

## INNEHÅLL

<b>1</b>	<b>Introduktion</b>	<b>2</b>
<b>2</b>	<b>Snabbinställningar</b>	<b>3</b>
<b>3</b>	<b>Resultatvy</b>	<b>3</b>
<b>4</b>	<b>Analyslägen</b>	<b>4</b>
<b>5</b>	<b>Avancerade inställningar</b>	<b>4</b>
<b>6</b>	<b>Filuppladdning</b>	<b>6</b>

# 1 INTRODUKTION

Sparv är Språkbankens annoteringsverktyg som används bland annat för att analysera korpusarna i Korp och texterna i Strix. Sparvs webbgränssnitt<sup>1</sup> kan användas för att annotera egna texter.

Denna användarhandledningen beskriver de flesta funktionerna som finns tillgängliga i webbgränssnittet. Det finns även ett par övningsuppgifter för Sparv som du kan ladda hem här<sup>2</sup>.

För att göra en analys räcker det att skriva eller klistra in en text i textfältet (alternativt trycka på en av exempelknapparna) och sedan trycka på den gröna **Kör**-knappen. När analysen är klar visas den som tabell under inmatningsfältet.

English Info Nytt Länkar

**Sparv**  
Språkbankens annoteringsverktyg

Språk  
BANKEN SWE-CLARIN

Analyspråk: svenska ▼ Ladda exempel: Drama Åtta sidor Talbanken Läsbart Ikea Exempelkorpus

Mata in Ladda upp

☒ Ren text ☐ XML

1

☒ Lexikalanalys ☒ Sammansättningsanalys ☒ Dependensanalys ☒ Attitydanalys ☐ Namntaggare ☒ Läsbarhetsvärden

Visa avancerade inställningar Återställ standardinställningar

Kör!

Figur 1: Sparvs webbgränssnitt

<sup>1</sup><https://spraakbanken.gu.se/sparv>

<sup>2</sup>[https://github.com/spraakbanken/sparv-docs/raw/master/exercises/sparvovningar\\_hw2017.pdf](https://github.com/spraakbanken/sparv-docs/raw/master/exercises/sparvovningar_hw2017.pdf)

## 2 SNABBINSTÄLLNINGAR

Med kryssrutorna under inmatningsfältet kan man snabbt välja eller välja bort analystyper. Om man t.ex. inte vill se dependensträden i resultatet, kan man klicka ur kryssrutan för dependensanalysen.

Ovanför inmatningsfältet väljer man om man vill mata in ren text eller XML. Standardläget är ren text. Om man matar in uppmärkt text måste man byta till XML för att inte uppmärkningen ska försvinna.

## 3 RESULTATVY

När analysen är klar visas resultatet som en tabell under inmatningsfältet. Varje kolumn i tabellen representerar en viss typ av analys på ordnivå. Man kan få en kort förklaring för en analystyp genom att hålla musen över namnet i tabellhuvudet. Även många andra förkortningar förklaras när man håller muspekaren över dem, t.ex. etiketterna i dependensanalysen, kategorierna i namntaggar och de morfosyntaktiska beskrivningarna.

Om man har valt att göra en dependensanalys så visas denna inte bara i tabellen utan även som ett dependensträd i början av varje mening. Strukturella attribut, dvs. analyser som kan sträcka sig över flera ord eller över hela texten (t.ex. läsbarhetsvärden eller namntaggar) visas som XML-taggar i tabellen.

Man kan även ladda ner hela analysen som XML genom att trycka på **XML**-knappen ovanför tabellen.

XML

```
<text lix="49.56" ovix="79.90" nk="2.68" title="Ikea"> [Visa XML]
<paragraph> [Visa XML]
<sentence id="8f746-8322b"> [Visa XML]
  SS: Subjekt (övrigt subjekt)
  ( Ikea ( namnet är bildat av initialerna för Ingvar Kamprad Elmtaryd Agunnaryd ) är ett multinationellt möbelföretag som
```

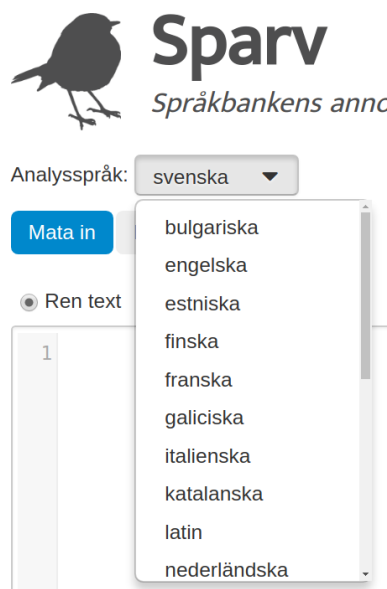
token	msd	lemma	lex	sense	deprel
Ikea	PM. NOM	Ikea	Ikea..pm.1	IKEA..1, Ikea..1	SS
(	PAD				IR
namnet	NN. NEU. SIN. DEF. NOM	namn	namn..nn.1	namn..1 (0.967), namn..2 (0.020), namn..3 (0.012)	SS
är	VB. PRS. AKT	vara	vara..vb.1	vara..1	MS
bildat	VB. SUP. AKT	bilda	bilda..vb.1	bilda..1 (0.992), bilda..2 (0.008)	+F
av	PP	av	av..pp.1	av..1	AG

Figur 2: En färdig sparvanalys

## 4 ANALYSLÄGEN

Under Sparvlogotypen finns det en analyspråksväljare. Med hjälp av denna kan man analysera texter som är skrivna på andra språk än svenska. I nuläget har Sparv stöd för 20 analyspråk, men för de flesta språken stöds enbart ordklasstaggning och lemmatisering.

Sparv har även ett analysläge för att annotera svenska texter från 1800-talet. Här används information från två äldre ordböcker (Dalin och Swedberg) för att få fram bättre analyser av ord med gammal stavning.



Figur 3: Språkväljarmenyn

## 5 AVANCERADE INSTÄLLNINGAR

När man vill analysera en text som inte har någon uppmärkning räcker det oftast att man skriver in texten i inmatningsfältet och sedan trycker på **Kör**-knappen. Om texten innehåller XML behöver man dock ändra på några inställningar för att berätta för Sparv vilka funktioner de olika taggarna i indatan fyller.

Inställningarna öppnas genom att klicka på **Visa avancerade inställningar**. Här kan man ange om styckes-, menings- och ordsegmenteringen ska ske automatiskt eller om befintliga attribut ska användas, man kan specificera vad dokumentelementet heter, samt ange ytterligare taggar och attribut som ska bevaras från indatan. Taggar och attribut som inte specificeras kommer att försvinna i resultatet.

Man har även kontroll över exakt vilka analyser som ska göras genom att klicka på de olika analysattributen. Är man t.ex. bara intresserad av läsbarhetsvärdet **lix** kan man klicka ur **ovix** och **nk** i de avancerade inställningarna. Man kan få en kort förklaring för en inställning genom att klicka på frågetecknet bredvid den. Med hjälp av återställningsknappen kan man få fram standardinställningarna.

Dölj avancerade inställningar
Återställ standardinställningar
Kör!

Korpusnamn:

Ordsegmentering
Segmenterare
whitespace linebreaks blanklines default\_tokenizer

Meningssegmentering
Segmenterare
Meningschunk
Fördefinierad attribut
text paragraph
Segmentera med
whitespace linebreaks blanklines default\_tokenizer punctuation

Styckessegmentering
Segmenterare
Styckeschunk
Dokumentelement
Segmentera med
whitespace linebreaks blanklines

Dokumentelement
Tagg:
Strukturella attribut
+

Ytterligare strukturella element
+

Positionella attribut
Lexikalanalys
pos msd lemma lex sense
Sammansättningsanalys
complemgram compwf prefix suffix
Dependensanalys
ref dephead deprel
Attitydanalys
sentiment

Namntaggar
ex type subtype

Textattribut
Läsbarhetsvärden
lix ovix nk

Visa Makefil
Visa formulärets JSON
Visa XML-taggar
Kör!

Figur 4: Sparvs avancerade inställningar i XML-läget

## 6 FILUPPLADDNING

Istället för att mata in text i Sparvs textfält kan man använda sig av filuppladdningen genom att trycka på **Ladda upp**-knappen. Accepterade filformat är txt (ren text) och XML. Dokument skapade i Microsoft Word och liknande kan inte analyseras.

Om man laddar upp XML-filer behöver man anpassa de avancerade inställningarna för att inte uppmärkningen ska tolkas som vanlig text. Även analyspråket måste väljas i filuppladdningsläget. Observera att man kan ladda upp och analysera flera filer åt gången. Inställningarna och språkvalet gäller då alla filer som laddas upp i samma körning.

Vissa analyser tar lång tid på sig och om man vill slippa vänta kan man skriva in sin mailadress i email-fältet. Man kommer då att få ett mail med en nedladdningslänk när analysen är klar. Om du har angett din mailadress går det bra att stänga webbläsarfönstret medan analysen körs.

När man använder sig av filuppladdningen visas resultatet inte som tabell utan man kan ladda ner en zip-fil med den annoterade texten i XML-format.

Analyspråk: svenska ▼

Mata in Ladda upp

+ Lägg till filer... Återställ

min\_text.txt ta bort test.txt ta bort

Maila mig när körningen är klar

e-mail

☒ Lexikalanalys ☒ Sammansättningsanalys ☒ Dependensanalys ☒ Attitydanalys ☐ Namntaggare ☒ Läsbarhetsvärden

⚙ Visa avancerade inställningar Återställ standardinställningar ▶ Ladda upp och kört!

Figur 5: Filuppladdningsläget