

---

# Sparv

Övningar  
Språkbankens höstworkshop 2017

---

<https://spraakbanken.gu.se/sparv>  
[sb-sparv@svenska.gu.se](mailto:sb-sparv@svenska.gu.se)

17 oktober 2017



# ÖVERSIKT

Sparv är Språkbankens annoteringsverktyg som används bland annat för att analysera korpusarna i Korp och texterna i Strix. Sparvs webbgränssnitt (<https://spraakbanken.gu.se/sparv>) kan användas för att annotera egna texter.

För att göra en analys räcker det att skriva eller klistra in en text i textfältet och sedan trycka på knappen **Kör**. När analysen är klar visas den som tabell under inmatningsfältet. Varje kolumn i tabellen representerar en viss typ av analys. Man kan få en kort förklaring för en analystyp genom att hålla musen över namnet i tabellhuvudet.

## ÖVNING 1 - MODERN SVENSKA

Använd Sparv för att analysera meningen:

*Vad har vi för olika analyser?*

- 1.1 I kolumnen **msd** finns den morfologiska beskrivningen. Låt muspekaren vila på en förkortning för att få en förklaring. Vad har ordet *Vad* fått för beskrivning?
- 1.2 I **lemma**-kolumnen kan vi se ordet normaliserat till grundform. Håller du med om normaliseringarna? Har något ordpar fått en gemensam analys?
- 1.3 Under **lex** kan lemgrammen slås upp i Karp och under **sense** ges ordbetydelser från SALDO samt sannolikhetsvärden i fall det finns flera analyser per ord. Vilka olika uppslag får vi för *ha* och *för*?

Prova att göra en Sparvanalys av meningen:

*Den gamla damen träffade killen med handväskan.*

- 1.4 Sparv analyserar dependensträdet med hjälp av parseern Malt tränad på Talbanken. Vem är det som har handväskan i den här meningen enligt analysen? Förklaringar visas när muspekaren är över de olika förkortningarna i trädet.
- 1.5 Vilken betydelse av *träffa* har fått det högsta sannolikhetsvärdet? (Detta visas i **sense**-kolumnen.)
- 1.6 Prova lite andra variationer av meningen för att se om handväskan kan analyseras att tillhöra någon annan. Kolla även om betydelseerna av ordet *träffa* kan få andra sannolikhetsvärden eller en annan rakning.
- 1.7 Under **complemgram** (sammansatta lemgram) och **compwf** (sammansatta ordformer) ses resultatet av sammansättningsanalysen. Ett av orden har fått en bra analys, men ett av dem en lite sämre. Vad kan det bero på?

Klicka i kryssrutan **Namntaggar** och gör en Sparvanalys av följande mening:

*Den brittiske författaren Kazuo Ishiguro, född i Japan 1954, har tilldelats Nobelpriset i litteratur.*

- 1.8 Hur många namnentiteter har analysen hittat? Vilka typer och undertyper har de? Finns det någon entitet som inte är ett namnuttryck?
- 1.9 Vilka ord har fått positiva/negativa/neutrala attityder? *Tips: när du håller muspekaren över attitydsiffran får du veta hur den ska tolkas.*

## ÖVNING 2 - 1800-TALSSVENSKA

Sparv har 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.

Gå till analyspråksmenyn och välj **svenska-1800-tal**. Ladda exemplet genom att trycka på Aftonbladet-knappen och kör analysen. Gå ner till meningen som börjar på "Lådan var sannolikt för liten...". I kolumnerna **lemma** och **lex** är det markerat vilket lexikon analysen kommer ifrån. Ordet *liten* t.ex. får tre analyser eftersom ordet förekommer i både Dalin, Swedberg och SALDO.

- 2.1 Från vilket lexikon får ordet *blifva* sin analys i den ovanstående meningen?
- 2.2 Vad får man för information om man klickar på ordet i **lex**-kolumnen?

## ÖVNING 3 - ANDRA SPRÅK

I nuläget har Sparv stöd för 20 analyspråk. De flesta analystyper finns bara för svenska. De andra språken har ordklasstagning och lemmatisering.

- 3.1 Välj ett språk från analyspråksmenyn, exempelvis finska, ryska eller latin. Skriv antingen in en egen text eller ladda ett exempel med knappen ovanför textfältet till höger. Gör nu en Sparvanalys för att få grundformer, en morfologisk analys och ordklasser. Ordklasser kommer från UD's taggmängd<sup>1</sup>, men språken skiljer sig lite emellanåt i hur den morfologiska analysen presenteras.

---

<sup>1</sup><http://universaldependencies.org/u/pos/>

## ÖVNING 4 - FILUPPLADDNING (AVANCERAD)

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 en xml-fil behöver man justera de avancerade inställningarna för att berätta för Sparv vilka funktioner de olika xml-taggar i indatan fyller. Om t.ex. dokumentelementet heter *korpus* så skriver man in det i textrutan bredvid **Dokumentelement > Tagg**. Attribut som inte specificeras kommer att försvinna i resultatet. Man kan få en kort förklaring för en inställning genom att klicka på frågetecknet bredvid den. Observera att man kan ladda upp och analysera flera filer åt gången. Inställningarna gäller då alla filer som laddas upp i samma körning.

- 4.1 Testa att ladda upp en txt- eller xml-fil. Om du vill kan du skriva in din mailadress i email-fältet för att få ett mail med en nedladdningslänk när körningen är klar. Om du väljer att ladda upp en xml-fil får du kontrollera så att inställningarna innehåller rätt information om dina xml-taggar.
- 4.2 När körningen är klar ladda ner zip-filen genom att klicka på filsymbolen under **Färdiga körningar** eller genom att klicka på länken i mailet du har fått. Öppna zip-filen och xml-filen som ligger inuti. Har styckes- och meningsindelningen blivit korrekt? Har något av orden fått en sammansättningsanalys?