

Korpuslingvistik (SV2119)

Instruktioner inför uppgift 2



Richard Johansson

`richard.johansson@svenska.gu.se`

15 november 2013

om uppgift 2

- ▶ kvantitativa undersökningar av syntaktiska fenomen
- ▶ vi kommer att söka i syntaktiskt annoterade korpusar (trädbanker)
- ▶ vi kommer att använda sökverktyget TIGERSearch



översikt uppgift 2

- ▶ välj minst en av 4 trädbanker
 - ▶ svenska, engelska, tyska, italienska
 - ▶ är du intresserad av något annat språk, kontakta oss
- ▶ lös alla uppgifter märkta **ALL** eller med din trädbanks språk
- ▶ lösningen:
 - ▶ redovisa de sökfrågor du använt
 - ▶ diskutera hur bra resultatet blev
- ▶ deadline: 13 december

https://svn.spraakdata.gu.se/repos/richard/pub/k12013_web/uppg2.pdf



sökverktyget TIGERSearch

- ▶ detta är ett sökverktyg för att söka i frasstrukturträdbanker
- ▶ i jämförelse med Korp:
 - ▶ frasstruktur, inte dependenser
 - ▶ i TIGERSearch söker man på struktur, i Korp på enskilda ord

att köra TIGERSearch i datorsalen

- ▶ I dockan längst ned till höger bredvid papperskorgen finns en applikations-mapp
 - ▶ välj där TIGERSearchTools och TIGERSearch.app

att installera på egen dator

- ▶ se utskickat mail
- ▶ fråga oss i slutet om du behöver hjälp!



översikt TIGERSearch

- ▶ välj korpus
- ▶ utforska korpus
- ▶ sökruta



dokumentation av TIGERSearch

- ▶ manual
- ▶ fuskklapp

`http://www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/TIGERSearch/doc/html/QueryLanguage\_QuickReference.html`

- ▶ också bra att ha dokumentation för trädbanken tillgänglig



sökning: grunderna

- ▶ sökning med avseende på ...
 - ▶ ord och ordklass (*särdrag* för enskilda ord)
[word="jag"]
[pos="NN"]
[word="att" & pos="SN"]
 - ▶ frastyp (*särdrag* för fras)
[cat="PP"]

namngivning

- ▶ för att kunna kombinera flera kriterier
- ▶ för att kunna samla statistik

```
#s:[word="som"]
```

```
#pp:[cat="PP"]
```

- ▶ om du har namngivit någon del av din sökning så kan du sammanställa statistik
- ▶ vilket namn? vilket särdrag?
- ▶ enskilda träffar, frekvens

mer flexibla sökningar

- ▶ sökmönster (reguljära uttryck):

`[word=/[Ää]ktenskap.*/]`

`[pos=/JJ|NN/]`

`[cat=/S|VP/]`

- ▶ att använda reguljära uttryck kan göra att sökningen blir långsammare

sökning med avseende på relationer

- ▶ söka efter meningar som innehåller ...
 - ▶ både *i* och *skolan*; både en prepositionsfras och en nominalfras
`[word="i"] & [word="skolan"]`
`[cat="PP"] & [cat="NP"]`
 - ▶ *i* direkt följt av *skolan*
`[word="i"] . [word="skolan"]`
 - ▶ *i* följt av *skolan*
`[word="i"] .* [word="skolan"]`
 - ▶ en prepositionsfras bildad av en nominalfras (*direkt dominans*)
`[cat="PP"] > [cat="NP"]`
 - ▶ en prepositionsfras som innehåller ett adjektiv (*allmän dominans*)
`[cat="PP"] >* [pos="JJ"]`

sökning med avseende på relationer (forts)

- ▶ söka efter meningar som innehåller ...
 - ▶ en sats vars objekt är en nominalfras
[cat="S"] >00 [cat="NP"]
 - ▶ vilket subjekt som helst
#x >SS #y
 - ▶ en prepositionsfras och en nominalfras, där båda har samma förälder (*syskonrelation*)
[cat="PP"] \$ [cat="NP"]
 - ▶ en prepositionsfras och en nominalfras som är syskon, där PP kommer före NP
[cat="PP"] \$ [cat="NP"]

kombination av kriterier

- ▶ sökkriterier kan kombineras med & (*och*), | (*eller*)
- ▶ sök efter nominalfraser omgivna av två prepositionsfraser

```
[cat="PP"] $.* #np:[cat="NP"]  
& #np:[cat="NP"] $.* [cat="PP"]
```

- ▶ vilka är de vanligaste transitiva verben?

```
#sats >00 #objekt  
& #sats >FV #verb
```

- ▶ det finns inte alltid ett enda rätt svar på frågan
- ▶ lösningarna kan se ganska olika ut beroende på vilken trädbank du använder
 - ▶ svenska och tyska trädbankerna mer “platta”, engelska och italienska mer “djupa”
 - ▶ den italienska har inga grammatiska funktioner utan bara fraser
- ▶ om du är osäker på hur något annoteras i din trädbank ...
 - ▶ ta en titt i manualen (om den finns)
 - ▶ försök hitta exempel i trädbanken och utgå från dem
 - ▶ t.ex.: hur ser relativsatser ut i Talbanken?
- ▶ fråga gärna om hjälp!