

SPRÅKBANKENTEXT

En forskningsinfrastruktur för språkliga data



SPRÅKBANKEN FÖR NYBÖRJARE

2023

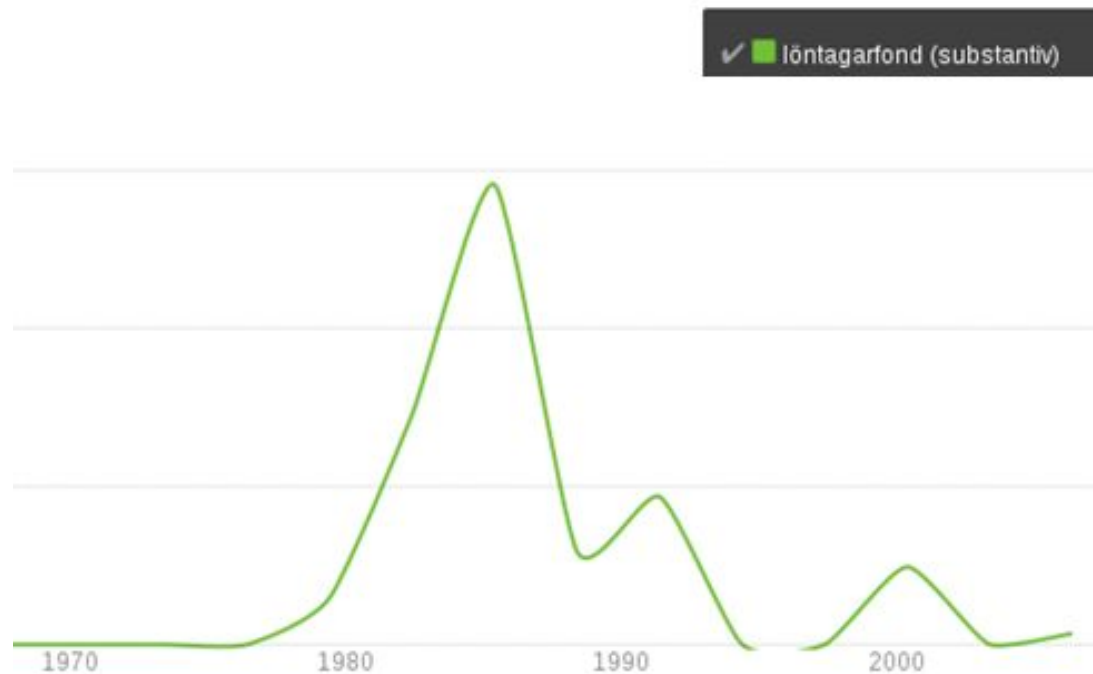
INNEHÅLL

- Vad är Språkbanken Text?
- Vår data
- Våra plattformar
- Exempel på tillämpningar
- Annat vi gör
- Vad vi kan hjälpa till med

BLOGGMIX 2005 (stödjer ej utökad kontext)
emien, Kungliga biblioteket och Språkbanken i Göteborg.
BLOGGMIX 2006 (stödjer ej utökad kontext)
För svenska har vi språkbanken – en stor gratis tjänst som måste vara gudarnas största gåva till mänskligheten kräftas empiriskt så jag kollade i Språkbanken och om man ska tro på deras data så var det faktiskt så att formen informatione Här är en länk till Språkbanken .
BLOGGMIX 2007 (stödjer ej utökad kontext)
Detta bekräftas om man går till Språkbanken : utan dess like finns överhuvudtaget inte vare sig i de tidigare tidningskorpora
BLOGGMIX 2010 (stödjer ej utökad kontext)
De röda symbolerna är Språkbanken , de blå träffar i Dagens Nyheter på Mediearkivet. skorpusen Press65, som finns på Språkbanken , så finns där 46 förekomster av äta men ingen alls av kåka, och talar är mer än t atar och talar i Mediearkivet och Språkbanken , resultaten syns i diagrammet nedan.
BLOGGMIX 2011 (stödjer ej utökad kontext)
har därför skapat en gemensam språkbank med frivilliga medlemmar och förtroendevalda.

VAD ÄR SPRÅKBANKEN TEXT?

- En **bank** skapar vinst
- Språk**banken** skapar kunskap



VEM BEHÖVER SPRÅKBANKEN TEXT?

Alla forskare vars område har att göra med mänsklig kommunikation:

- datavetare
- filologer
- genusforskare
- historiker
- hälsoforskare
- kulturvetare
- lingvister
- litteraturvetare
- medievetare
- pedagoger
- psykologer
- socialantropologer
- språkteknologer
- statsvetare
- med flera!

DET FINNS FLERA SPRÅKBANKER

- **Nationella språkbanken** är en forskningsinfrastruktur som finansieras gemensamt av Vetenskapsrådet och 10 universitet och myndigheter.
- Den består av tre avdelningar: **Språkbanken Sam**, **Språkbanken Tal** och **Språkbanken Text**.
- Det här bildspelet berättar om Språkbanken Text, som arbetar med skriftliga språkliga data.


NATIONELLA
SPRÅKBANKEN
SPRÅKBANKENTEXT SPRÅKBANKENTAL
SPRÅKBANKENSAM SPRÅKBANKENTEXT

VÅR DATA

- Fler än 1000 korpusar
- Fler än 50 lexikon

Du kan söka bland och ladda ner dem på vår hemsida:

spraakbanken.gu.se/resurser

SPRÅKBANKENTEXTEnglish






[AKTUELLT](#) [BLOGG](#) [FORSKNING](#) [VERKTYG](#) [DATA](#) [FRÅGELÅDA](#) [OM OSS](#) [KONTAKT](#)

[Hem](#) / [Språkresurser](#)

Språkresurser

Här kan du leta och filtrera bland våra korpusar och lexikon. Du kan klicka på en rad för att få fram nedladdningsbara filer. Genom att klicka på Korp- eller Karp-ikonen kommer du direkt till gränssnittet där du kan utforska resursen.

Alla **Samlingar** **Korpusar** **Lexikon** **Tränings- och utvärderingsdata** **Modeller**

Namn eller beskrivning	Språk			
	- Alla -			
Resurs	Utforska	Typ	Språk	Ladda ner
1734 års förarbeten Material från lagkommissionen till 1734 års lag		Korpus	svenska	forarbeten1734.xml.bz2 2014-12-08 – 9.11 MB stats_FORARBETEN1734.txt 2015-06-25 – 4.11 MB
1734 års lag Materialet utgörs av balkarna i själva lagtexten, förordet samt domarreglerna. Materialet är inskrivet för hand och korrekturläst, men en del fel finns fortfarande kvar.		Korpus	svenska	lag1734.xml.bz2 2014-09-29 – 565.61 KB stats_LAG1734.txt 2015-06-25 – 589.55 KB
8 Sidor Nyhetsartiklar från 8 SIDOR.		Korpus	svenska	attasidor.xml.bz2 2023-08-16 – 145.14 MB stats_attasidor.csv 2023-08-16 – 4.7 MB
Akademisk ordlista Akademisk ordlista		Lexikon	svenska	ao.xml 2017-09-13 – 265.72 KB
Akademiska texter: Humaniora En korpus med akademiska texter		Korpus	svenska	sweachum.xml.bz2 2017-05-19 – 208.67 MB stats_SWEACHUM.txt 2017-05-21 – 25.63 MB

VAD ÄR EN KORPUS?

- En stor sökbar samling texter.
- För att utöka sökmöjligheter brukar korpusar vara försedda med uppmärkning, det vill säga olika slags lingvistisk information.
- Uppmärkningen (annoteringen) är ofta gjord av datorn och därmed inte felfri.

Annotering ger information:

- **Morfologisk:** *Boken* är ett substantiv, singular bestämd form, samma ord som *bok*, *böcker* och *böckerna*
- **Syntaktisk:** *Jag läser en bok*: *bok* är objekt
- **Semantisk:** Vilken betydelse av *bok* gäller: en skrift eller ett träd?
- **Metadata:** I vad slags text förekommer ordet? Vem skrev texten? Och när?

NÅGRA AV SPRÅKBANKENS KORPUSAR

- skönlitteratur
- tidningar
- tidskrifter
- sociala medier
- bloggar
- forum
- Twitter
- myndighetstexter
- Wikipedia
- finlandssvenska
- historiska texter
- fornsvenska
- nysvenska
- parallella texter
- samma texter på 25+ språk
- Gigaword
- en stor balanserad korpus av olika genrer av nusvenskan från 1950 och framåt

NÅGRA AV SPRÅKBANKENS LEXIKON

- SALDO: ett semantiskt och morfologiskt lexikon
- Hellquists svenska etymologiska ordbok
- Bliss-symbollexikon
- Schlyters och Söderwalls ordböcker över medeltidssvenskan
- attitydlexikon
- lånordslexikon
- Swesaurus (ett svenskt ordnät, ett svenskt begreppslexikon)

Våra lexikon är skapade för automatisk textanalys.

SPRÅKBANKENS VIKTIGASTE PLATTFORMAR

Korp 
Språkbankens ordforskningsplattform

- **Korp:** en sökmotor som ger tillgång till 30 miljarder ord i Språkbankens korpusar.

Karp 
Språkbankens lexikala plattform

- **Karp:** en plattform för att arbeta med våra lexikon.

Sparv 
Språkbankens analysplattform

- **Sparv:** en annoteringsplattform som kan berika texter med morfologisk, syntaktisk och semantisk information.

Strix 
Språkbankens textforskningsplattform

- **Strix:** en sökmotor som är dokumentcentrerad och kan ta större hänsyn till dokumentets semantiska innehåll.

Mink 
Språkbankens dataplattform

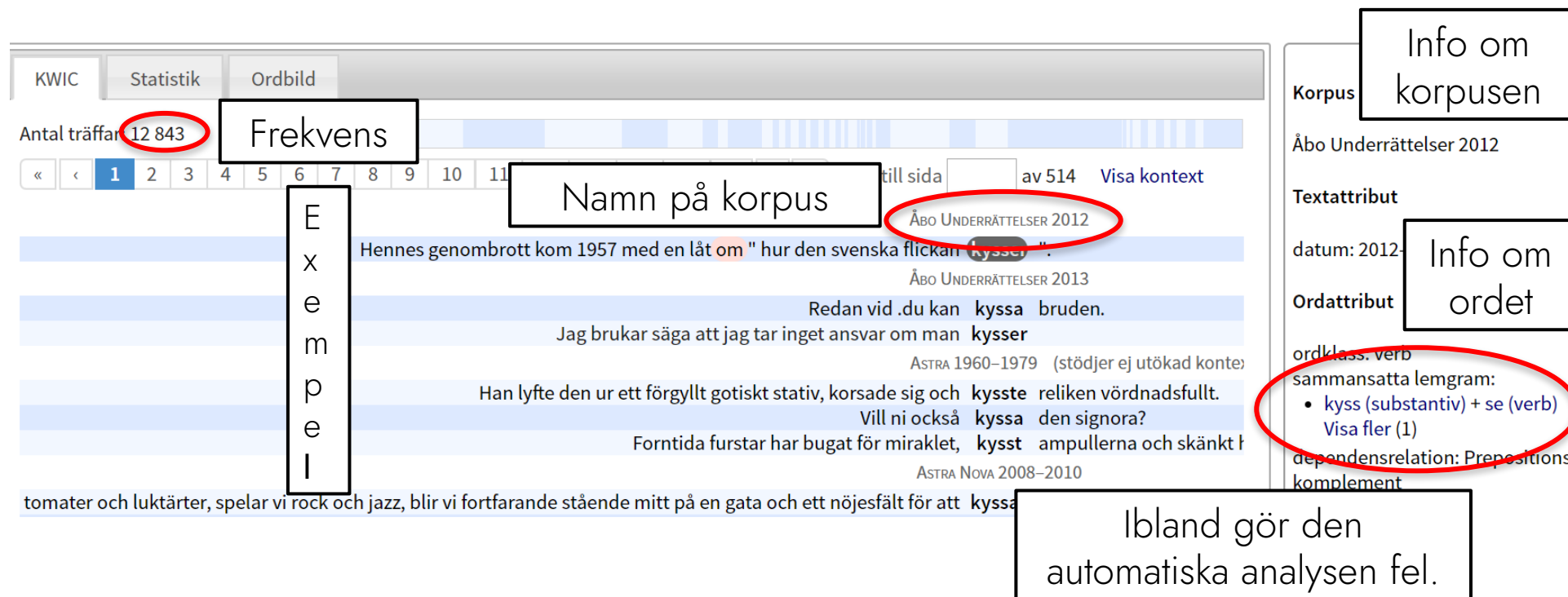
- **Mink:** en plattform där du kan tillämpa våra språkteknologiska metoder på texter som du själv har samlat in.

VAD ÄR KORP?

- en sökmotor
- avsedd för undersökningar av språkliga fenomen

ORDSÖKNING

- Ger information om ordets frekvens och visar alla dess förekomster i kontext. Ordet *kyssa*:



The screenshot shows the Korp search results for the word "kyssa". The interface includes a navigation bar with tabs for "KWIC", "Statistik", and "Ordbild". The "Statistik" tab is active, showing a bar chart of the word's frequency across different corpora. The search results are displayed in a table with columns for "Antal träffar" (12 843), "Frekvens", "Namn på korpus", and "Visa kontext". The word "kyssa" is highlighted in red in the search results. The right sidebar contains information about the corpus and the word, including the text "Info om korpusen", "Info om ordet", and "Ibland gör den automatiska analysen fel." (circled in red).

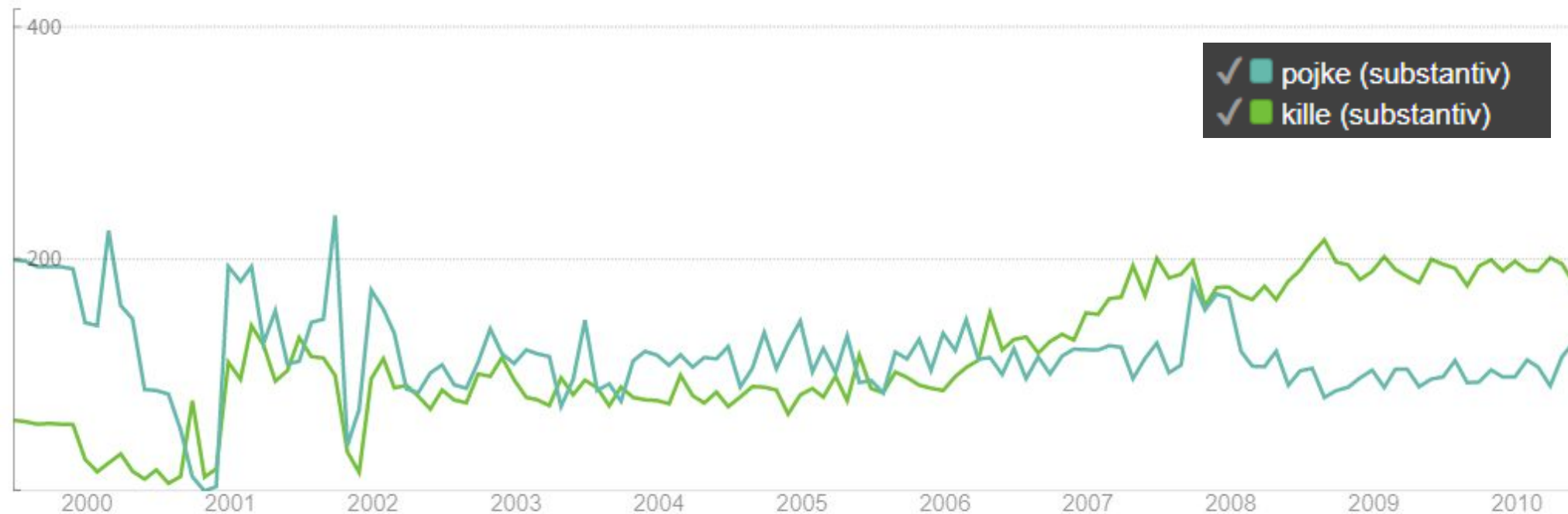
Info om korpusen

Info om ordet

Ibland gör den automatiska analysen fel.

TRENDDIAGRAM

■ Språkförändring:



JÄMFÖRELSE BASERADE PÅ SANNOLIKHET

- Vilka partier fokuserar på utbildning och vilka på sjukvård i sina program och manifest?

Utmärkande för *utbildning*

Centerpartiet	215
Vänsterpartiet	88
Piratpartiet	6
Moderata samlingspartiet	121
Folkpartiet liberalerna	181
Sveriges liberala parti	4
Lantmanna- och borgarepartiet	3

Utmärkande för *sjukvård*

Kristdemokraterna	62
Miljöpartiet de gröna	48
Ny demokrati	6
De rödgröna	6
Sverigedemokraterna	1
Alliansen	21
Sveriges socialdemokratiska arbetareparti	45



ORDBILDER


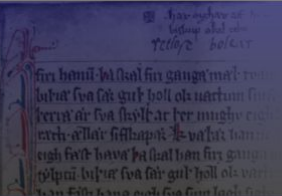


- Vem kysser vem/vad i Sverige och på vilket sätt?

Subjekt	kyssa	Objekt	Adverbial
1. pojke	41 📄	1. tjej	181 📄
2. främling	25 📄	2. främling	97 📄
3. kille	42 📄	3. fot	136 📄
4. man	46 📄	4. groda	93 📄
5. kvinna	39 📄	5. flicka	103 📄
6. hand	18 📄	6. hand	107 📄
7. kille ²	23 📄	7. kille	97 📄
8. tjej	26 📄	8. klubbmärke	45 📄
9. oskuld	10 📄	9. kille ²	73 📄
10. gaypar	8 📄	10. kompis	71 📄
11. gång ²	22 📄	11. kind	51 📄
12. hemlandshustru	6 📄	12. tomt	51 📄
13. sol	20 📄	13. tomt	51 📄
14. sol ²	20 📄	14. läpp	57 📄
15. påve	11 📄	15. farväl	48 📄
		1. på kind	86 📄
		2. på mun	72 📄
		3. i regn	52 📄
		4. i nacke	49 📄
		5. senast	63 📄
		6. på panna	33 📄
		7. på hand	33 📄
		8. första gången	38 📄
		9. gång ²	72 📄
		10. gång	55 📄
		11. gång ³	55 📄
		12. sen	68 📄
		13. i arsele	18 📄
		14. på hals	17 📄
		15. under vecka	25 📄

VAD ÄR KARP?

- Ett verktyg för att arbeta med Språkbankens lexikon.

 Svenska | English Välj en resurs nedan eller sök i Karps standardurval. 

<h2>SALDO</h2> <p>Semantiskt och morfologiskt lexikon för språkteknologi.</p>	 <h2>Dalin</h2> <p>Dalins ordbok - Ordbok över 1800-talspråket</p>
 <h2>Fornsvenska</h2> <p>Schlyter och Söderwall</p>	<h2>Konstruktikon</h2> <p>Svenskt konstruktikon är en fritt tillgänglig konstruktionsdatabas, dvs. en samling beskrivningar av svenska konstruktionsmönster. Dessa konstruktioner kan vara allt från väldigt generella strukturer till högst specifika uttrycksätt.</p>
 <h2>SweFN</h2> <p>Svenskt frasnät</p>	 <h2>Svenskt kvinnobiografiskt lexikon</h2>

SALDO


- Hur ser det semantiska grannskapet ut för ordet *hus*?












SALDO ▾ 77 TRÄFFAR (VISAR 25)							
	BETYDELSE	LEMGRAM	ORDKLASS	PRIMÄR	SEKUNDÄR	BARN (PRIMÄRA)	BARN (SEKUNDÄRA)
⚙	balkong	balkong (substantiv) 🐦 204656	substantiv	hus		piskbalkong	balkongdörr balkongdörr ² balkonglåda balkongräcke
⚙	bankhus	bankhus (substantiv) 🐦 3153	substantiv	hus	bank		
⚙	bod ²	bod (substantiv) 🐦 116910	substantiv	hus			bodlänga
⚙	bollhus	bollhus (substantiv) 🐦 1347	substantiv	hus	bollspel		

ETYMOLOGISK ORDBOK

- Från vilka språk har svenskan lånat mest?

Hitta ingångar där inte [etymon-språk](#) är lika med

och där också [etymon-språk](#) finns 


fra	200	
deu	114	
lat	88	
ita	48	
eng	40	
dan	33	
ell	33	
lty	23	
nor	21	
spa	17	
medellågtyska	13	

Sökfrågan: "hitta alla ingångar där det är angett vilket språk ordet kommer från, och det där språket inte är fornsvenska"

DIAKRONISK PIVOT

- Jag vill söka efter ordet *kvinna* i äldre texter, men jag vet inte riktigt hur det stavades under den period jag är intresserad av:

[kvinna \(substantiv\)](#)

 6395274

Dalin/1800-tal

Exakta träffar:

[qvinna \(substantiv\)](#)

 55743

[qvinno \(substantiv\)](#)

 121

Breda träffar:

[bondqvinna \(substantiv\)](#)

 80

[fästeqvinna \(substantiv\)](#)

 167

[hjelteqvinna \(substantiv\)](#)

 3

[dagsverksqvinna \(substantiv\)](#)

 0

[grannqvinna \(substantiv\)](#)

 156

Fornsvenska

Exakta träffar:

[grankuna \(substantiv\)](#)

 0

[granqvinna \(substantiv\)](#)

 0

BLISS

- Ett system av symboler där begreppen representeras av stiliserade bilder. Vilka Bliss-bilder motsvarar *farfar* och *mormor*?

Hitta ingångar [där](#) [vad som helst](#) [är lika med](#) eller

	BCI-AV#	GLOSS	SALDO	WORDNET
  	14626	grandfather, granddad, grandpa	grandfather granddad grandpa	farfar morfar
<hr/>   	14629	grandmother, grandma, granny	grandmother grandma granny	farmor mormor

MER OM KORP

- Låt oss söka efter ordet *grym*:

Olika betydelser!

	GP 2013	
- Nja, det vet jag inte, men Sean Banan är Berättelserna om tortyr, övergrepp, skräck blir just så obegripligt	grym	- som i verkligheten.
Vi tappade bollen - och de hade sådan	grymma	kvalitet i det som de gjorde.
Mika har gjort	grym	mycket poäng och Kari är en av allsvenskans bästa vänsterbackar, säger han.
n, Vägen mot Bålberget, ger sig Therése Söderlind i kast med detta	grymma	och svårbegripliga kapitel i Sveriges historia.
Häcken är en	grym	klubb, en skön klubb.
en varje månad ser arrangörerna till att ge besökarna tung bas och	grym	electro.
ndratals beagles till laboratorier i Storbritannien för att användas i	grymma	experiment när avelsanläggningen i Örkelljunga stänger ner, skriver hon.
Vi tappade bollen - och de hade sådan	grym	kvalitet i det som de gjorde.
GÖTEBORG:	Grymt	gott!
Och då är hon	grym	.
Det hade varit	grymt	att vinna två raka guld, samtidigt måste fokus ligga på varje enskild match.
Vecka efter vecka presenterar de den ena	grymma	bokningen efter den andra.
Som programledaren Timo Räisänen konstaterade: " Fatta vad	grymma	vi är på musik i det här landet! "
Den sista löpningen är	grym	och berör mig starkt.
ndratals beagles till laboratorier i Storbritannien för att användas i	grymma	experiment när avelsanläggningen i Örkelljunga stänger ner, skriver hon.
Det är	grymt	svårt att få en hyfsad storlek på dem.
glehundarna i England kommer att utsättas för experiment som är	grymmare	än de i Sverige.
nu långt kvar tills ridån går upp för kvällens föreställning, komedin	Grymt	galet på Skandiateatern i Norrköping.
Under EM 2006, vilken	grym	publik det var.
"	Grymt	!" säger han.
Man blir	grymt	effektiv att rulla bollar.
Häcken är en	grym	klubb, en skön klubb.
- Vi fick en pangstart på matchen, 73 poäng i slutspelsmatch är	grymt	bra, säger lagets coach Johanna Ericsson.
Tillsammans spelar de systrar i komedin	Grymt	Galet.

AUTOMATISK TEXTANALYS

■ Betydelseerna i Korp:

Text	Form	Betydelse
GP 2013		
- Nja, det vet jag inte, men Sean Banan är grym	grym	- som i verkligheten.
Berättelserna om tortyr, övergrepp, skräck blir just så obegripligt grymma	grymma	- som i verkligheten.
Vi tappade bollen - och de hade sådan grym	grym	kvalitet i det som de gjorde.
Mika har gjort grymt	grymt	mycket poäng och Kari är en av allsvenskans bästa vänsterbackar, säger han.
an, Vägen mot Bålberget, ger sig Therése Söderlind i kast med detta grymma	grymma	och svärbegripliga kapitel i Sveriges historia.
Häcken är en grym	grym	klubb, en skön klubb.
gen varje månad ser arrangörerna till att ge besökarna tung bas och grym	grym	electro.
indrattals beagles till laboratorier i Storbritannien för att användas i grymma	grymma	experiment när avelsanläggningen i Örskelljunga stänger ner, skriver hon.
Vi tappade bollen - och de hade sådan grym	grym	kvalitet i det som de gjorde.
GÖTEBORG: Grymt	Grymt	gott!
Och då är hon grym	grym	.
Det hade varit grymt	grymt	att vinna två raka guld, samtidigt måste fokus ligga på varje enskild match.
Vecka efter vecka presenterar de den ena grymma	grymma	bokningen efter den andra.
Som programledaren Timo Räisänen konstaterade: " Fatta vad grymma	grymma	vi är på musik i det här landet! "
Den sista löpningen är grym	grym	och berör mig starkt.
indrattals beagles till laboratorier i Storbritannien för att användas i grymma	grymma	experiment när avelsanläggningen i Örskelljunga stänger ner, skriver hon.
Det är grymt	grymt	svårt att få en hyfsad storlek på dem.
glehundarna i England kommer att utsättas för experiment som är grymmare	grymmare	än de i Sverige.
inu långt kvar tills ridån går upp för kvällens föreställning, komedin Grymt	Grymt	galet på Skandiateatern i Norrköping.
Under EM 2006, vilken grym	grym	publik det var.
" Grymt	Grymt	! " säger han.
Man blir grymt	grymt	effektiv att rulla bollar.
Häcken är en grym	grym	klubb, en skön klubb.
- Vi fick en pangstart på matchen, 73 poäng i slutspelsmatch är grymt	grymt	bra, säger lagets coach Johanna Ericsson.
Tillsammans spelar de systrar i komedin Grymt	Grymt	Galet.

Gå till sida av 19

Korpus

GP 2013

Textattribut

artikelförfattare: Torbjörn Skarhed
artikelavdelning: Kultur Nöje
datum: 2013-02-22

Ordattribut

ordklass: adjektiv
sammansatta lemgram: [tom]
dependensrelation: Subjektspredikativ (subjektiv predikatsfyllnad)
förled: [tom]
efterled: [tom]
betydelse:
• grym²
Visa fler (1)
msd: JJ.POS.UTR.SIN.IND.NOM ⓘ
sammansatta ordformer: [tom]
grundform:
grym
lemgram:
grym (adjektiv)
Visa dependensträd





AUTOMATISK TEXTANALYS

- Betydelseerna i Korp kommer från lexikonet SALDO:

GP 2013

- Nja, det vet jag inte, men Sean Banan är grym .
Berättelserna om tortyr, övergrepp, skräck blir just så obegripligt grymma - som i verkligheten.
Vi tappade bollen - och de hade sådana grym kvalitet i det som de gjorde.
Mika har gjort grymt mycket poäng och Kari är en av allsvenskans bästa vänsterbackar, säger han.
n, Vägen mot Bälberget, ger sig Therése Söderlind i kast med detta grymma och svårbegripliga kapitel i Sveriges historia.

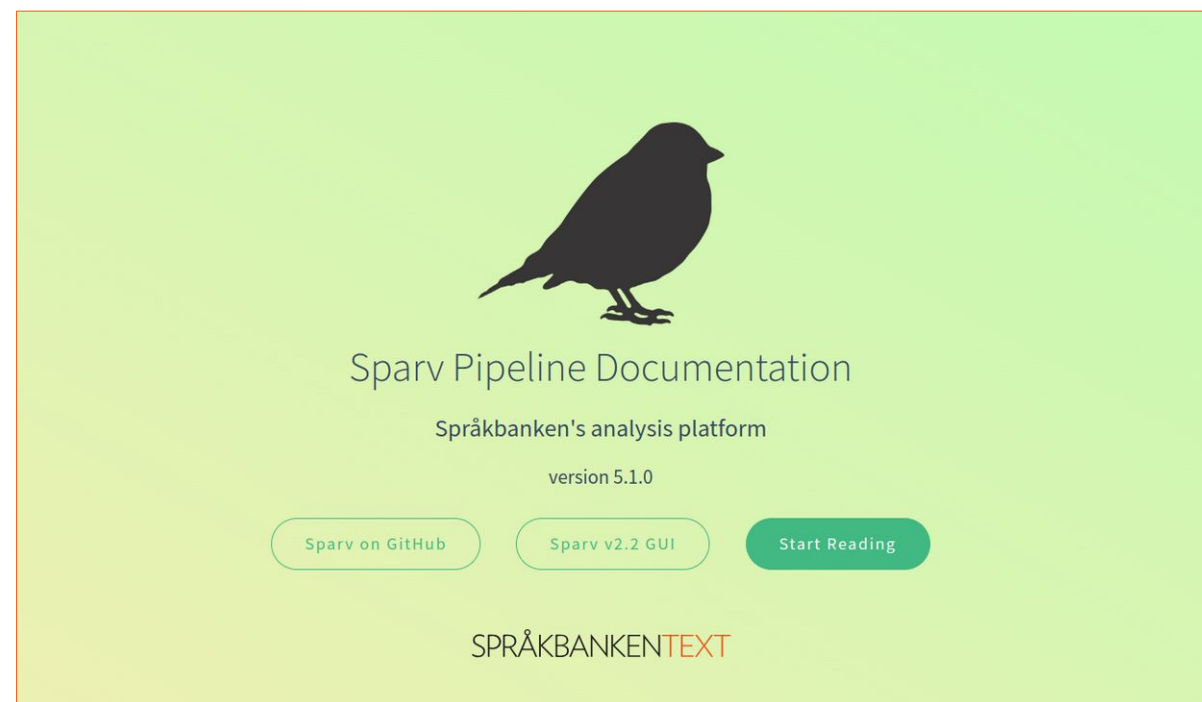
SALDO ▾ 2 TRÄFFAR (VISAR 2)

BETYDELSE	LEMGRAM	ORDKLASS	PRIMÄR	SEKUNDÄR	BARN (PRIMÄRA)	BARN (SEKUNDÄRA)
 grym	 grym (adjektiv) 1087340	adjektiv	hård		bestialisk grymhet kärlekslös	despot
 grym ²	 grym (adjektiv) 1087340	adjektiv	bra		grymhet ² lexikograf lexikolog	

VAD ÄR SPARV?

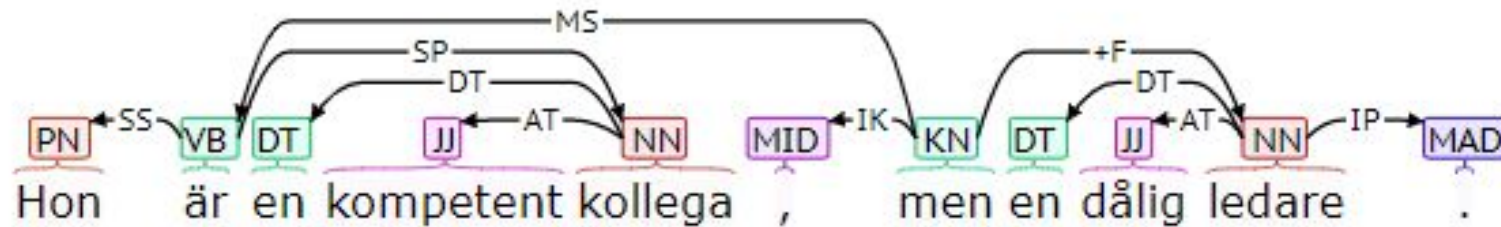
Sparv kan märka upp dina texter och förse dem med:

- lexikal analys
- syntaktisk analys
- semantisk analys
- attitydanalys
- läsbarhetsanalys
- och mer!



SYNTAKTISK ANALYS

Hon är en kompetent kollega, men en dålig ledare.



SEMANTISK ANALYS

Hon är en kompetent kollega, men en dålig ledare.

Ordet ledare kan ha minst tre betydelser:

Betydelse	Sannolikheten i denna mening (enligt Sparv)
chef	0,65
strömledare	0,21
tidningsartikel	0,14

SENTIMENTANALYS

Hon är en kompetent kollega, men en dålig ledare.

token	sentimentclass
Hon	
är	neutral
en	
kompetent	positive
kollega	neutral
men	
en	
dålig	negative
ledare	neutral

Sentimentanalys kallas även för attitydanalys.

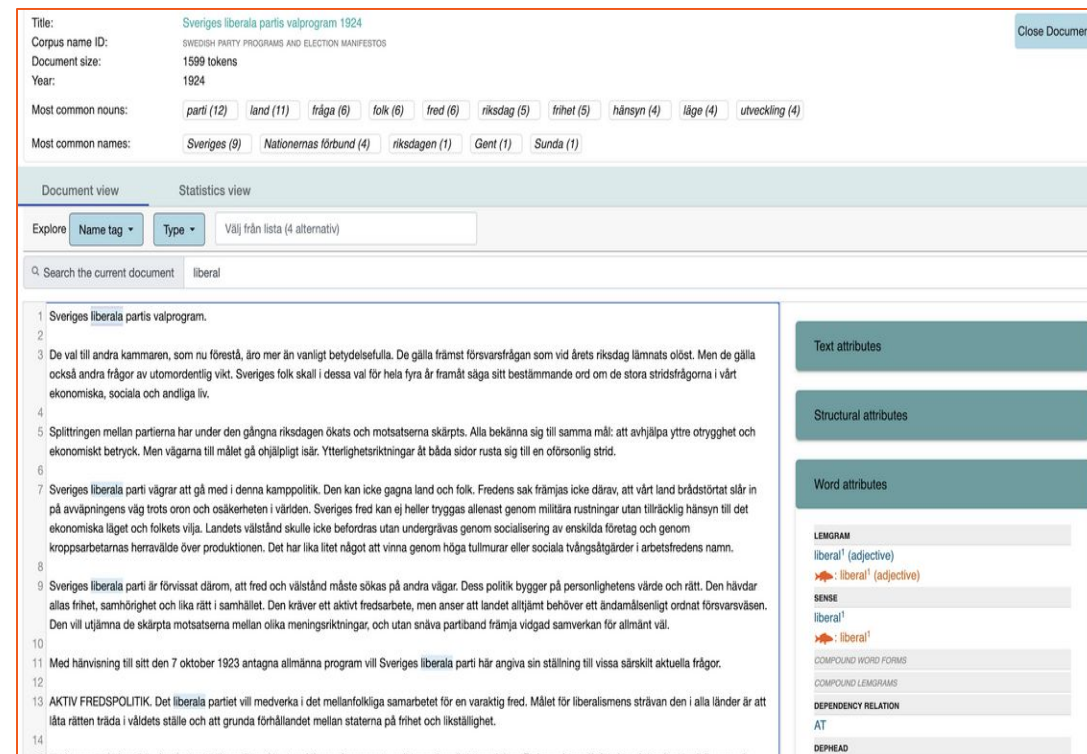
VAD ÄR STRIX?

En sökmotor som:

- inte fokuserar så mycket på enstaka ord (som Korp gör) utan på hela dokument
- tar hänsyn till dokumentets semantiska innehåll

VAD KAN STRIX GÖRA?

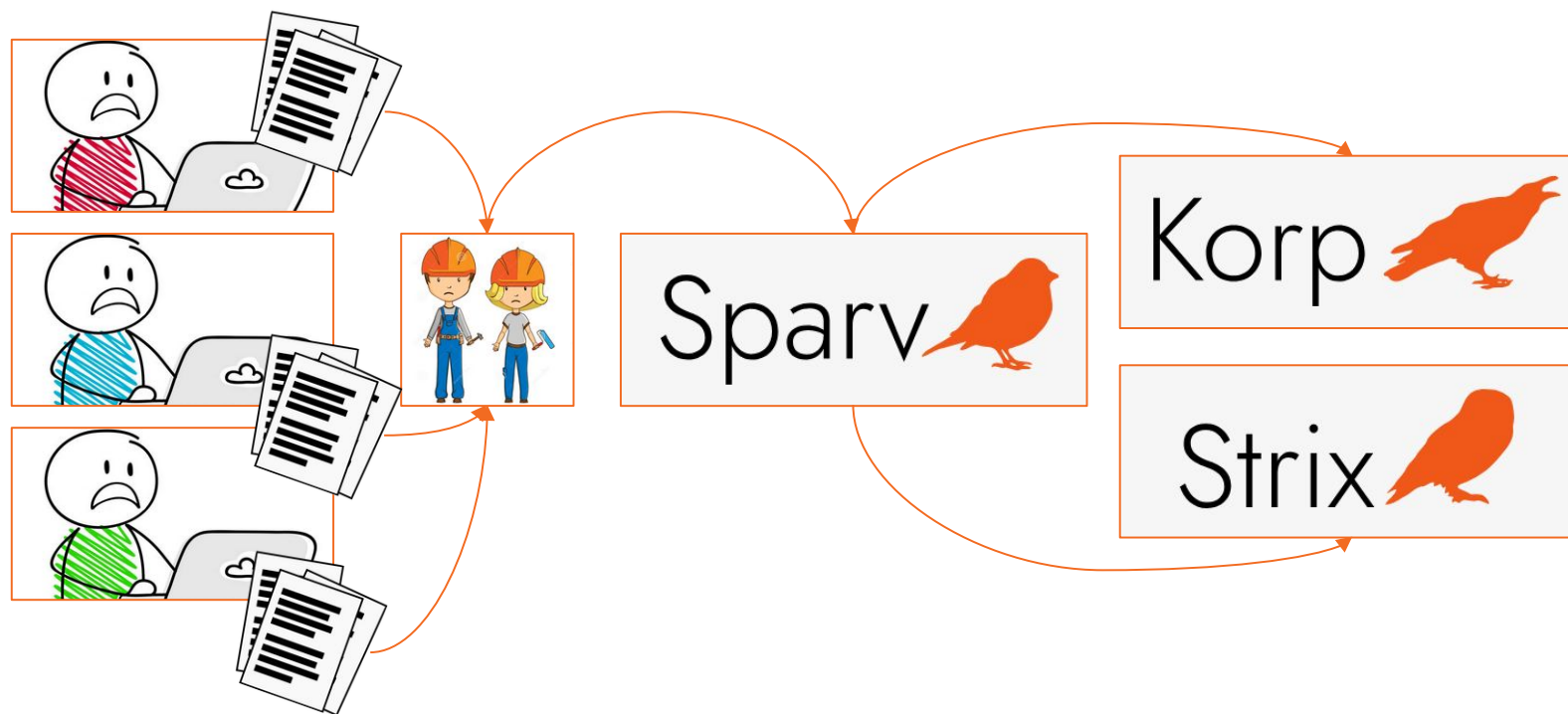
- textanalys
- ta fram nyckelord
- ta fram namn
- sentimentanalys
- ämnesklassificering
- visa hela annoterade dokument
- arbeta med dokumentsamlingar
- visa metadata



The screenshot displays the Strix web interface for a document titled "Sveriges liberala partis valprogram 1924". The interface includes a header with document metadata (Corpus name ID, Document size, Year) and a navigation bar with "Document view" and "Statistics view" tabs. Below the navigation bar, there are search and filter options, including a search bar with the query "liberal" and a "Name tag" dropdown. The main content area shows a list of search results for the word "liberal", with the first result being "Sveriges liberala partis valprogram." The right sidebar contains several panels for analyzing the word "liberal", including "Text attributes", "Structural attributes", and "Word attributes". The "Word attributes" panel shows the word "liberal" with its part of speech (adjective) and other linguistic information.

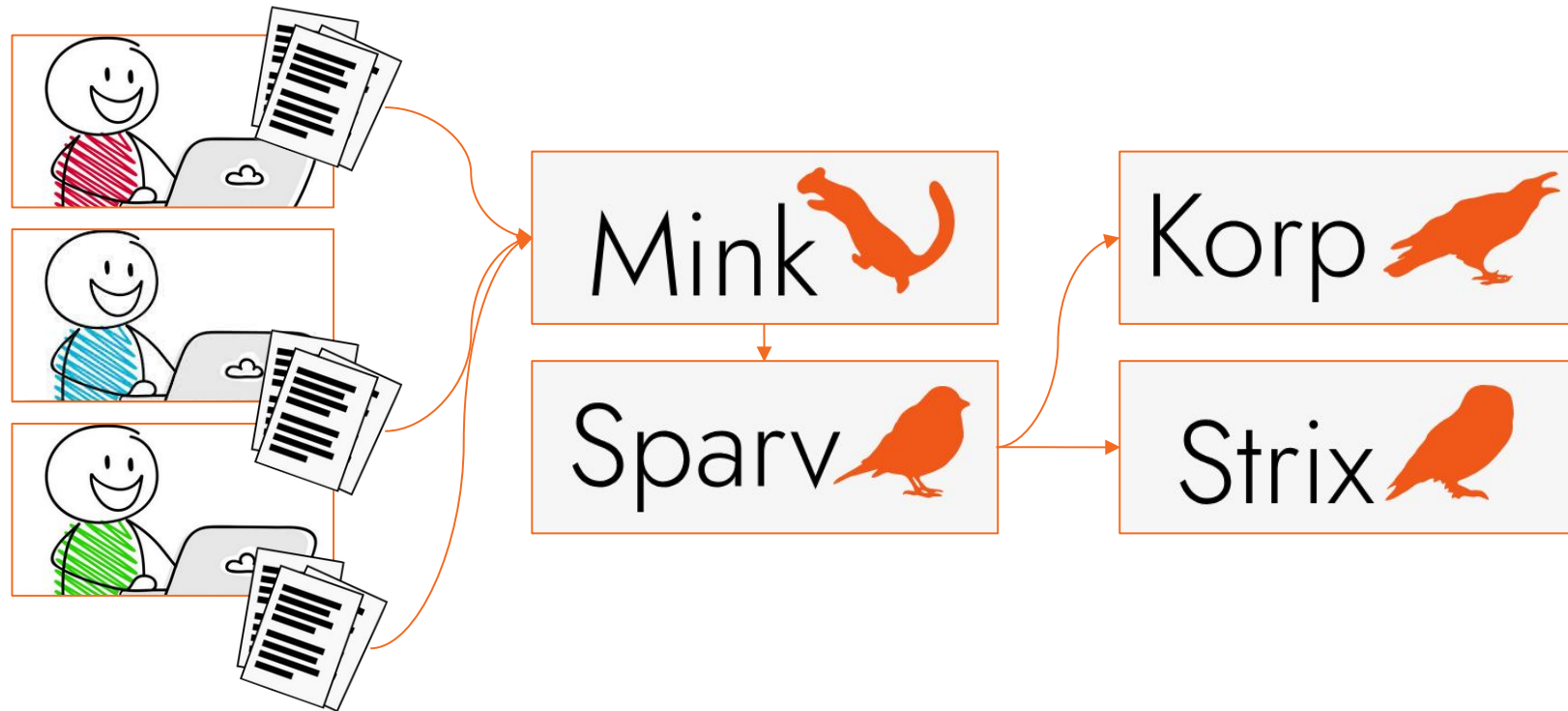
VAD ÄR MINK?

Att lägga till nya korpusar involverar Språkbanken Text.



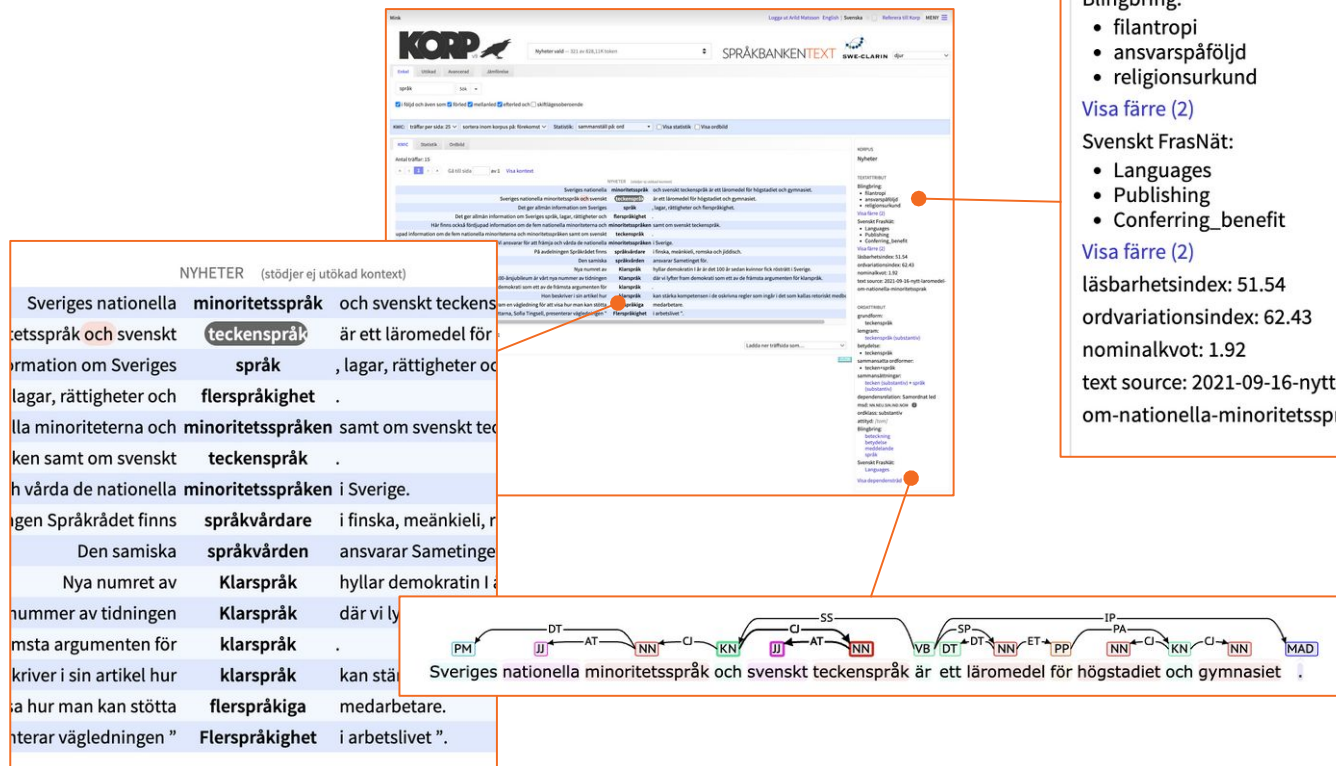
VAD ÄR MINK?

Med Mink kan du själv lägga till och analysera dina texter!



VAD ÄR MINK?

Sedan kör Sparv i bakgrunden och annoterar din text.



NYHETER (stödjer ej utökad kontext)

Sveriges nationella **minoritetsspråk** och svenskt teckenspråk och svenskt **teckenspråk** är ett läromedel för information om Sveriges **språk**, lagar, rättigheter och lagar, rättigheter och **flerspråkighet**. Alla minoriteterna och **minoritetsspråken** samt om svenskt teckenspråk samt om svenskt **teckenspråk**. I vårda de nationella **minoritetsspråken** i Sverige. I språkrådet finns **språkvårdare** i finska, meänkieli, romani och den samiska **språkvården** ansvarar Sametinget. Nya numret av **Klarspråk** hyllar demokratin i nummer av tidningen **Klarspråk** där vi ly msta argumenten för **klarspråk**. kraver i sin artikel hur **klarspråk** kan stå a hur man kan stötta **flerspråkiga** medarbetare. nterar vägledningen ” **Flerspråkighet** i arbetslivet ”.

Syntaktiskt träd: PM → DT → AT → NN → CI → KN → JJ → AT → NN → VB → DT → DT → NN → ET → PP → IP → PA → NN → CI → KN → CI → NN → MAD

TEXTATTRIBUT

Blingbring:

- filantropi
- ansvarspåföljd
- religionsurkund

Visa färre (2)

Svenskt FrasNät:

- Languages
- Publishing
- Conferring_benefit

Visa färre (2)

läsbarhetsindex: 51.54

ordvariationsindex: 62.43

nominalkvot: 1.92

text source: 2021-09-16-nytt-lar om-nationella-minoritetssprak

ORDATTRIBUT

grundform:
teckenspråk

lemgram:
teckenspråk (substantiv)

betydelse:

- teckenspråk

sammansatta ordformer:

- tecken+språk

sammansättningar:
tecken (substantiv) + språk (substantiv)

dependensrelation: Samordnat led

msd: NN.NEU.SIN.IND.NOM ⓘ

ordklass: substantiv

attityd: [tom]

Blingbring:

- beteckning
- betydelse
- meddelande
- språk

Svenskt FrasNät:
Languages

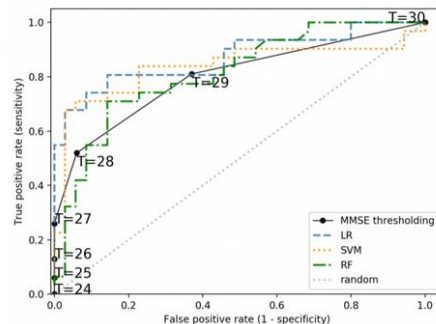
TEKNISK ANVÄNDNING AV PLATTFORMARNA

- De flesta av våra verktyg är öppen källkod.
- Korp och Sparv kan användas genom ett API.
- Korp kan anpassas till andra korpusar och andra språk och används i Italien, Estland, Finland, Island, Danmark och Norge.

EXEMPEL PÅ TILLÄMPNINGAR

Språkteknologi stödjer medicinsk diagnostik:

- Ett vanligt test som används för att diagnostisera lindrig kognitiv störning (MCI) har träffsäkerheten ("AUC score") 68%.
- Om man förstärker testet med automatisk analys av patientens språk förbättras träffsäkerheten till 87%.

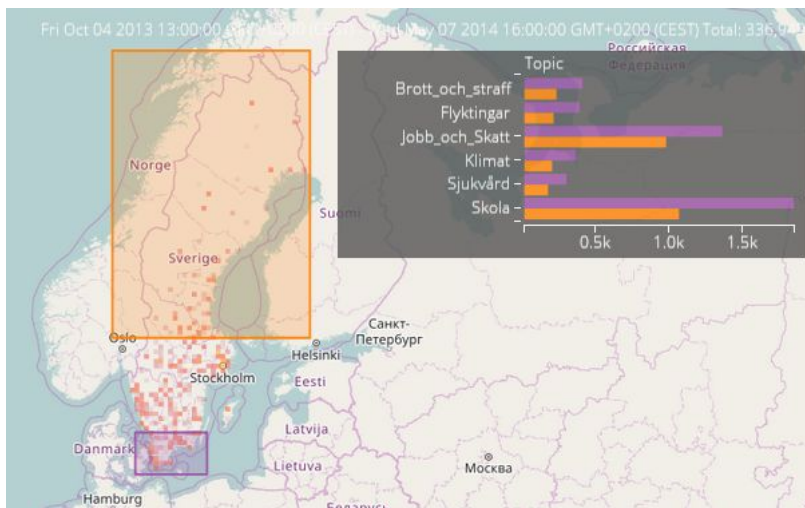


Fraser, K., Lundholm Fors, K., Eckerström, M., Themistocleous, C., & Kokkinakis, D. (2018). Improving the sensitivity and specificity of MCI screening with linguistic information.

EXEMPEL PÅ TILLÄMPNINGAR

Vad diskuterar svenskar på X (Twitter) just nu?

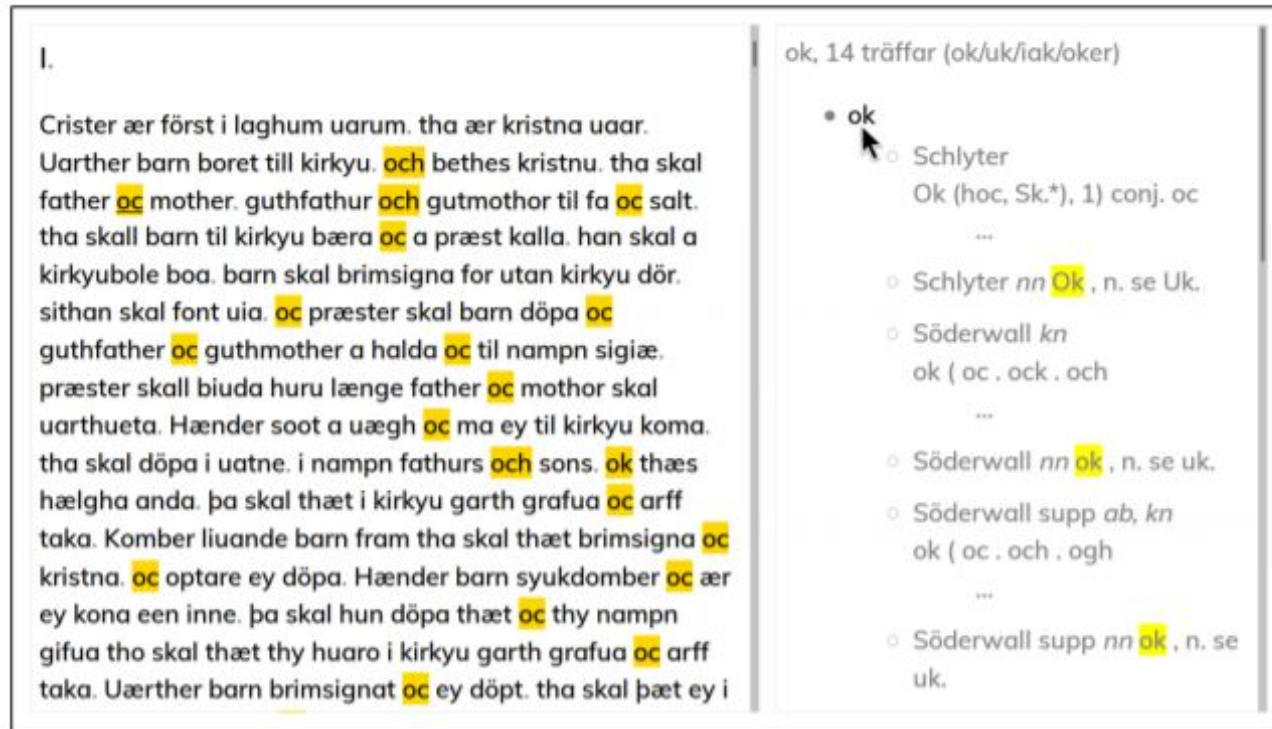
- Orange versus lila: Hur mycket diskuteras temat i norra versus södra Sverige?
- Man ser inga stora skillnader i fördelningen.



Borin, L., & Kosiński, T.
2016. Towards interactive
visualization of public
discourse in time and space

EXEMPEL PÅ TILLÄMPNINGAR

Att förenkla läsning av historiska texter.



I.

Crister ær först i laghum uarum. tha ær kristna uaar.
Uarther barn boret till kirkyu. **och** bethes kristnu. tha skal
father **oc** mother. guthfathur **och** gutmothor til fa **oc** salt.
tha skall barn til kirkyu bæra **oc** a præst kalla. han skal a
kirkyubole boa. barn skal brimsigna for utan kirkyu dör.
sithan skal font uia. **oc** præster skal barn döpa **oc**
guthfater **oc** guthmother a halda **oc** til namn sigiæ.
præster skall biuda huru længe father **oc** mothor skal
uarthueta. Hænder soot a uægh **oc** ma ey til kirkyu koma.
tha skal döpa i uatne. i namn fathurs **och** sons. **ok** thæs
hælgha anda. þa skal thæt i kirkyu garth grafua **oc** arff
taka. Komber liuande barn fram tha skal thæt brimsigna **oc**
kristna. **oc** optare ey döpa. Hænder barn syukdomber **oc** ær
ey kona een inne. þa skal hun döpa thæt **oc** thy namn
gifua tho skal thæt thy huaro i kirkyu garth grafua **oc** arff
taka. Uærther barn brimsignat **oc** ey döpt. tha skal þæt ey i

ok, 14 träffar (ok/uk/iak/oker)

- ok
 - Schlyter
Ok (hoc, Sk.*), 1) conj. oc
...
 - Schlyter nn **ok**, n. se Uk.
 - Söderwall kn
ok (oc . ock . och
...
 - Söderwall nn **ok**, n. se uk.
 - Söderwall supp ab, kn
ok (oc . och . ogh
...
 - Söderwall supp nn **ok**, n. se uk.

Fig. 2. The FSvReader highlights all text words linked to the same lexicon entry, in this case *ok* ('and') in the start passages of the *Younger Västgötalagen*

Adesam, Y., Ahlberg, M., & Bouma, G. 2018. FSvReader – Exploring Old Swedish cultural heritage texts.

EXEMPEL PÅ TILLÄMPNINGAR

Fler exempel på forskning där språkteknologi spelar en roll:

- Hur kan man se konsumtionssamhällets framväxt i äldre svenska romaner (1830–1860)?
- Grammatikböcker innehåller värdefull information om tusentals språk. Kan man extrahera informationen automatiskt?
- Hur kan man använda *big data* och språkteknologi för att analysera språk- och samhällsförändring över tid?

spraakbanken.gu.se/swe/forskning

spraakbanken.gu.se/swe/publikationer

ANNAT VI GÖR

- Undervisar i språkteknologi
- Handleder doktorander
- Informerar om våra resurser, vår forskning och språk över huvud taget
- Svarar på frågor från forskare och allmänhet
- Arrangerar workshoppar
- Håller användardagar
- Bloggar

VAD VI KAN HJÄLPA TILL MED

Andra saker vi kan göra:

- Arrangera en workshop för att hjälpa dig eller dina studenter att komma igång med Språkbanken Text och språkteknologi.
- Hjälpa dig att använda Språkbanken Text för din forskningsfråga.
- Samarbeta för att hitta nya forskningsfrågor.

SPRÅKBANKENTEXT

En forskningsinfrastruktur för språkliga data



GÖTEBORGS
UNIVERSITET



spraakbanken.gu.se