

# SALDO

Beskrivningen av SALDO finner du på adressen

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

För att söka i SALDO går du till den här adressen:

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

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

Hur det går till att söka i SALDO kan enklast visas med några konkreta 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; *nn* 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 och om A har en morfologiskt enklare struktur än B (detta gäller framför allt ordbildningsrelation). 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 *yx*, vilket också är hyperonymen till *stenyxa*, men detta är förstas 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 en semantisk granne med specificerande funktion. Den sekundära deskriptorn är inte obligatorisk; den saknas hos drygt hälften av alla lexem i SALDO.

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

3. Klicka nu på lemmat.

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

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 t.ex. som förled inte bara substantivet *sten* utan också verbet *stena*, som efterled inte bara substantivet *yxa* utan också verbet *yxa* (jfr *yxa till*).

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 *yxa* med dubbel uppåtpil, den sekundära *sten* med enkel uppåtpil. Pilarnas riktning anger att deskriptorerna är överordnade i förhållande till nyckelordet (lexemet). Samtliga i lexikonet inlagda lexem utgör nämligen ett sammanhängande, hierarkiskt ordnat nätverk.

På raden under deskriptorramen återfinner 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 lemmat 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* och *trindyxa* med den sekundära deskriptorn *rundad*. 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 *stenyxcultur*.

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, 51, 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 *yx*a 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 *yx*a som sekundär deskriptor. Även dessa rader är sorterade efter den vänstra (svarta) kolumnen.

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 indexsiffra, men vi skall förstå det så att vi har att göra med *damm*<sup>1</sup> (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 *damm*<sup>ar</sup>. Skriv in detta ord i sökfältet.

4. Presentationsraden för *damm*<sup>2</sup> är nu borta, men i stället har det tillkommit två verb, *damma* och *damma av*, varav det sistnämnda är uppdelat på två betydelser, en ”egentlig” och en metaforisk. *Damm*<sup>2</sup> dyker här upp som deskriptor till *damma* och *damma av*<sup>1</sup>.

## Exempel 3

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*<sup>2</sup>, *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*<sup>2</sup> betyder att det också finns ett lexem *puls*<sup>1</sup>. Detta återfinner du som sekundär deskriptor till tredje radens lexem, *pulsar*. Lexemet *puls* står här utan indexsiffra, men vi skall – som sagt – tänka oss det försett med siffran 1.

2. Du kan nu klicka på denna deskriptor, dvs *puls* (dvs *puls*<sup>1</sup>), och får upp dess lexemsida. Du ser att *puls* har deskriptorerna *stöt* och PRIM. I nedre högra tabellen finner du *pulsar*.

3. Gå nu tillbaka till den tomma SALDO-rutan överst och skriv in ordet *puls*. Du får nu upp presentationsraderna för *puls* (dvs *puls*<sup>1</sup>) och *puls*<sup>2</sup>. I första fallet ges bara en

deskriptor; för att presentationsraden skall bli så enkel som möjligt utelämnas den ”tomma” deskriptorn PRIM i de fall den är sekundär deskriptor.

4. Till sist kan du skriva in formen *pulser*. Du får upp bara en presentationsrad. Detta beror på att *pulser* fungerar som pluralisform bara till *puls*<sup>1</sup>, inte till *puls*<sup>2</sup>. Detta förhållande avspeglas också i att *puls* inte enbart är uppdelat på två lexem, utan också på två lemman: *puls*<sup>1</sup> (nn) och *puls*<sup>2</sup> (nn).

#### Exempel 4

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*<sup>1</sup> (nn) och *kur*<sup>2</sup> (nn).
3. Skriv in formen *kurer*. Du får upp de två presentationsrader vars lemman innehåller denna form, alltså med lexemen *kur*<sup>1</sup> och *kur*<sup>2</sup>.
4. Skriv in formen *kurar*. Som väntat får du upp presentationsraden för lexemet *kur*<sup>3</sup>, men – hoppsan – också för verbet *kura*!

#### Exempel 5

1. Skriv in ordet *rå*.
2. Du får upp inte mindre än elva presentationsrader. De två sista, *rå för* och *rå om*, visar att man kan söka inte bara på hela lexem utan också på delar av flerordsuttryck. Dessa kallar vi i detta lexikon också för lexem. Som 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*.
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*<sup>1</sup>. Hos substantiven *rå*<sup>2</sup> och *rå*<sup>3</sup> förekommer den som alternativ form till *rået* och *rån*. Alternativa former utgör inte grund för uppdelning i olika lemman.

#### Exempel 6

1. Skriv in ordet *pisksnärtoperation*. Endast en presentationsrad kommer upp.
2. 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 det finns 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.
3. Klicka nu på den sekundära deskriptorn, dvs *pisksnärtskada*. Observera att även deskriptorerna är klickbara redan i presentationsraden; man behöver alltså inte gå via lexemsidan.
4. Som synes har *pisksnärtskada* sammansättningen *nackskada*, inte bara det enkla ordet *skada*, som primär deskriptor.
5. Klicka på *nackskada*.
6. Primärdeskriptorn till *nackskada* är substantivet *skada*<sup>2</sup>. *Skada*<sup>1</sup> är (det semantiskt mer primära) verbet.
7. När det gäller den sekundära deskriptorn till *pisksnärtskada* måste man välja mellan *pisksnärt* och *påkörning*. Man måste alltså avstå från en betydelsekomponent. Du kan klicka på ordet och själv konstatera vilket alternativ som f.n. ligger ”inne”. Inte heller redan gjorda analyser är slutgiltiga i SALDO; de är som redan påpekats föremål för fortlöpande revidering. – Ett liknande val som mellan *pisksnärt* och *påkörning* ställs man för övrigt inför i nybildningen *curlingförälder*; här är dock den mer informativa deskriptorn *överbeskyddande* att föredraga.

## Exempel 7

1. Skriv in ordet *som*.
2. Du får upp ett stort antal presentationsrader. Detta beror på att *som* ingår i en mängd flerordsuttryck, och sökprogrammet hittar samtliga dessa fall.
3. Klicka på lexemet *som*<sup>2</sup>. Du kan söka upp raden för detta lexem, men ett kanske enklare sätt är att titta i radernas mellanfält, eftersom just detta ord ofta förekommer som sekundär deskriptor.
4. Du får nu upp den semantiska tabellen för *som*<sup>2</sup>. I avdelningen markerad med enkel nedåtpil finner du ett stort antal flerordsuttryck av typen *som en varg*. Den fullständiga frasen är ju *hungrig som en varg*, men i SALDO har vi valt att utelämna *hungrig*, eftersom detta ord fungerar som första deskriptor. Hela tabellen är ju för övrigt sorterad efter de icke-klickbara (svarta) orden i första kolumnen, dvs primärdeskriptorerna.

5. Gå ned till primärdeskriptorn *röd* och klicka på *korallröd*.
6. I deskriptorraden för *korallröd* finner du inte bara en utan två sekundära deskriptorer, *korall* och *som<sup>2</sup>*. Dessa är åtskilda med ett plustecken.
7. Klicka på primärdeskriptorn *röd*. I nedre vänstra tabellen ser du att sekvensen *korall+som<sup>2</sup>* är behandlad som en enhetlig sekundärdeskriptor.
8. Skriv nu i stället in ordet *korallröd* i sökrutan.
9. Du får upp presentationsraden för *korallröd*. Obs att skillnaden mellan primärdeskriptor och sekundärdeskriptor är markerad genom att det plustecken som åtskiljer sekundärdeskriptorerna inte är omgivet av blanktecken.

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