



# Introduktion till Språkbankens historiska material i Korp

Joakim Lilljegen 2018

MEIJERBERGS INSTITUT

FÖR  
SVENSK ORDFORSKNING

# Innehåll

|                                                                           |    |
|---------------------------------------------------------------------------|----|
| Inledning.....                                                            | 3  |
| Tidsavgränsning.....                                                      | 4  |
| Textavgränsning.....                                                      | 5  |
| OCR-läsning.....                                                          | 5  |
| Ord- och textattribut.....                                                | 6  |
| Lexikaliska resurser.....                                                 | 8  |
| Korpusar och korpussamlingar.....                                         | 8  |
| August Strindberg.....                                                    | 9  |
| Bellmans samlade verk.....                                                | 9  |
| Betänkande angående likformig uppställning av grammatiska läroböcker..... | 10 |
| Ekeblads brev.....                                                        | 10 |
| Fornsvenska textbanken.....                                               | 11 |
| Lagar.....                                                                | 11 |
| Prosa och verser.....                                                     | 12 |
| Nysvenska.....                                                            | 12 |
| Gustav Vasas brevproduktion .....                                         | 13 |
| Kubhist.....                                                              | 13 |
| Kvinnotidningar.....                                                      | 15 |
| Litteraturlbanken.....                                                    | 15 |
| Medeltidsbrev, Svenskt diplomatarium.....                                 | 16 |
| Nils Matsson Kiöpings resor.....                                          | 17 |
| Runeberg.....                                                             | 17 |
| Svensk prosafiktio n 1800–1900.....                                       | 18 |
| Åkerbruk och gödsel.....                                                  | 18 |
| Äldre biblar.....                                                         | 19 |
| Äldre finlandssvenska .....                                               | 20 |
| Äldre lagtexter.....                                                      | 22 |
| Äldre svenska romaner.....                                                | 23 |

# Inledning

Med Språkbankens verktyg [Korp](#) kan man söka i stora textsamlingar (över 13 miljarder token). Materialet spänner över tretton olika sekler och kan därför vara till stor nytta i olika språkhistoriska forskningsprojekt. Korpusarna där det historiska materialet ingår har emellertid mycket varierande egenskaper och kvalitet, eftersom det digitaliserats med olika metoder, vid olika tidpunkter och med olika syften. Själva källtexterna är också av så pass skiftande karaktär (allt från utgåvor av handskrivna medeltida landskapslagar till tryckta dagstidningar från slutet av 1800-talet) att de lämpar sig för olika sorters språkhistoriska undersökningar.

I denna introduktion presenteras de olika korpusarna som ingår i Korps historiska material (före ungefär år 1900), vad det är för texter och hur de gjorts tillgängliga digitalt. De olika annoteringarna av korpusarna tas också upp, t.ex. om de är märkta med olika ord- eller textattribut, vilka möjliggör mer avancerade sökningar i korpusarna. För handledning att genomföra avancerade sökningar rekommenderas, förutom [användarhandledningen](#) på Språkbankens webbplats, Klas Hjortstams Att söka i Korp med CQP och RegExp.

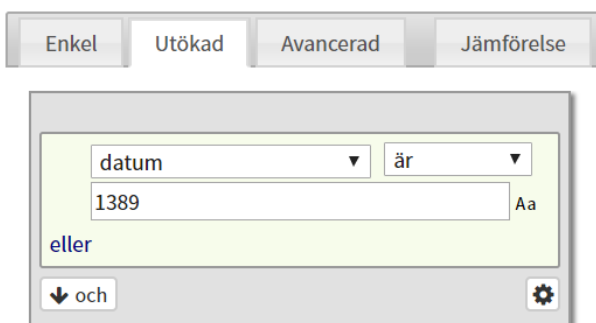
Eftersom arbetet med Språkbankens korpusar är ständigt pågående lämnas reservationer för framtida förändringar i Korp. Nyheter i Korp brukar presenteras på webbplatsen. Information om de olika korpusarna finns även dels kortfattat direkt i Korp (låt muspekaren vila över namnet på korpusen eller korpusamlingen så visas en informationsruta) och under rubriken [Resurser](#) på Språkbankens webbplats (där datum för senaste uppdatering och uppgifter om bl.a. tidsperiod och antalet token och meningar återfinns).

I detta dokument finns länkar till KB:s nätkatalog [Libris](#) med för att användaren lätt ska kunna hitta de tryckta utgåvorna av verken som digitaliserats. Länkar till digitala fulltextsversioner eller faksimil har också tagits med, liksom länkar till andra relevanta webbplatser.

Denna introduktion till Språkbankens historiska material i Korp har utarbetats under våren 2018 inom ett projekt finansierat av Meijerbergs institut för svensk ordforskning vid Göteborgs universitet. Ett stort tack riktas till institutets styrelse och till alla dem som delat med sig av upplysningar om de olika korpusarna som beskrivs här.

# Tidsavgränsning

Vissa korpusar omfattar bara ett enskilt år, medan andra sträcker sig över flera sekler. I Korp kan man under ”utökad sökning” göra avgränsningar utifrån tid, vilket är en mycket användbar funktion om det bara är material från vissa perioder man är intresserad av.



Korpusamlingarna *Fornsvenska textbanken*, *Medeltidsbrev*, *Svenskt diplomatarium* och *Kubhist* har variablerna ”datum” och ”tidsintervall”. För *Fornsvenska textbanken* finns inga dateringar mer detaljerade än enskilda årtal, så där är det inte meningsfullt att söka på djupare nivå än år, vilket skrivs enligt modellen ÅÅÅÅ, t.ex. 1442 (årtalet för Kristoffers landslag).

För *Medeltidsbrev*, *Svenskt diplomatarium* skrivs datumet enligt modellen ÅÅÅÅMMDD, t.ex. 14140528 för 28 maj 1414. Ibland finns bara uppgifter om år och månad eller enbart år, t.ex. 128206 för juni 1282 eller endast 1389 när ett brev inte kunnat dateras säkrare än årtalet.

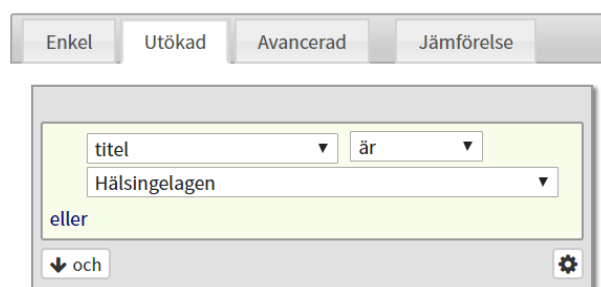
I *Kubhist* skrivs datumen med bindestreck mellan enheterna: ÅÅÅÅ-MM-DD, exempelvis 1855-10-27 för 27 oktober 1855.

För variabeln ”tidsintervall” väljer man en start- och sluttid. Dessa avgränsar sökresultatet så att enbart texter vars datum helt ingår i intervallet kommer med i resultatet. Väljer man t.ex. tidsintervallet 1270–1300 i samlingen *Fornsvenska textbanken* kan man få träffar i äldre västgöotalagen, då den är märkt med ”1280–1290” i Korp. Väljer man däremot tidsintervallet 1285–1300 utesluts äldre västgöotalagen, eftersom hela dateringen inte ingår i avgränsningen.

För andra korpusar, t.ex. *Litteraturbanken*, finns variabeln ”tryckår”. Här bör man vara uppmärksam på att själva textens ålder kan vara mycket äldre än tryckets. Om man exempelvis vill undersöka ett visst språkligt fenomen under de senaste två seklerna kan nytryckta texter av 1800-talsförfattare råka tolkas som belägg från 2000-talet. (Litteraturbanken arbetar med att revidera denna metadata så att både ”tryckår” och ”årtal för första utgivning” ska finnas som valbara variabler.)

# Textavgränsning

Under ”utökad sökning” kan man också välja ”titel” som variabel. För vissa korpussamlingar, t.ex. *Fornsvenska textbanken*, redovisas då samtliga tillgängliga titlar i en rullgardinsmeny varifrån man kan välja den text man vill ha med i sin sökning (eller utesluta genom att välja ”är inte”). Man kan också välja flera texter genom att klicka på ”eller” och kombinera flera olika sökkriterier. Också textattributet ”författare” som finns i vissa korpusar, bl.a. *Litteraturbanken*, kan användas för att avgränsa vilka texter man söker i.



## OCR-läsning

En stor del av det historiska materialet i Korp bygger på inskanningar eller fotokopior av tryckta sidor, där bildfilerna sedan maskinlästs med OCR-teknik (*optical character recognition*) för att skapa digitala texter. Resultatet kan variera av flera olika orsaker, t.ex. beroende på vilken programvara som använts och kvaliteten på det tryckta materialet som skannats in.

Vanligt förekommande fel kan t.ex. vara att *n* läses som *u* eller tvärtom. Således förekommer exempelvis *iuue* och *ntnr* flera gånger i Korp för det som stått som *inne* och *utur* i de tryckta originalen. Felläsningarna är inte jämt fördelade, varken inom eller mellan korpusarna. I korpussamlingen *Kubhist* förekommer t.ex. *inne* mer än dubbelt så mycket som *utur*. Ändå ger felläsningen *ntnr* mer än dubbelt så många träffar som *iuue*. Bokstaven *u* tycks alltså mycket oftare bli feltolkad som *n* än tvärtom. För olika strategier att hantera detta, se avsnitt 7.5 i Hjortstams handledning Att söka i Korp med CQP och RegExp.

# Ord- och textattribut

Korps historiska material består inte bara av själva texterna, utan det finns även en stor mängd information som är knuten till dem på olika nivåer. Exempelvis de ovan nämnda möjligheterna till avgränsningar utifrån tid och titel bygger på annoteringar som tillförts de olika korpusarna på textnivå. De enskilda orden kan i sin tur vara märkta med uppgifter såsom ordklass, vilket möjliggör sökningar på t.ex. ett visst verb följt av valfritt substantiv. Hos [Språkbanken](#) finns ett kort sammandrag över olika ordattribut. Se också del 4 i Hjortstams Att söka i Korp med CQP och RegExp för en mer utförlig introduktion om de olika ord- och textattributen.

Ett vanligt förekommande och på många olika sätt användbart ordattribut är den morfo-syntaktiska beskrivningen, förkortad ”msd” för *morphosyntactic description*. För denna variabel används en rad koder på två eller tre bokstäver för att beskriva orden, exempelvis ordklass, numerus, genus o.s.v. En lista över msd-koderna finns hos [Språkbanken](#) (en länk till denna sida finns också jämte varje msd-kod i det högra fältet i Korp när man markerar en sökträff).

Ett annat praktiskt ordattribut är *lemgram*, en term som myntats på Språkbanken och som innefattar samtliga böjningsformer av ett visst ord eller flerordsuttryck samt ordets sammansättningsform. Detta ordattribut går att välja också i enkel sökning. Istället för att direkt klicka på sök efter att ha skrivit t.ex. *stol* i sökfältet kan man vänta någon sekund och sedan få fram en lista med förslag på lemgram. Lemgrammen är också uppdelade mellan olika epoker; *fornsvenska*, *1600-tal*, *1800-tal* och nutida (de moderna har inte någon särskild markering, såsom det översta och näst näst översta exemplet i bilden till höger visar). För man muspekaren över alternativen får man en siffra på hur många träffar just det lemgrammet ger i ens markerade korpusar, t.ex. 103 förekomster för fornsvenska *stol (substantiv)* i samma exempel. För

The screenshot shows the Korp search interface with the word "stol" entered in the search field. The search results are displayed in a list, showing the word "stol" with its grammatical category (substantiv) and the number of occurrences (103) for the fornsvenska corpus. The results are organized by epoch (1800-tal, 1600-tal, fornsvenska) and grammatical category (substantiv, verb, e). The results are as follows:

| Epoch       | Grammatical Category           | Occurrences |
|-------------|--------------------------------|-------------|
| 1800-tal    | stol (substantiv)              |             |
| 1800-tal    | sticka under stol (verb)       |             |
| fornsvenska | stol <sup>2</sup> (e)          |             |
| fornsvenska | stol (substantiv)              | 103         |
| fornsvenska | stol <sup>2</sup> (substantiv) |             |
| fornsvenska | biskups stol (substantiv)      |             |
| 1600-tal    | stol (substantiv)              |             |
| 1600-tal    | stol <sup>2</sup> (substantiv) |             |
| fornsvenska | predika stol (substantiv)      |             |
| fornsvenska | predika stol (e)               |             |
| fornsvenska | bryggio stol (substantiv)      |             |

de fornsvenska texterna, där den stora stavningsvariationen gör det extra svårt för den automatiska annoteringen att pricka in rätt lemgram, finns även ordattributet ”närliggande lemgram” med ingångar till [Karps fornsvenska läge](#) (digitaliserade fornsvenska ordböcker, se vidare avsnittet ”Lexikaliska resurser” nedan). Om man i dessa korpusar markerar ett ord i sökresultatet i Korp visas i det högra fältet ett eller flera lemgram och ortografiskt närliggande lemgram. Användaren får själv avgöra vilket lemgram som är det riktiga.

Söker man i flera olika korpusar samtidigt presenteras alla ord- och textattribut som finns med i åtminstone en av de markerade korpusarna som valbara variabler under ”utökad sökning”. Väljer man att söka utifrån ett av dessa kriterier utesluts de korpusar som saknar den annoteringen. Om man exempelvis söker i flera korpusar efter verbet *röra* följt av valfritt substantiv får man noll träffar i *Medeltidsbrev*, *Svenskt Diplomatarium*, men detta beror inte på att relevanta förekomster saknas i den samlingen utan bara på att de inte är annoterade med uppgifter om ordklass.

För de flesta korpusar har annoteringen gjorts automatiskt, vilket innebär att den kan innehålla fel (särskilt om texten också innehåller OCR- eller korrekturfel). T.ex. kan homografer som substantivet *hem* och adverbet *hem* sammanblandas i den automatiska annoteringen. Att ett visst ord- eller textattribut finns med som valbar variabel för en korpus behöver heller inte betyda att texterna konsekvent är annoterade med den informationen.

I slutet av varje avsnitt nedan där Korps historiska samlingar beskrivs finns tabeller med uppgifter om vilka ord- och textattribut som är valbara för respektive korpus eller korpussamling.<sup>1</sup> Där står också huruvida korpusen stöder utökad kontext, vilket innebär att sökningar kan genomföras över meningsgränser. Korpusarnas texter är uppdelade i stycken och meningar, och vanligtvis genomförs sökningarna inom de enskilda meningarna.<sup>2</sup> I de korpusar som stöder utökad kontext kan man dessutom genomföra avancerade sökningar där t.ex. ord i slutet på en mening och i början på efterföljande mening tas med. Att söka på ord som står i olika stycken är däremot inte möjligt i det historiska materialet.

---

<sup>1</sup> Ett stort tack riktas till Klas Hjortstam som sammanställt dessa tabeller.

<sup>2</sup> I standardläget i Korp, KWIC (*keyword in context*), presenteras sökresultaten meningsvis i en lista, där det man sökt efter står i fetstil centrerat till mitten, medan resten av meningen står i vanlig stil till vänster och höger. Trycker man på ”Visa kontext” redovisas istället hela stycket där sökträffen förekommer.

# Lexikaliska resurser

Till stöd för läsningen av de äldsta texterna i Korps material rekommenderas det fornsvenska läget hos Språkbankens lexikaliska resurser i [Karp](#). Där har [Schlyters](#) och [Söderwalls](#) fornsvenska ordböcker digitaliserats och tillgängliggjorts. De ordböckerna är också sökbara i [Fornsvensk lexikalisk databas](#), som även den utvecklats vid Språkbanken men inte längre underhålls.

I Karp finns även [Swedbergs Svensk ordabok](#), som bygger på Lars Holms utgåva av Jesper Swedbergs manuskript *Swensk Ordabok* (2009, även tillgänglig hos [Litteraturbanken](#)). Också [Dalins ordbok](#), en digitaliserad version av [Ordbok öfver svenska språket](#) (1850–1853), finns i Karp. Det är ur morfologierna till dessa två ordböcker (som för tillfället är under utarbetning vid Språkbanken) som 1600-tals- respektive 1800-talslemgrammen i Korp är hämtade.<sup>3</sup>

Ord ur svenska texter som är yngre än år 1521 beskrivs i [Svenska Akademiens ordbok](#) (SAOB). Utgivningen av ordboken påbörjades år 1893 på bokstav *a* och i skrivande stund har arbetet nått bokstaven *v*. Ord som börjar på de allra sista bokstäverna i alfabetet har därför ännu inte behandlats i SAOB.

## Korpusar och korpussamlingar

Nedan följer beskrivningar av de olika korpusarna och korpussamlingarna med historiskt material i Korp (äldre än ca år 1900). De presenteras i alfabetisk ordning utifrån namnen som de tilldelats i Korp. Totalt sett består de av över 1,35 miljarder token, ungefär en tiondel av Korps hela material, fördelade på 175 olika korpusar. De flesta av dessa korpusar har i Korp sammanförts i olika större samlingar och i de fallen är det namnet på samlingen som står som huvudrubrik nedan.

---

<sup>3</sup> Också ordattributet ”dalin-lemgram” bygger på Dalins ordbok, men detta är ett äldre attribut som inte längre underhålls av Språkbanken. Dess funktion har ersatts med 1800-talsvarianterna under ”lemgram”.



# August Strindberg

I samlingen *August Strindberg* finns två korpusar: *August Strindbergs samlade verk* och *August Strindbergs brev*. Den första av dem innehåller material från de 59 volymer av [nationalupplagan](#) som hade utgivits fram till år 2003. Hela detta material ingår också i korpusen *Litteraturbanken*, där även texter ur de senare volymerna återfinns.

Korpusen *August Strindbergs brev* bygger på Kerstin Dahlbäcks arbetsmaterial med samtliga tryckta och otryckta brev av Strindberg som var tillgängliga 1 augusti 1991. Materialet kan innehålla både korrekturfel och tolkningsfel utifrån de handskrivna förlagorna. Ett kommenterat urval av Strindbergs brev finns också i Dahlbäcks utgåva [Min eld är den största: Brev 1858–1912](#) (1999).

På grund av att denna korpusamling är bristfällig på olika sätt, rekommenderas hellre *Litteraturbanken* med avgränsningen ”författare är Strindberg, August” under ”utökad sökning” för den som vill genomföra korpusstudier på Strindbergs verk.

| August Strindbergs samlade verk |                                                                                                                        |
|---------------------------------|------------------------------------------------------------------------------------------------------------------------|
| Ordattribut                     | ordklass, msd, grundform, lemgram, dependensrelation, förled, efterled, uttryck, namntyp namnundertyp, namn, betydelse |
| Textattribut                    | författare, namn, namntyp, namnundertyp, sida, tidsintervall, titel, uttryck, verksnummer, år                          |
| Stöder utökad kontext?          | Ja, upp till stycke                                                                                                    |

| August Strindbergs brev |                                                                                                                        |
|-------------------------|------------------------------------------------------------------------------------------------------------------------|
| Ordattribut             | ordklass, msd, grundform, lemgram, dependensrelation, förled, efterled, uttryck, namntyp namnundertyp, namn, betydelse |
| Textattribut            | band, dag, författare, mottagare, månad, namn, namntyp, namnundertyp, uttryck, år                                      |
| Stöder utökad kontext?  | Ja, upp till stycke                                                                                                    |

## Bellmans samlade verk

Korpusen *Bellmans samlade verk* bygger på den s.k. standardupplagan, [Carl Michael Bellmans skrifter](#) (1921–2003). Digitaliseringen ger ett bearbetat intryck och tycks sakna de slags fel som ofta kan förekomma i OCR-tolkningar. Bellmans samlade verk finns också

digitaliserade hos [Bellmansällskapet](#), där faksimil och OCR-lästa texter i olika format tillgängliggjorts i samarbete med The Internet Archive. Texter av Bellman ingår även i korpusen *Litteraturbanken*, där de härrör från ytterligare en annan digitalisering.

| Bellmans samlade verk  |                                                                                        |
|------------------------|----------------------------------------------------------------------------------------|
| Ordattribut            | msd, grundform, lemgram, dalin-lemgram, betydelse, förled, efterled, dependensrelation |
| Textattribut           | författare, sida, titel                                                                |
| Stöder utökad kontext? | Ja, upp till stycke                                                                    |

## Betänkande angående likformig uppställning av grammatiska läroböcker

Korpusen bygger på en digitalisering som Göteborgs universitetsbibliotek genomfört av [Betänkande angående likformig uppställning af grammatiska för rikets allm. läroverk afsedda läroböcker jemte dertill hörande grammatiska utkast](#) (1882). Den tryckta utgåvan har skannats och OCR-lästs, men ej korrekturlästs.

| Betänkande angående likformig uppställning av grammatiska läroböcker |                                                                                                  |
|----------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|
| Ordattribut                                                          | grundform, ordklass, msd, lemgram, dalin-lemgram, dependensrelation, betydelse, förled, efterled |
| Textattribut                                                         | plats, sida, tidsintervall, titel, utgivare, år                                                  |
| Stöder utökad kontext?                                               | Ja, upp till stycke                                                                              |

## Ekeblads brev

Korpusen *Ekeblads brev* bygger på den digitala utgåvan [Breven till Claes 1639–1655](#), som i sin tur bygger på Sture Alléns textkritiska utgåva [Johan Ekeblads brev till brodern Claes Ekeblad, 1639–1655](#) (1965). Den digitala utgåvan och korpusen innehåller enbart brevtexterna, inte inledningen eller kommentarerna som finns i 1965 års tryckta utgåva.

Den digitala utgåvan har även överförs till Litteraturbanken, varför samma text också återfinns i korpusen *Litteraturbanken*. Breven finns också i en moderniserad och kommenterad utgåva, både i tryck och digitalt: [Breven till Claes: Om livet och hovet på Kristinas tid](#) (red. Sture Allén, 2004).

| Ekeblads brev          |                                    |
|------------------------|------------------------------------|
| Ordattribut            | ordklass, grundform, lemgram       |
| Textattribut           | datering, datum, författare, titel |
| Stöder utökad kontext? | Nej                                |

## Fornsvenska textbanken

[Fornsvenska textbanken](#) är ett projekt som Lars-Olof Delsing vid Lunds universitet drivit sedan början av 1990-talet. Där har olika vetenskapliga utgåvor av medeltidstexter, främst Collins och Schlyters [Samling af Sweriges gamla lagar](#) (SSGL, 1827–1877) och [Samlingar utgivna av Svenska fornskriftsällskapet](#) (SSFS, olika serier sedan 1844), skrivits av eller skannats in och OCR-lästs.

De äldsta digitaliseringarna härstammar från projektet Källtext som bedrevs vid Göteborgs universitet i slutet av 1980-talet, men sedan dess har [flera personer](#) vid olika institutioner bidragit till textsamlingen. Texterna är granskade och korrekturlästa i varierande omfattning, varför användare själva uppmanas att ”förvissa sig om att textens kvalitet är hållbar för den typ av undersökning hon/han vill göra”.

Det äldsta materialet i *Fornsvenska textbanken* finns också tillgängligt i fulltextversioner på Språkbankens [FSvReader](#). Där kan man markera orden i texten och få förslag på ordboksartiklar ur Söderwalls och Schlyters fornsvenska ordböcker till stöd för läsningen. Äldre västgötalagen finns även i en utgåva med parallelltexter på fornsvenska och moderniserad svenska samt faksimil av handskriften (red. Per-Axel Wiktorsson), tillgänglig hos [Litteraturbanken](#). Skånelagen, eller *Skånske Lov*, har också digitaliserats av [Det Danske Sprog- og Litteraturselskab](#), men då utifrån *Codex Runicus* (AM 28, till skillnad från Fornsvenska textbankens version som bygger på utgåvan av Holm B 76). Där finns dels faksimil av handskriften, dels en text där handskriftens runor transkriberats till latinska alfabetet.

### Lagar

Fornsvenska textbanken innehåller flera lagar från början av 1200-talet till mitten av 1400-talet. De flesta av dessa är på fornsvenska, medan skånelagen är på forndanska och gutalagen på forngutniska. Några lagar innehåller också inslag på latin (t.ex. inledningar och rubriker).

Texterna bygger till större delen på del 1–12 av [SSGL](#). Lejonparten av dessa är inskrivna för hand och korrekturlästa varierande antal gånger. Fragment C, D, E, H, M och N av östgötalagen är inskrivna manuellt utifrån Emil Olssons utgåva ([SSFS 41](#)). *Upplandslagen enligt Codex Holm B 199* och *Tillägg till upplandslagen* är inskannade utifrån Samuel Hennings utgåva ([SSFS 70](#)). Texten har sedan OCR-tolkats och korrekturlästs.

| Äldre lagar och Yngre lagar |                                                        |
|-----------------------------|--------------------------------------------------------|
| Ordattribut                 | homografmängd, grundform, lemgram, närliggande lemgram |
| Textattribut                | datum, heltext, språk, stad, tidsintervall, titel      |
| Stöder utökad kontext?      | Nej                                                    |

## Prosa och verser

*Fornsvenska textbanken* innehåller såväl religiös som profan prosa samt flera olika verser. Dessa bygger uteslutande på utgåvorna av [SSFS](#). Precis som med lagarna varierar digitaliseringsmetoderna; vissa är inskannade och OCR-lästa, andra inskrivna manuellt. Vissa texter har korrekturlästs, andra inte. Mer detaljerade uppgifter om varje text finns hos [Fornsvenska textbanken](#).

| Äldre religiös prosa, Yngre religiös prosa, Profan prosa och Verser |                                                        |
|---------------------------------------------------------------------|--------------------------------------------------------|
| Ordattribut                                                         | grundform, homografmängd, lemgram, närliggande lemgram |
| Textattribut                                                        | datum, tidsintervall, titel                            |
| Stöder utökad kontext?                                              | Nej                                                    |

## Nysvenska

I Fornsvenska textbanken ingår en avdelning som kallas [nysvenska annexet](#) som också överförts till Korp. Där ingår texter yngre än 1526, bl.a. ett par böcker ur Gustav Vasas bibel (1541) och Dalins *Then swänska Argus* (1732–1734, som även finns i två olika digitala versioner, åtkomliga via [Libris](#)). Är man endast intresserad av texter från fornsvensk tid bör man se till att denna delsamling är avmarkerad när man genomför sina sökningar i *Fornsvenska textbanken*.

| Nysvenska (fem korpusar) |                                                   |
|--------------------------|---------------------------------------------------|
| Ordattribut              | ordklass                                          |
| Textattribut             | datum, heltext, språk, stad, tidsintervall, titel |
| Stöder utökad kontext?   | Nej                                               |

# Gustav Vasas brevproduktion

Korpusen *Gustav Vasas brevproduktion* bygger på Riksarkivets utgåva av [Konung Gustaf den förstes registratur](#) (1861–1916) som Göteborgs universitetsbibliotek började digitalisera 2014. Fotokopierade versioner av den tryckta utgåvan är OCR-lästa och kan därför innehålla fel. Faksimil av den tryckta utgåvan för jämförelser finns åtkomliga via [Libris](#).

| Gustav Vasas brevproduktion |                                                                                        |
|-----------------------------|----------------------------------------------------------------------------------------|
| Ordattribut                 | msd, grundform, lemgram, dalin-lemgram, betydelse, förled, efterled, dependensrelation |
| Textattribut                | publicerat, tidsintervall, titel, år (slut), år (start)                                |
| Stöder utökad kontext?      | Ja, upp till stycke                                                                    |

## Kubhist

Samlingen *Kubhist* består av OCR-läst tidningstext, större delen från 1800-talet, men även stora delar från seklerna före och efter. Samlingen är indelad först utifrån tidningstitel, sedan utifrån tryckdecennium. *Kubhist* är den till textmängden överlägset största korpusamlingen bland det historiska materialet i Korp och den bygger på material som tillgängliggjorts genom projektet [Digidaily](#), ett samarbete mellan Kungl. biblioteket, Mittuniversitetet och Riksarkivet, där miljontals svenska tidningsnummer digitaliserats.

Texternas kvalitet är mycket varierande, bl.a. beroende på ojämnt tryck och fläckar på de fotokopierade förlagorna. Någon traditionell korrekturläsning av så stora mängder text har inte genomförts, men väl fortlöpande justeringar av OCR-inställningarna för att anpassa till de olika typsnitt och utformningar som använts i de olika tidningarna genom åren. På så vis har man kunnat anpassa mjukvaran för att undvika vanligt förekommande fel läsningar.

Många tidningsnummer inleds med märkliga haranger av tecken då OCR-läsningen försökt tolka vinjetter och dekorationer kring tidningsnamnen som bokstäver. Den digitala texten blir emellertid genast läsligare så fort den nått den del av förstasidan där artiklarna börjar. Se vidare avsnitt 7.5 i Hjortstams handledning Att söka i Korp med CQP och RegExp för närmare beskrivning av OCR-läsningen i *Kubhist*.

Genom KB:s databas [Svenska Dagstidningar](#) finns flera hundra tidningstitlar, bl.a. de som ingår i *Kubhist*, digitaliserade. Där är äldre material som saknar upphovsrätt fritt tillgängligt i både digitalt faksimil och OCR-text, medan de nyare, upphovsrättsskyddade tidningarna bara är åtkomliga från KB och några andra bibliotek. (Också finländska [Nationalbiblioteket](#) har stora samlingar digitaliserade svenskspråkiga tidningar.<sup>4</sup> En annan liknande databas är [Svensk-amerikanska tidningar](#), där ett tjugotal tidningar som gavs ut på svenska i USA åren 1859–2007 har digitaliserats.)

Materialet i *Kubhist* är försett med en mängd metadata, t.ex. uppgifter om tidningarnas pris, tryckort, utgivningsfrekvens, politiska tendens och publikationsperiod. Dessa uppgifter visas i fältet till höger när man markerar en sökträff. Informationen är hämtad från KB:s [Nya Lundstedt dagstidningar](#), en databas över tidningar som getts ut i Sverige fr.o.m. år 1900 och som bygger vidare på Bernhard Lundstedts [Sveriges periodiska litteratur](#) (1895–1902), en bibliografi över alla periodiska publikationer som utgavs i Sverige 1645–1899. Bara enstaka delar av alla tillgängliga uppgifter i denna databas är överförda till Korp, så för den som vill ha ännu mer upplysningar om de olika tidningarna går det att söka efter dem genom länkarna ovan.

| <b>Kubhist: Aftonbladet (en korpussamling med fyra korpusar)</b> |                                                                                             |
|------------------------------------------------------------------|---------------------------------------------------------------------------------------------|
| Ordattribut                                                      | grundform, ordklass, lemgram, dalin-lemgram, dependensrelation, betydelse, förled, efterled |
| Textattribut                                                     | ISSN, datum, sida, tidsintervall, titel                                                     |
| Stöder utökad kontext?                                           | Nej                                                                                         |

| <b>Kubhist: alla utom Aftonbladet (20 dagstidningar med var sin korpussamling, sammanlagt 78 korpusar)</b> |                                                                                                                                                                                                                                                                                 |
|------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Ordattribut                                                                                                | grundform, ordklass, lemgram, dalin-lemgram, dependensrelation, betydelse, förled, efterled                                                                                                                                                                                     |
| Textattribut                                                                                               | ISSN, antal sidor, datum, fullständig titel, kommentar, källa, licenshavare, ort, politisk tendens, publikationsfrekvens, publikationsperiod, publikationsdagar, publikationstyp, redaktör, sida, storlek, tidsintervall, titel, tryckt av, tryckt i, utgivare, utgåva, årspris |
| Stöder utökad kontext?                                                                                     | Nej                                                                                                                                                                                                                                                                             |

---

<sup>4</sup> Ur Nationalbibliotekets samlingar har också delar av materialet i korpussamlingen *Äldre finlandssvenska* hämtats.

# Kvinnotidningar

Samlingen *Kvinnotidningar* innehåller texter ur tidskrifter från 1880-talet till 1930-talet, tillgängliggjorda av [KvinnSam](#) vid Göteborgs universitetsbibliotek, som själva digitaliserat vissa av tidningarna, nämligen *Dagny*, *Idun* och *Morgonbris*. Tidningarna *Hertha* och *Rösträtt för kvinnor*<sup>5</sup> har digitaliserats av Autotext i Norberg, *Kvinnornas tidning* av Uppsala universitetsbibliotek och *Tidevarvet* av Kulturföreningen Fogelstad. Tidningarna är inskannade eller fotokopierade – de digitala faksimilen är tillgängliga via länken ovan. De har sedan OCR-tolkats, men ej blivit korrekturlästa.

| Kvinnotidningar        |                                                                                                  |
|------------------------|--------------------------------------------------------------------------------------------------|
| Ordattribut            | grundform, ordklass, msd, lemgram, dalin-lemgram, dependensrelation, betydelse, förled, efterled |
| Textattribut           | sida, tidsintervall, titel, år                                                                   |
| Stöder utökad kontext? | Nej                                                                                              |

## Litteraturbanken

[Litteraturbanken](#) är en ideell förening som tillgängliggör svensk skönlitteratur (både original och översättningar) och litteraturvetenskapliga texter. Föreningens medlemmar är Kungl. biblioteket, Kungl. Vitterhetsakademien, Språkbanken, Svenska Akademien, Svenska litteratursällskapet i Finland och Svenska Vitterhetssamfundet.

I korpusen *Litteraturbanken* ingår såväl de skönlitterära texterna från tidigt 1600-tal och framåt som de litteraturvetenskapliga texterna från senare tid. De olika texterna har överförts till digitalt format på olika sätt; de flesta har skannats in och sedan OCR-lästs, medan enstaka texter skrivits in manuellt. I vissa fall genomförs korrekturläsning, varvid texterna publiceras i Litteraturbanken som digital text (XML-filer) och som nedladdningsbara EPUB-filer. I de fall korrekturläsning inte genomförs, publiceras bara de digitala faksimilen med OCR-tolkningen som sökbar text. Enstaka texter (framför allt moderna litteraturvetenskapliga verk) har

---

<sup>5</sup> *Rösträtt för kvinnor* har också kopierats från KvinnSam till Projekt Runeberg, varifrån texten överförs till Korps samling *Runeberg* (dock utan sista årgången, 1919). Söker man i båda dessa samlingar riskerar man därför dubletter bland träffarna. Den i samlingen *Kvinnotidningar* är då att föredra eftersom den har samtliga årgångar.

överförts direkt från förlagets eller författarens digitala versioner, desamma som legat till grund för de tryckta utgåvorna.

Det angivna året för texterna gäller tryckår för boken som texten hämtats från. Således återfinns t.ex. många av August Strindbergs verk från perioden 1889–1913, den period då *Nationalupplagan* gavs ut. (Vid Litteraturbanken arbetar man för tillfället med att särskilja årtal för originalutgivning och senare tryckning, liksom för det skönlitterära och det litteraturvetenskapliga materialet.) Markerar man ett ord bland sökresultaten i Korp visas i fältet till höger bl.a. en direktlänk till fulltexten och/eller faksimilen i Litteraturbanken.

| Litteraturbanken       |                                                                                                                         |
|------------------------|-------------------------------------------------------------------------------------------------------------------------|
| Ordattribut            | ordklass, msd, grundform, lemgram, dependensrelation, förled, efterled, uttryck, namntyp, namnundertyp, namn, betydelse |
| Textattribut           | författare, källa, länk till boken, namn, namntyp, namnundertyp, sida, sidlänk, tidsintervall, titel, tryckår, uttryck  |
| Stöder utökad kontext? | Nej                                                                                                                     |

## Medeltidsbrev, Svenskt Diplomatarium

Inom serien [Svenskt diplomatarium](#) har svenska medeltidsdokument utgivits sedan 1829. [Riksarkivet](#) har varit huvudman för utgivningen sedan 1976. Efter inskanning, OCR-tolkning och korrekturläsning i början på 2000-talet gjordes samtliga dittills utgivna texter tillgängliga och sökbara i [SDHK](#) (Svenskt Diplomatariums huvudkartotek) vid årsskiftet 2004/2005. Kvarvarande OCR-fel rättas allteftersom de upptäcks. Texterna i nyutgivna häften tillförs i SDHK vid utgivningen.

Texterna har överförts till Korp där samlingen är indelad i fem olika korpusar utifrån språk; svenska, norska, tyska, latin och övriga (i den sista ingår bl.a. brev på danska, franska och italienska samt flerspråkiga brev med olika kombinationer av de ovan nämnda språken). Viss felsortering av språken förekommer, t.ex. finns enstaka svenska texter också med i korpusen *Medeltidsbrev – Latin*.

| Medeltidsbrev, Svenskt Diplomatarium (fem korpusar) |                                            |
|-----------------------------------------------------|--------------------------------------------|
| Ordattribut                                         | –                                          |
| Textattribut                                        | datum, heltext, språk, stad, tidsintervall |
| Stöder utökad kontext?                              | Nej                                        |



# Nils Matsson Kiöpings resor

Korpusen *Nils Matsson Kiöpings resor* består av två språkligt olika versioner av samma reseskildring. De är manuellt inskrivna av Erik Magnusson Petzell vid Institutet för språk och folkminnen utifrån [1674](#) och [1743](#) års utgåvor, men ej korrekturlästa.

En kommenterad edition med parallelltexter ur samma utgåvor gavs ut av Ture Johannisson i serien [Svenska texter](#) 1961. Den första utgåvan av Kiöpings reseskildring från [1667](#) finns digitaliserad på Projekt Runeberg av Martin Rundkvist.

| Nils Matsson Kiöpings resor |                                                                                                   |
|-----------------------------|---------------------------------------------------------------------------------------------------|
| Ordattribut                 | msd, grundform, lemmgram, betydelse, förled, efterled, dependensrelation, typografisk information |
| Textattribut                | datum, kapitel, tidsintervall, titel                                                              |
| Stöder utökad kontext?      | Ja, upp till stycke                                                                               |

## Runeberg

[Projekt Runeberg](#) är ett ideellt initiativ för digitalisering av nordisk litteratur som p.g.a. ålder saknar upphovsrätt. Materialet från Projekt Runeberg som är överfört till denna samling i Korp utgörs av texter ur diverse tidningar och tidskrifter, de flesta från 1800-talet men vissa från 1900-talets början.<sup>6</sup> Texterna är inskannade och OCR-lästa och i de flesta fall inte korrekturlästa.

Samlingen består av korpusarna *Biblioteksbladet*, *Folkbiblioteksbladet*, *Rösträtt för kvinnor*,<sup>7</sup> *Svensk Tidskrift*, *Tiden*, *Ur Dagens Krönika* och *Diverse tidningar*. Den sistnämnda innehåller

---

<sup>6</sup> Också en del andra texter som digitaliserats inom Projekt Runeberg har överförts till Korp via projektet *Svenskan i Finland – i dag och i går*. Se korpussamlingen *Äldre finlandssvenska* nedan. Även korpusen *Bibeln 1917* som ingår i samlingen *Äldre biblar* bygger på digitalisering som genomförts inom Projekt Runeberg.

<sup>7</sup> *Rösträtt för kvinnor* finns också som korpus under samlingen *Kvinnotidningar*. Tidningen digitaliserades av KvinnSam vid Göteborgs universitetsbibliotek och har därifrån överförts till Projekt Runeberg. Korpusen under *Kvinnotidningar* har emellertid större textmängd och omfattar samtliga utgivna nummer, 1912–1919, medan den under *Runeberg* saknar sista årgången. Genomförs sökningar i båda korpussamlingarna rekommenderas därför att *Rösträtt för kvinnor* under *Runeberg* avmarkeras för att undvika dubletter.

texter ur *Brand*, *De ungas tidning*, *Det nya Sverige*, *Elegant*, *Hvar 8 dag*, *Nyare Conversations-Bladet*, *Sundsvalls tidning* och *Varia*. Faksimil, den OCR-lästa texten och uppgifter om vilka texter som korrekturlästs finns hos [Projekt Runeberg](#) (sök efter respektive titel i sökfältet).

| Runeberg               |                                                                                                  |
|------------------------|--------------------------------------------------------------------------------------------------|
| Ordattribut            | msd, grundform, lemgram, betydelse, förled, efterled, dependensrelation, typografisk information |
| Textattribut           | datum, tidsintervall, titel                                                                      |
| Stöder utökad kontext? | Ja, upp till stycke                                                                              |

## Svensk prosafiktion 1800–1900

Korpusen *Svensk prosafiktion 1800–1900* innehåller all skönlitteratur som utgivits i svenskt original vid sex olika årtal; 1800, 1820, 1840, 1860, 1880 och 1900. Texterna har sammanställts inom projektet [Spf1800–1900](#), en del av Litteraturbanken, och digitaliseringsmetoderna har varit liknande dem som använts för korpusen *Litteraturbanken*: Materialet är fotokopierat eller inskannat, därefter OCR-läst och sedan korrekturläst ”i den utsträckning som visar sig nödvändig”. På [sidan](#) där tillvägagångssättet för *Spf1800–1900* redovisas, konstateras att särskilt materialet från början på 1800-talet ofta genererar ett stort antal OCR-fel.

| Svensk prosafiktion 1800–1900 |                                                                                   |
|-------------------------------|-----------------------------------------------------------------------------------|
| Ordattribut                   | ordklass, msd, grundform, lemgram, betydelse, dependensrelation, förled, efterled |
| Textattribut                  | författare, källa, länk till boken, sida, sidlänk, tidsintervall, titel, tryckår  |
| Stöder utökad kontext?        | Nej                                                                               |

## Åkerbruk och gödsel

Samlingen består av två korpusar, *Åkerbruk* och *KVAH*, som är inskrivna manuellt och korrekturlästa av Lena Rogström vid Göteborgs universitet. Korpusen *Åkerbruk* innehåller texterna ur två böcker; Jacob Serenius [Den Engelska åker-mannen och fåra-herden](#) (1727) och Magnus Stridsbergs [En grundelig kundskap om swenska åkerbruket](#) (1727). Serenius bok har också getts ut i en digital faksimilutgåva av [Lunds universitetsbibliotek](#).

KVAH (Kungl. Vetenskapsakademiens handlingar) innehåller 18 artiklar om åkerbruk och gödsel från tidsperioden 1740–1778. Samtliga artiklar från KVAH under perioden 1739–1854 finns också digitaliserade hos [Devo](#).

| Åkerbruk och KVAH      |                                                                         |
|------------------------|-------------------------------------------------------------------------|
| Ordattribut            | msd, grundform, lemgram, betydelse, förled, efterled, dependensrelation |
| Textattribut           | datum, kapitel, tidsintervall, titel                                    |
| Stöder utökad kontext? | Ja, upp till stycke                                                     |

## Äldre biblar

Samlingen innehåller tre olika korpusar från tre olika sekler. Den äldsta, *Gustaf Vasas bibel – Nya testamentet*, är en OCR-läst digitalisering av Natan Lindqvists utgåva [Nya Testamentet i Gustaf Vasas Bibel under jämförelse med texten av år 1526](#) (1941). I korpusen ingår både Lindqvists förord och utgåvan av 1541 års bibelöversättnings nya testamente, med skillnaderna gentemot 1526 års nya testamente anförda i fotnoter.

| Gustav Vasas bibel – nya testamentet |                                                                                        |
|--------------------------------------|----------------------------------------------------------------------------------------|
| Ordattribut                          | msd, grundform, lemgram, dalin-lemgram, betydelse, förled, efterled, dependensrelation |
| Textattribut                         | publicerat, tidsintervall, titel, utgivare, år                                         |
| Stöder utökad kontext?               | Ja, upp till stycke                                                                    |

Korpusen *Bibel 1873* bygger på [Bibelen eller Den heliga skrift](#), en språkligt moderniserad utgåva av Karl XII:s bibel där de gammaltestamentliga apokryferna utelämnats. Digitaliseringen ger ett bearbetat intryck och tycks sakna de slags fel som ofta kan förekomma i OCR-tolkningar. Texten finns också tillgänglig hos Life.Churchs [YouVersion](#). (Karl XII:s bibel från 1703 finns hos Litteraturbanken, både som [text](#) och [faksimil](#).)

| Bibeln 1873            |                                                                        |
|------------------------|------------------------------------------------------------------------|
| Ordattribut            | ordklass, msd, grundform, lemgram, förled, efterled, dependensrelation |
| Textattribut           | kapitel, tidsintervall, titel, vers, år                                |
| Stöder utökad kontext? | Nej                                                                    |

Korpusen *Bibeln 1917* bygger på [Projekt Runebergs](#) digitalisering av 1917 års bibelöversättning. Den genomfördes 1994–1996 av ett tjugotal frivilliga som skrev in texten manuellt och korrekturläste. Resultatet utvärderas på projektets webbplats som ”förhållandevis gott”. De brister som senare upptäckts är främst att hänvisningar och fotnoter

inte tagits med i alla kapitel och att det finns vissa skillnader i bl.a. kommatering p.g.a. att personerna i projektet använde olika tryckta utgåvor. År 2005 publicerade Projekt Runeberg också en [digital fakismilutgåva](#) av 1917 års översättning, baserad på en utgåva tryckt 1927 av Evangeliska Fosterlandsstiftelsen. Texten finns också hos [Svenska Bibelsällskapet](#) (där även Bibel 2000 och delar av Karl XII:s bibel från 1703 återfinns).

| <b>Bibeln 1917</b>     |                                                                                   |
|------------------------|-----------------------------------------------------------------------------------|
| Ordattribut            | ordklass, msd, grundform, lemgram, betydelse, förled, efterled, dependensrelation |
| Textattribut           | kapitel, tidsintervall, titel, vers, år                                           |
| Stöder utökad kontext? | Nej                                                                               |

## Äldre finlandssvenska

Korpusarna som sorteras under *Äldre finlandssvenska* sammanställdes under projektet [Svenskan i Finland – i dag och i går](#) 2010–2017. Texterna sträcker sig över många olika genrer, t.ex. brev, dagböcker, skönlitteratur och tidningstexter. Många av texterna skannades in och OCR-lästes inom projektet, medan andra erhöles från olika håll, t.ex. från Nationalbiblioteket, Projekt Runeberg och Projekt Gutenberg. Allt material har korrekturlästs minst en gång.

Annoteringen på ord- och textnivå skiljer sig mycket åt för de olika korpusarna, vilket syns i de många inforutorna nedan. Ingen av korpusarna stöder utökad kontext.

| <b>Brev (fyra korpusar; 1700-tal, 1800–1849, 1850–1899 och 1900–1959)</b> |                                                                                   |
|---------------------------------------------------------------------------|-----------------------------------------------------------------------------------|
| Ordattribut                                                               | ordklass, msd, grundform, lemgram, betydelse, förled, efterled, dependensrelation |
| Textattribut                                                              | arkivkod, avsändare, datum, källa, mottagare, tidsintervall, titel                |

| <b>Dagböcker 1700-tal</b> |                                                                                   |
|---------------------------|-----------------------------------------------------------------------------------|
| Ordattribut               | ordklass, msd, grundform, lemgram, betydelse, förled, efterled, dependensrelation |
| Textattribut              | arkivkod, datum, källa, plats, titel                                              |

| <b>Dagböcker 1800–1849 och Frågebrevsvar 1900–1949</b> |                                                                                   |
|--------------------------------------------------------|-----------------------------------------------------------------------------------|
| Ordattribut                                            | ordklass, msd, grundform, lemgram, betydelse, förled, efterled, dependensrelation |
| Textattribut                                           | arkivkod, datum, källa, plats, tidsintervall, titel                               |

| <b>Protokoll vid lantdagen i Borgå år 1809</b> |                                                                                   |
|------------------------------------------------|-----------------------------------------------------------------------------------|
| Ordattribut                                    | ordklass, msd, grundform, lemgram, betydelse, förled, efterled, dependensrelation |
| Textattribut                                   | källa, tidsintervall, titel, år                                                   |

| <b>Sakprosa (fem korpusar; 1700–1749, 1750–1799, 1800–1849, 1850–1899 och 1900–1949)</b> |                                                                                   |
|------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|
| Ordattribut                                                                              | ordklass, msd, grundform, lemgram, betydelse, förled, efterled, dependensrelation |
| Textattribut                                                                             | författare, sida, tidsintervall, titel, url, utgivare, år                         |

| <b>Skönlitteratur (tre korpusar; 1800–1849, 1850–1899 och 1900–1959)</b> |                                                                                   |
|--------------------------------------------------------------------------|-----------------------------------------------------------------------------------|
| Ordattribut                                                              | ordklass, msd, grundform, lemgram, betydelse, förled, efterled, dependensrelation |
| Textattribut                                                             | författare, tidsintervall, titel, utgivare (endast för 1900–1959), år             |

| <b>Tidningstexter (elva korpusar)</b> |                                                                                   |
|---------------------------------------|-----------------------------------------------------------------------------------|
| Ordattribut                           | ordklass, msd, grundform, lemgram, betydelse, förled, efterled, dependensrelation |
| Textattribut                          | datum, tidsintervall                                                              |

| <b>Tidskrifter 1800–1849: Spanska Flugan</b> |                                                                                   |
|----------------------------------------------|-----------------------------------------------------------------------------------|
| Ordattribut                                  | ordklass, msd, grundform, lemgram, betydelse, förled, efterled, dependensrelation |
| Textattribut                                 | tidsintervall, titel, år                                                          |

| <b>Tidskrifter 1850–1899: Ateneum</b> |                                                                                   |
|---------------------------------------|-----------------------------------------------------------------------------------|
| Ordattribut                           | ordklass, msd, grundform, lemgram, betydelse, förled, efterled, dependensrelation |
| Textattribut                          | tidsintervall, titel, utgåva, år                                                  |

| <b>Tidskrifter 1850–1899: Filosofia.fi 1850–1899</b> |                                                                                   |
|------------------------------------------------------|-----------------------------------------------------------------------------------|
| Ordattribut                                          | ordklass, msd, grundform, lemgram, betydelse, förled, efterled, dependensrelation |
| Textattribut                                         | författare, källa, tidsintervall, titel, url, år                                  |

| <b>Tidskrifter 1850–1899: Finsk tidskrift 1878–1899, Landtmannen 1877–1879, Litterär tidskrift utgifven i Helsingfors, Typografitidning 1889–1890</b> |                                                                                   |
|-------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|
| Ordattribut                                                                                                                                           | ordklass, msd, grundform, lemgram, betydelse, förled, efterled, dependensrelation |
| Textattribut                                                                                                                                          | tidsintervall, utgåva, år                                                         |

| <b>Tidskrifter 1850–1899: Typografiskt minnesblad 1891</b> |                                                                                   |
|------------------------------------------------------------|-----------------------------------------------------------------------------------|
| Ordattribut                                                | ordklass, msd, grundform, lemgram, betydelse, förled, efterled, dependensrelation |
| Textattribut                                               | tidsintervall, år                                                                 |

| <b>Tidskrifter 1900–1959: Astra 1920–1959, Argus 1907–1911, Euterpe 1900–1905, Finsk Tidskrift 1900–1912, Husmodern 1903–1912</b> |                                                                                   |
|-----------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|
| Ordattribut                                                                                                                       | ordklass, msd, grundform, lemgram, betydelse, förled, efterled, dependensrelation |
| Textattribut                                                                                                                      | tidsintervall, utgåva, år                                                         |

| <b>Tidskrifter 1900–1959: Ateneum 1900–1903</b> |                                                                                   |
|-------------------------------------------------|-----------------------------------------------------------------------------------|
| Ordattribut                                     | ordklass, msd, grundform, lemgram, betydelse, förled, efterled, dependensrelation |
| Textattribut                                    | tidsintervall, titel, utgåva, år                                                  |

| <b>Tidskrifter 1900–1959: Filosofia.fi 1900–1959</b> |                                                                                   |
|------------------------------------------------------|-----------------------------------------------------------------------------------|
| Ordattribut                                          | ordklass, msd, grundform, lemgram, betydelse, förled, efterled, dependensrelation |
| Textattribut                                         | författare, källa, tidsintervall, titel, url, år                                  |

## Äldre lagtexter

Samlingen innehåller fyra olika korpusar: *Stockholms stads tänkeböcker*, *1734 års lag*, *1734 års förarbeten* och *Lagar från 1800-talet*. Den första av dem berör år 1626 och innehåller tre versioner av tänkeboken; en mer utförlig konceptversion, en stramare renskrift som var den juridiskt gällande texten, och en s.k. notariatversion som enbart innehåller material från årets första och sista kvartal (under ”utökad sökning” redovisas dessa som titlar). [Stadsarkivet](#) i Stockholm ger ut stadens tänkeböcker och utgåvan som innehöll 1626 års handlingar var [del 15](#) i serien (1990, red. Lars Wikström). Denna utgåva användes som underlag för digitaliseringen, som genomfördes inom projektet *Digital litteratur* år 1993 under ledning av Donald Broady vid KTH. 1626 års tänkebok behandlas också i Theresia Petterssons doktorsavhandling [Stockholms stads tänkeböcker: Funktionell texthistoria 1476–1626](#) (2017).

| Stockholms stads tänkeböcker (1626) |                                            |
|-------------------------------------|--------------------------------------------|
| Ordattribut                         | homografmängd, grundform, lemgram          |
| Textattribut                        | datum, marginalnoter, tidsintervall, titel |
| Stöder utökad kontext?              | Nej                                        |

Korpusen *1734 års lag* utgörs av balkarna i själva lagtexten, förordet samt domarreglerna. Materialet är inskrivet manuellt och korrekturläst, men en del fel finns fortfarande kvar. Korpusen *1734 års förarbeten* utgörs av en digitalisering av Vilhelm Sjögrens utgåva [Förarbetena till Sveriges rikes lag 1686–1736](#) (1900–1909). Materialet är OCR-läst med manuell efterarbetning.

| 1734 års lag och 1734 års förarbeten |                                             |
|--------------------------------------|---------------------------------------------|
| Ordattribut                          | grundform, lemgram, typografisk information |
| Textattribut                         | datum, tidsintervall, titel                 |
| Stöder utökad kontext?               | Nej                                         |

Korpusen *Lagar från 1800-talet* består av regeringsformen 1809 med ändringar 1809–1974. Denna bygger på en digitalisering av [Sveriges konstitutionella urkunder](#) (SNS 1999) som ger ett bearbetat intryck och tycks sakna de slags fel som ofta kan förekomma i OCR-tolkningar.

| Lagar från 1800-talet  |                                                                |
|------------------------|----------------------------------------------------------------|
| Ordattribut            | homografmängd, grundform, lemgram, betydelse, förled, efterled |
| Textattribut           | datum, marginalnoter, tidsintervall, titel                     |
| Stöder utökad kontext? | Nej                                                            |

# Äldre svenska romaner

Korpusen *Äldre svenska romaner* består av 56 stycken OCR-lästa romaner som Språkbanken erhållit genom samarbete med Textalk AB i Mölndal. Romanerna är skrivna av fjorton olika författare från 1830-talet till 1940-talet. En sammanställning över vilka romaner som ingår finns hos [Språkbanken](#). Texterna är stavningskontrollerade med hjälp av ett vid Språkbanken utvecklat system och därefter manuellt korrigerade av Textalk AB.

| Äldre svenska romaner  |                                                                                                                         |
|------------------------|-------------------------------------------------------------------------------------------------------------------------|
| Ordattribut            | ordklass, msd, grundform, lemgram, dependensrelation, förled, efterled, uttryck, namntyp, namnundertyp, namn, betydelse |
| Textattribut           | författare, namn, namntyp, namnundertyp, tidsintervall, titel, uttryck, år                                              |
| Stöder utökad kontext? | Ja, upp till stycke                                                                                                     |