in ordet *puls*. Vi noterar att *puls* ingår i en flerordning, en nominalfras markerad (nfm) samt utgör två enordslexem, *puls* (dvs *puls*¹) och *puls*². Flerordningens kärna, dvs *puls*, hänför sig till *puls*¹. Lexemet *puls*¹ har bara en deskriptor. Vi ser också att *puls*¹ och *puls*² tillhör olika lemman, alltså har olika böjning. Flerordningen är försedd med tre deskriptorer, en primär och två sekundära.
3. Du kan nu klicka på *puls* (dvs *puls*¹), och får upp dess lexemsida. De uppåtriktade pilarna visar att *puls* har deskriptorerna *stöt* och PRIM. I den nedre vänstra tabellen, markerad med nedåtriktad dubbelpil, finner du två ord, dels *pulsera*, som saknar sekundärdeskriptor (i dess ställe står PRIM), dels flerordningen *elektromagnetisk puls*, med en komplex sekundärdeskriptor (mer om sådana deskriptorer kommer nedan).
4. Klicka på *pulsera*. Du ser att detta verb har *puls*¹ som överordnat led och *puls*² som underordnat led. Dessutom är *pulsera* sekundärdeskriptor till *pulsar*.
5. Till sist kan du skriva in formen *pulser*. Du får upp två presentationsrader, dock inte *puls*². Detta beror på att *pulser* fungerar som pluralisform bara till *puls*¹, inte till *puls*². *Puls* i flerordningen *elektromagnetisk puls* hör, som nämnts, till *puls*¹.

Exempel 6

1. Skriv in ordet *pisksnärtoperation*. Endast en presentationsrad kommer upp.

2. För mer fullständig information, klicka på lexemfältet. Du ser, ovanför ordrosen, att det också finns en variant med binde-s.

3. Detta är ett ord hämtat direkt från ”verkligheten”, alltså inte från något lexikon. En stor mängd sådana ord flyter kontinuerligt in i SALDO och får sin semantiska analys. I detta fall är *operation* (med den underförstådda indexsiffran 1) primär deskriptor, men resten av ordet, *pisksnärt*, duger inte som sekundär deskriptor. Det får i stället bli *pisksnärtskada*. Denna typ av komprimering är mycket vanlig i svenska sammansättningar. Men när man går vidare finns det ett par andra betydelsekomponenter, som man gärna vill ha med: att det är frågan om en nackskada som uppstår vid påkörning av en annan bil.

4. Klicka nu på den sekundära deskriptorn, dvs *pisksnärtskada*. Observera för övrigt att även deskriptorerna är klickbara redan i presentationsraden; man behöver alltså inte gå via lexemsidan. Om du specifikt är ute efter presentationsraden, måste du dock skriva in ordet i sökrutan.

5. Som synes har *pisksnärtskada* sammansättningen *nackskada*, inte bara det enkla substantivet *skada* (*skada*², till skillnad från verbet, *skada*¹), som primär deskriptor. Som sekundär deskriptor anges ”pisksnärt+påkörning”. Intill helt nyligen var man i SALDO hänvisad till att utse bara en sekundär deskriptor, men nu har vi öppnat för möjligheten att utse två eller flera; vi kan alltså få med såväl *pisksnärt* som *påkörning*, förutom, naturligtvis, huvuddeskriptorn *nackskada*. Vi har redan träffat på ett sådant fall, nämligen flerordningen *elektromagnetisk puls*. Här är ytterligare ett par exempel där en tredje deskriptor nu har kunnat tilläggas:

Hiroshima:stad + Japan+atombombsfällning

curlingförälder:förälder + curling+överbeskyddande

Införandet av en tredje deskriptor föranleddes dock främst av ord av typen *fembarnsmor* och *tolvtaggare*:

fembarnsmor:mor + barn+fem

tolvtaggare:älg + tagg+tolv

I sådana fall blir tilldelning av bara två deskriptorer klart otillfredsställande.

Distinktionen mellan primär och sekundär deskriptor upprätthålles fortfarande. Detta kommer till synes i presentationsraden (som du ju får fram om du skriver in ”pisksnärtskada” eller ”pisksnärtsskada” i sökrutan); här är plustecknet som står efter huvuddeskriptorn avskilt av mellanslag, till skillnad från plustecknet som åtskiljer de båda sekundärdeskriptorerna.

6. Klicka nu på *nackskada*. Du återfinner de båda sekundära deskriptorerna *pisksnärt* och *påkörning* sammanskrivna med plustecken och placerade på en rad. De behandlas alltså i visst hänseende fortfarande som en enhet.

7. Gå tillbaka till sidan för *pisksnärtskada* och klicka på *pisksnärt*. Här saknas deskriptorn *påkörning*. På samma sätt: om du klickar på *påkörning* får du inte fram sekundärdeskriptorn *pisksnärt*. Denna ordning är praktisk bl.a. när man söker på förnamn, som delas av flera personer. Skriver man in *Hus* får man i nedre höger-tabellen upp ”förnamn Jan”, ingenting annat.

Exempel 7

Här fanns i tidigare versioner exempel med deskriptorn *som*² (*som* i komparativ betydelse). Eftersom just dessa ord och fraser i SALDO f.n. befinner sig under omarbetning, utelämnas i föreliggande version detta avsnitt.

Du är nu väl förberedd för fortsatta strövtåg i SALDO. Lycka till!

TILLDELNING AV DESKRIPTORER: STRATEGIER OCH PROBLEM

I de enklaste fallen får man hjälp av ordens morfologiska struktur. T.ex. *torrhet* får den enda deskriptorn *torr*. *Muntorrhet* får primären *torrhet* och sekundären *mun*:

torrhet:torr

muntorrhet:torrhet+mun

Något mer komplicerat kan det bli om avledningen sker utan suffix, t.ex.:

väta:våt

torka:torr

I dessa exempel blir det också lexemuppdelning pga av olika ordklass, (vb) kontra (mn).

Större krav ställs när man måste lita enbart till semantiken. Ett oderiverat ord, som inte är ett av de 43 PRIM-orden, måste få en morfologiskt orelaterad deskriptor, t.ex. *damm*²:torr

Även deriverade ord kan tilldelas en sådan deskriptor, t.ex.

skärgård:ö

järnväg:tåg.

Ishink kan mycket väl analyseras som

ishink:hink+is

men mer informativ blir en icke-morfematisk beskrivning:

ishink:kyla+champagne

Om associationen till *hink* är viktig, kan även detta beaktas:

ishink:hink + kyla+champagne

I vissa fall kan sammansättningsleden användas, men ofullständigt eller indirekt. Jfr följande fall:

pendelur:ur2+pendel

pendelkaos:kaos+pendeltågstrafik

pendeltågstrafik:trafik+pendeltåg

pendeltåg:tåg+pendla2

pendla2:resa+arbetsplats

I vissa sammansättningar är det ena ledet semantiskt tomt, t.ex.

isvak:vak

skogsdunge:dunge

Isvak är synonymt med *vak*, liksom *skogsdunge* är synonymt med *dunge*. Om det enkla ordet, som här, är mer centralt har det i sin tur sammansättningens förled som deskriptor. Denna deskriptor kan vara primär, som i fallet *skogsdunge*, eller sekundär, som i fallet *isvak*:

dunge:skog+smärre

vak:hål+is

Vid tydlig frekvensskillnad kan ordningen bli den omvända:

skulle:höskulle

Mindre klara fall är följande:

framtidsvision:vision (jfr vision:syn+framtid)

hedersutmärkelse:utmärkelse (jfr utmärkelse:belöning+heder)

Tomma förled kan förekomma även vid verb:

återupprepa:upprepa

När det gäller tomma efterled är vissa abstrakter vanliga i denna roll; jfr:

återuppbyggnadsarbete:återuppbyggnad

återuppbyggnadsprocess:återuppbyggnad

sömnlöshetstillstånd:sömnlöshet

åtstrammingsåtgärd:åtstramning

I ett fall som *rödfärgad* är efterledet tomt endast i den ena betydelsen av ordet:

rödfärgad¹:röd (jfr röd:färg¹)

Men: rödfärgad²:färga+röd (jfr färga:färg²)

Abstrakta substantiv som är syntaktiska derivat har samma betydelse som sina utgångsord:

ridning:rida

lätta:lat

Därför kan det kanske tyckas att man vid sammansättningar som har dessa ord som efterled lika gärna kan välja verbet eller adjektivet som primär, t.ex. genom att låta *kamelridning* få *rida* som primär. Men detta bör man inte göra. Det skulle innebära att man skapar mer heterogena grupper, i vilka t.ex. ord med efterleden *arbeta* och *arbete* blandas. Och en viktig princip i SALDO är att skapa så homogena grupper som möjligt. Alltså bör de nu inlagda

renskötsel:sköta+ren2

oljeborrning:borra+olja

ändras till

renskötsel:skötsel+ren2

oljeborrning:borrning+olja

Fall som

återbetalningsskyldighet:återbetalningsskyldig

berörs givetvis inte av denna princip. Huruvida *djupborrning* skall bli

djupborrning:borrning+djup

eller

djupborrning:djupborra

beror på frekvensförhållandet mellan *djupborrning* och *djupborra*. Om det senare bedöms lågfrekvent får det analysen

djupborra:djupborrning

När vi inte har att göra med syntaktiska derivat kan verb mycket väl användas som primärer. Tvetydigheten hos ordet *träskruv* löses t.ex. elegant på följande sätt:

träskruv¹:skruva+trä

skruva:vrida+skruv

träskruv²:skruv+trä

skruv:fästa

*Träskruv*², det är förstås "skruv av trä".

Vid val av förled råder en helt annorlunda situation. Ofta kan man inte avgöra ordklassen, t.ex. *badplats*: *bad* eller *bada*? Ibland vägleder semantiken; jfr *borrdimension* och *borrdjup*. Här är det ofta lämpligt att "hoppa över" ett derivationssteg. T.ex. såväl *ankringsförbud* som *badförbud* kan få verb – *ankra* resp. *bada* – som sekundärer, vilket ökar homogeniteten. Verb är här också syntaktiskt motiverade, jfr *förbjudet att bada*, *kompromissvillig* – *villig att kompromissa*. I fall som *regeringsduglig* och *regeringstid* är verbet *regera* det enda rimliga som sekundär deskriptor.

Ordningsföljden mellan primär och sekundär är inte alltid självklar. Som huvudregel gäller naturligtvis att sammansättningens efterled får bli primär. Men i bl.a. följande fall har avsteg från detta gjorts:

ridkunnig:rida+kunna
grottlänkande:grotta+likna
klotformig:klot+form
elefantunge:elefant+unge
björnhona:björn+hona

Anledningen är att dessa sekundärer har modifierande karaktär.

Mellan två ord bör endast en väg tillåtas. Multipla relationer bör undvikas, även om detta inte alltid är möjligt. Ett exempel är *segel*, vilket i ett första försök analyserades som

segel:båt (meronymi)

Problemet är att *segelbåt* måste analyseras som

segelbåt:båt+segel

Segel får därmed såväl en vertikal som en horisontal relation till *båt*. Detta kan elegant undvikas genom att omanalys:

segel:vind

I själva verket finns ju segel även på isjakter och - kanske i framtiden - rymdfarkoster. Brandsegel och gomsegel blir inga problem i sammanhanget, det är inga riktiga segel.

En rent horisontell multipel relation hotar i fall som *fingerring* och *ringfinger*:

ringfinger:finger+ring

fingerring:ring+finger

Problemet kan kringgå, snarare än elimineras, genom omanalysen

ringfinger:finger+fingerring

Har man tur råder synonymi mellan orden, och då löses problemet lätt, t.ex.:

regeringskoalition:koalitionsregering

koalitionsregering:regering+koalition

Multipel relation verkar ibland inte kunna undvikas:

borrhål:hål+borra

borra:vrida+hål

älgskog:skog+älg

älg:djur+skog

skattebetalande:betala+skatt

skatt:betala

En deskriptor kan återkomma lägre ner i hierarkin med en nivå emellan. Brottet mot multipel relation märks då mindre, t.ex.:

torskfiske:fiske+torsk

torsk:fisk

fiske:fiska
fiska:fånga+fisk
fransktalande:tala+franska
franska:språk+Frankrike
språk:tala

LITET FILOSOFI PÅ SLUTET

Orden i SALDO är organiserade i ett stort hierarkiskt nätverk. Principerna för uppbyggandet av detta har i oförändrat skick övertagits från SAL, med två undantag:

1. Tillhörighet till olika ordklasser utgör numera grund för uppdelning i olika lexem.
2. Som ovan nämnts har det blivit möjligt att ange mer än två deskriptorer.

Trots att lexikonet är konstruerat med hjälp av vertikala relationer är det egentligen de horisontella som är viktigast. SALDO kan ses som en stor grupperingsmaskin. Det gäller att genom en adekvat tilldelning av deskriptorer skapa så homogena och intuitivt tilltalande horisontella lexemklasser som möjligt.

Den tesaurusmodell som går tillbaka på Rogets pionjärbete i början av 1800-talet kan karaktäriseras som global. Ordmassan organiseras efter ett relativt litet antal på förhand uppställda begreppsklasser. Något påminner detta om att beskriva ordens betydelse genom semantiska särdrag. Begreppsklasser och särdrag är instrument som står utanför det som skall ordnas och beskrivas, alltså själva orden.

I SALDO är varje ord inte bara föremål för beskrivning, det är också potentiellt ett instrument för beskrivning, en deskriptor. Sammanställarens enda roll är att fatta ett mycket stort antal beslut, som vardera bara berör en liten punkt i lexikonet. Det är ett lokalt, inte globalt, tillvägagångssätt. Större strukturer i lexikonet växer fram organiskt, utan kontroll ”uppifrån” eller ”utifrån”.

Särskilt bland evolutionsforskare har det vuxit fram en övertygelse om att de allra mest komplexa strukturerna i naturen uppkommer på ett sådant sätt, genom små, små steg. Man jämför ibland med de blixtnabba och radikala förändringarna i form och riktning hos en fågelflock eller ett fiskstim; det är svårt att tro att det faktiskt inte finns någon annan styrning än att varje individ anpassar sig efter de närmaste grannarna enligt mycket enkla ”instruktioner”.

Men den lokala principen råder ju även i vanliga definitionslexika? Ja, förvisso, men även här finns enorma skillnader om vi jämför med SALDO. För det första är orden i traditionella lexikon inte strukturerade i nätverk. För det andra är definitionerna i dessa ordböcker rent semantiska, medan utseendet av den viktigaste och hierarkiska pande deskriptorn i SALDO sker också med hjälp av andra kriterier – även om semantisk närhet alltid är ett villkor. För det tredje bryter ordboksdefinitioner ofta mot en princip som uppställdes redan av Aristoteles: definiens skall vara mer känt än definiendum. Bara med avseende på perifera ordskikt uppfylls den naturliga informerande funktionen, medan frekventa och därmed välkända ord får beskrivningar som innehåller mindre kända komponenter. Sålunda definieras *ligga* ofta med hjälp av

ordet *horisontell*. – SALDO följer genom hela lexikonet Aristoteles princip, men måste ge upp vid ett litet antal ord, som lämnas utan beskrivning.

Principen att definiens skall vara mer frekvent än definiendum är det viktigaste villkoret bakom SALDO:s enkelhetskriterium. Men det är samtidigt en svårbestämd egenskap. Frekvensordböcker är till liten nytta. Det saknar betydelse om ett ord i en stor korpus förekommer fem gånger och ett annat åtta. Det viktiga är hur den relativa vanligheten uppfattas av språkbärarna. Det är bara mycket klara frekvensskillnader som kan användas i enkelhetskriteriet. I sådana fall får t.ex. hierarkiska semantiska relationer – hyponymi och meronymi – stå tillbaka. *Hovdjur* kan inte accepteras som huvuddescriptor till *häst*, *trias* kan inte få *mesozoikum* som huvuddescriptor.

Med undantag av semantisk närhet och dess ytterlighetsfall synonymi är interlexematiska relationer assymetriska och hierarkiska. De skapar liksom pyramider: förgreningarna nedåt är oftast fler än förgreningarna uppåt. Men räckvidden hos relationerna är begränsade till semantiska och derivationella fält samt synonymiräckor. Här finns dock ett markant undantag: frekvensen. Dennas särställning baseras på att den är fri från semantiska och morfologiska band och därmed kan omfatta hela ordmassan. Men pyramidformen finns även här.

Med tanke på att de relationer som binder samman lexemen är hierarkiska och pyramidskapande är det inte förvånande att hela det resulterande nätverket får liknande egenskaper. När man från någon punkt blickar in mot centrum ser man konvergerande linjer och färre enheter, blickar man ut mot periferin ser man divergerande linjer och fler enheter.

Man kan fråga sig om kriterierna närhet, enkelhet och precisering verkligen är så lätthanterliga att de kan tillämpas på ett mycket stort material. Blir inte resultatet oerhört godtyckligt, subjektivt och inkonsekvent? Det har visat sig möjligt att med allt större eftertryck svara nej på den invändningen, och detta av främst två skäl. För det första är en mycket stor del av besluten oproblematiske. Vilken annan descriptor än *brokig* kan ges till *brokighet*, och hur kan *riskanalis* definieras, annat än genom *analys* och *risk*? För det andra – och detta är det viktigare skälet – är varje beslut att utse en descriptor preliminärt. SALDO revideras ständigt, och här återstår ett enormt arbete. Ett förfinat gränssnitt är en avgörande faktor. När man ser hela listan av ord som delar en eller flera descriptorer faller rader som stör homogenitetet i ögonen. Det är ofta påfallande lätt att inse inte bara hur en klar felaktighet kan rättas till utan också hur en tveksam lösning kan förbättras. Just härigenom rör vi oss hela tiden bort från godtycklighet och subjektivitet i riktning mot ett idealtillstånd, där varje lexem har knuffats in på sin optimala plats i nätverket.