

Övning på Språkbankens verktyg Korp

Lars Ahrenberg (modellerade efter material på Korps webbplats)

Vad är Korp?

Korp är Språkbankens verktyg för att utforska textkorpusar. Det finns som en webbtjänst och nås på adressen <https://spraakbanken.gu.se/korp/>

Korps funktionalitet

Den primära funktionen i Korp är sökning efter ord och fraser i textkorpusar. En sökning går till så att man

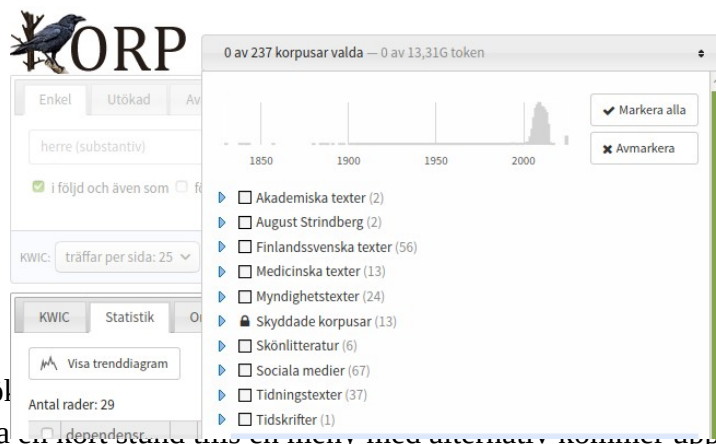
1. Väljer en eller flera korpusar man vill söka i
2. Anger ett sökord eller sökuttryck
3. Väljer vilka typer av resultat man vill ha och några parametrar som påverkar resultatet. Med de förinställda valen får man 1/ alla meningar i de valda korpusarna som innehåller sökordet/sökuttrycket, 2/ statistik över träffarna.
4. Trycker på knappen Sök

En översikt över alla funktioner finns i Korps användarhandledning som du hittar via ikonerna uppe till höger på Korps söksida eller här: <https://spraakbanken.gu.se/verktyg/korp/anvandarhandledning>

Övning 1

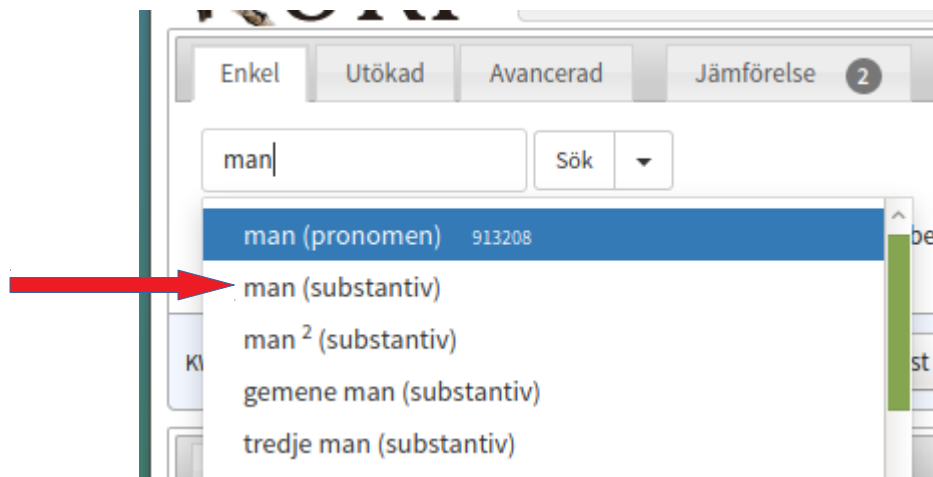
Gör en sökning på alla ordformer av substantiven *man* och *kvinna* i texter från Göteborgsposten och ta med ordbilder i resultatet. Gången är:

1. Öppna korpusmenyn (det grå fältet omedelbart till höger om namnet KORP). Börja med att trycka på knappen **Avmarkera** för att eliminera alla förhandsval. Tryck sedan på den lilla blå pilen framför rubriken Tidningstexter. Detta tar dig till en undermeny där du avmarkerar alla alternativ utom GP. Stäng korpusmenyn genom att klicka någonstans utanför den.

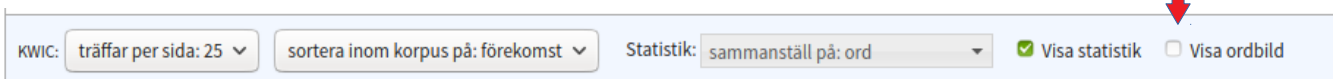


2. Gå till sökrutan och vänta på resultatet. Välj det alternativ som anger *man* (eller *kvinna*) i sökrutan

substantiv i parentes efter sökordet. I Korps terminologi kallas ett sådant sökuttryck för ett 'lemgram' och garanterar att alla böjningsformer (*mannen, män, männen, etc. resp. kvinnan, kvinnor, etc.*) kommer att hittas.

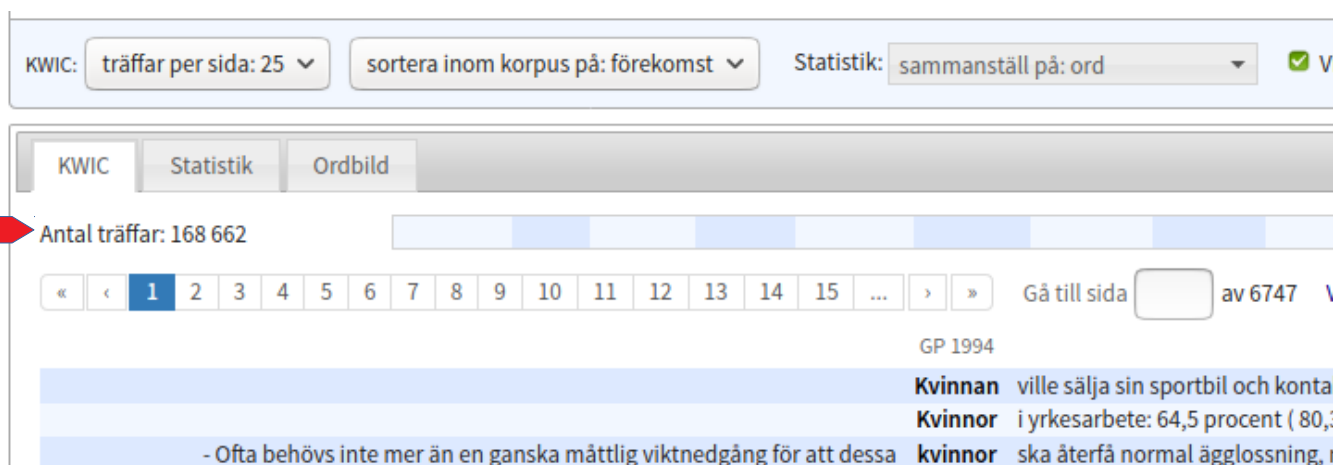


3. Markera alternativet **Visa ordbild** innan du startar sökningen. Det gör du längst ut till höger i det blå fältet som har ordet KWIC längst till vänster.



4. Tryck på **Sök**. (**Obs!** Det kan ta en stund innan Korp svarar.)

Som första resultat får du en lista med träffar. Dessa visas normalt 25 stycken i taget och med delkorpusarna i en bestämd ordning. Du får också besked om totala antalet träffar



Fråga: Vilket av de två substantiven *man* resp. *kvinnor* har flest träffar? Är skillnaden förväntad?

I KWIC-läget kan du studera alla träffar och det finns också möjlighet att ladda ner resultatet till sin egen dator (se användarhandledningen). Om vi i stället tittar under fliken *Statistik* får vi data för alla förekommande böjningsvarianter. I defaultläget ("sammanställ på ord") presenteras frekvensstatistik för de olika ordformerna, men man kan ställa om denna parameter så att man får en annan indelning, t.ex. på dependensrelation, där man får frekvenser för ordets förekomst som subjekt eller direkt objekt och andra grammatiska relationer.

<input type="checkbox"/>	ord		Totalt	GP 1994	GP 2001
<input checked="" type="checkbox"/>	Σ		620,1 (168 662)	655,7 (13 987)	535,3 (9 320)
<input type="checkbox"/>	kvinnor		238,2 (64 798)	253 (5 397)	217,2 (3 782)
<input type="checkbox"/>	kvinna		138,6 (37 706)	129,8 (2 768)	112,8 (1 964)
<input type="checkbox"/>	kvinnan		88,8 (24 164)	83,6 (1 783)	70,1 (1 220)

Fråga: Finns det någon skillnad mellan orden *man* och *kvinna* vad gäller den relativa frekvensen hos olika böjningsformer?

Det tredje läget, *Visa ordbild*, ger listor på ord och uttryck som tenderar att förekomma tillsammans med sökordet. Orden är uppdelade på olika ordklasser. Ordningen i en lista är inte baserad på absolut frekvens utan på ett mått som anger graden av samförekomst mellan sökordet och listordet.

Preposition	Attribut	kvinna	Efterställt Attribut		
1. av	7150	1. ung	4844	1. vara	1282
2. för	6163	2. gravid	778	2. i ålder	285
3. mot	2458	3. gammal	1724	3. ha	760
4. bland	1245	4. fler	834	4. utsätta	176
5. hos	919	5. muslimsk	451	5. få	447
6. om	1705	6. 23-åriga	254	6. i 40-årsåldern	176

Fråga: Vilka skillnader och likheter kan du se i ordbilden för *man* resp. *kvinna*? Hur kan dessa förklaras?

Fundera på följande:

1. Varför är det viktigt att söka på *man* (*substantiv*) och inte bara på ordet *man*?
2. Hur skulle man kunna använda Korp för att undersöka om det skett någon förändring över tid när det gäller skillnader i frekvens mellan referenser till män resp. kvinnor i GP-texterna?
3. Hur skulle man kunna använda Korp för att ta reda på genretypiska drag i tidningstext?