

Språkbanken är en språkteknologisk forskningsenhet vid Göteborgs universitet. En central del av våra aktiviteter består i att utveckla en språkteknologisk infrastruktur som ska underlätta forskning på stora svenska textmassor från alla tider och även möjliggöra att man ställer nya sorters forskningsfrågor.

Här kommer *Korp* in i bilden. *Korp* är ett verktyg för att utforska svenska texter, där en stor språkteknologiskt annoterad textsamling – närmare tio miljarder ord totalt – görs tillgänglig via avancerade sökfunktioner och analysverktyg.

Nedan ger vi några smakprov på typer av frågor man kan söka svar på med hjälp av *Korp*. För mer information, besök: <<http://spraakbanken.gu.se>>.

Det är allmänt känt att August Strindbergs kvinnoösyn lämnade en del övrigt att önska, men lämnar den även spår i hans verk?

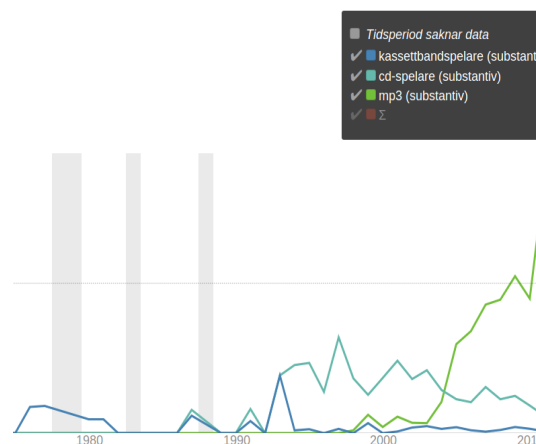
I detta exempel har vi sökt efter "kvinna" och "man" i August Strindbergs verk och brev. Det som visas här är så kallade ordbilder, vilket enkelt uttryckt är sökordens samförekomster.

Preposition	Attribut	kvinna	Efterställt Attribut	kvinna	verb	Verb	kvinna
1. för	114	1. ung	64	1. älska	7	1. älska ²	34
2. av	142	2. gift	12	2. älska ²	7	2. älska	34
3. hos	31	3. gammal	42	3. för porterstämman	4	3. älska ²	14
4. med	111	4. olycklig	13	4. med barn	6	4. älska	14
5. åt	28	5. ogift	9	5. vara	18	5. frigöra	5
6. bland	14	6. samn	11	6. vilja	8	6. äro	139
7. mot	29	7. ofruktam	5	7. bo	4	7. känna ²	10
8. till	64	8. stackars	12	8. offra	3	8. skola ²	17
9. genom	18	9. ond	16	9. kunna	10	9. vilja	11
10. om	32	10. skön	13	10. på jordstycke	2	10. känna	11
11. i närheten	2	11. lös ²	8	11. åt lugngingame	2	11. gifta sig	6
12. i alla händelser	1	12. lös	8	12. i hämmännen	2	12. kunna	15
13. i den mån	1	13. grovlemmad	4	13. av chevaleriet	2	13. föda	7
14. inför	4	14. charmanta	4	14. från polisuppsikten	2	14. exploatera	3
15. till och med	3	15. oavhängig	4	15. vid höte	2	15. uppfinna	4

Något tillspetsat ser vi t ex att mannen sägs vara *hederlig*, *rättfärdig*, *trovärdig* och *klok* (samt möjligen *fattig* eller *död*), medan kvinnan oftare är *ond*, *lös* eller beskrivs på sätt som relaterar till mannen (*gift*, *ogift*, *ofruktam* och *oavhängig*).

Preposition	Attribut	man	Efterställt Attribut	man	verb	Verb	man
1. av	180	1. ung	342	1. av ära	12	1. vise	24
2. för	154	2. äkta	85	2. ha	26	2. se	68
3. hos	48	3. rik	57	3. kunna	31	3. älska ²	17
4. mot	54	4. vis	20	4. vilja	19	4. älska	17
5. åt	49	5. hederlig	25	5. vara	43	5. vora	24
6. över	24	6. gammal	97	6. veta	9	6. känna	19
7. emot	8	7. stackars	24	7. få	14	7. exploatera	5
8. bland	10	8. rättfärdig	10	8. med utseende	7	8. förmedla	5
9. till	98	9. fattig	21	9. vilga	7	9. mörda	7
10. per	7	10. död	18	10. i guldhjälm	4	10. finna	22
11. med	132	11. svensk	35	11. bli	12	11. locka	7
12. liksom	6	12. trovärdig	7	12. ära	4	12. äro	30
13. utav	4	13. fyrtioårig	5	13. med skönhetsinne	4	13. träffa	9
14. gentemot	3	14. klok	10	14. med hjärta	5	14. bedra	5
15. inför	6	15. obetydlig	11	15. göra	12	15. känna ²	12

Medan mannen är aktiv och *arbetar*, *ser* och *säger*, så *älskar* och *känner* kvinnan.



Tekniken går framåt: mp3 har ersatt CD:n som i sin tur ersatt kassetbanden. Låt oss ta en titt på hur denna utveckling ser ut i Korps textsamling med hjälp av ett trenddiagram, som visar trender i ords och andra språkliga fenomenens användning över tid.

Här kan vi se att kassetbanden har funnits sedan 1975 och har använts fram till runt år 1995. CD:n började ta fart år 1990 och var en stor succé runt år 2000. Därefter tar mp3:n vid.

Inför valet 2010 så bildade de borgerliga partierna Alliansen, vilket utmynnade i ett gemensamt partiprogram vid sidan av partiernas individuella program. Låt oss jämföra delarna med helheten. Till vänster återfinns de ord som tonats ned i det gemensamma partiprogrammet, och till höger ser vi ord som lyfts fram, jämfört med de individuella programmen.

Utmärkande för M+C+FP+KD_subst (2010)	Utmärkande för Alliansen_subst (2010)
centerparti	allians
folkparti	sid
region	mandatperiod
rättighet	förslag
jobb	överskott
högskola	familjepolitik
ersättning	grund
lägkonjunktur	utredning
energi	krona

Swe-Clarin

Termen eVetenskap används ofta för att beteckna vetenskap som med hjälp av modern informationsteknik kan angripa problem av en omfattning och komplexitet som inte skulle kunna hanteras utan elektroniska hjälpmedel och möjligheten att dra fördel av geografiskt spridda resurser. Korpuslingvistik (som med hjälp av språkteknologi bedriver forskning på stora insamlade textmassor) är ett mycket tidigt exempel på eVetenskap. Språkteknologi är också ett område som utvecklats starkt över femtio år, och den teknologi och teoribildning som utvecklats i anslutning till detta forskningsfält har idag fått en rik flora av applikationer utanför den rent lingvistiska intressesfären.

Det stora europeiska infrastrukturprojektet CLARIN (Common Language Resources and Technology Infrastructure) syftar till att göra digitala språkresurser (textsamlingar, inspelningar av ljud och bild, lexikon och så vidare), tillsammans med de språkteknologiska verktyg som behövs för att hantera dem, tillgängliga för forskare inom alla discipliner. Särskild tonvikt läggs på forskare inom humaniora och samhällsvetenskap. Med andra ord: CLARIN erbjuder en möjlighet för alla forskare inom humaniora och samhällsvetenskap (och även andra discipliner som t.ex. medicin och vårdvetenskap) att använda sig av eVetenskapliga metoder i sin verksamhet. För mer information om det europeiska samarbetet, se <<http://www.clarin.eu/>>.

När CLARIN nu efter en inledande förberedelsefas konsolideras till ett permanent tillstånd som ett europeiskt konsortium för forskningsinfrastruktur skall projektet Swe-Clarin möjliggöra för Sverige som land, och för Sveriges språkteknologiska forskargemenskap, att dra sitt strå till stacken i det europeiska samarbetet och förstås specifikt att arbeta med utveckling och underhåll av resurser och verktyg för svenska och de officiella svenska minoritetsspråken.

Swe-Clarin bygger upp den nationella infrastrukturen inom området i form av ett antal centra, baserade på väletablerade forsknings- och utvecklingsmiljöer. Dessa centra skall dels skapa, uppgradera och underhålla de digitala resurserna och verktygen, och dels fungera som utbildnings-, rådgivnings- och stödenheter för forskare utanför språkteknologigemenskapen. Med en sådan infrastruktur på plats, enkelt tillgänglig för landets forskare i deras dagliga verksamhet, hoppas vi att kunna erbjuda bra eVetenskapligt forskningsstöd.

Swe-Clarins kick-off

Swe-Clarin höll sin kick-off i början av februari i Göteborg. Information och mötesanteckningar finns att hitta på <http://sweclarin.se/swe/Swe-Clarin-kick-off>

Kalender

9–10 mars: Nordic Clarin Network-möte i Köpenhamn
17 april: HS-workshop i Göteborg. Alla är välkomna.
9–10 juni: Nordic Clarin Network-workshop
5–6 oktober: Nordic Clarin Network-workshop i anslutning till Språkbankens höstworkshop med tema historiska resurser
19–20 november: Swe-Clarin-stormöte i Stockholm, Swe-Clarin-workshop i anslutning på fredag em. Alla är välkomna.

Swe-Clarins första HS-workshop

Den 17 april håller Swe-Clarin en första workshop för humaniora och samhällsvetenskap (HS). Vårt arbete med Swe-Clarin syftar till att bygga verktyg och resurser för att främja forskning inom digital humaniora och samhällsvetenskap. Vår definition av HS är bred och vi välkomnar alla som arbetar, eller vill börja arbeta, med digitala, språkvetenskapliga verktyg för sin forskning inom humaniora, samhällsvetenskap, juridik, medicin, mfl. Denna workshop kommer primärt att hållas på svenska.

Tid och plats:

17 april, Lennart Torstensongatan 6, rum K333, kl 13–19
Anmälan på: <http://goo.gl/forms/tuchRuyjD9>
<http://sweclarin.se/>

Partners

Swe-Clarin har 9 partners från Lund, Göteborg, Linköping, Stockholm och Uppsala och inkluderar både universitet och myndigheter.
Hitta en lista och beskrivning av alla våra partners på: <http://sweclarin.se/swe/centrum>

Uppdateringar om Swe-Clarin via http://lists.sweclarin.se/mailman/listinfo/news_lists.sweclarin.se