

SPRÅKBANKENTEXT

A research infrastructure for language data
and a language technology research unit



UNIVERSITY OF
GOTHENBURG

SPRÅKBANKEN FOR BEGINNERS

2025

CONTENTS

- What is Språkbanken Text?
- Our data
- Our platforms
- Example applications
- Other things we do
- What we can help you with

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 information 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 tidningskorpusar

BLOGGMIX 2010 (stödjer ej utökad kontext)

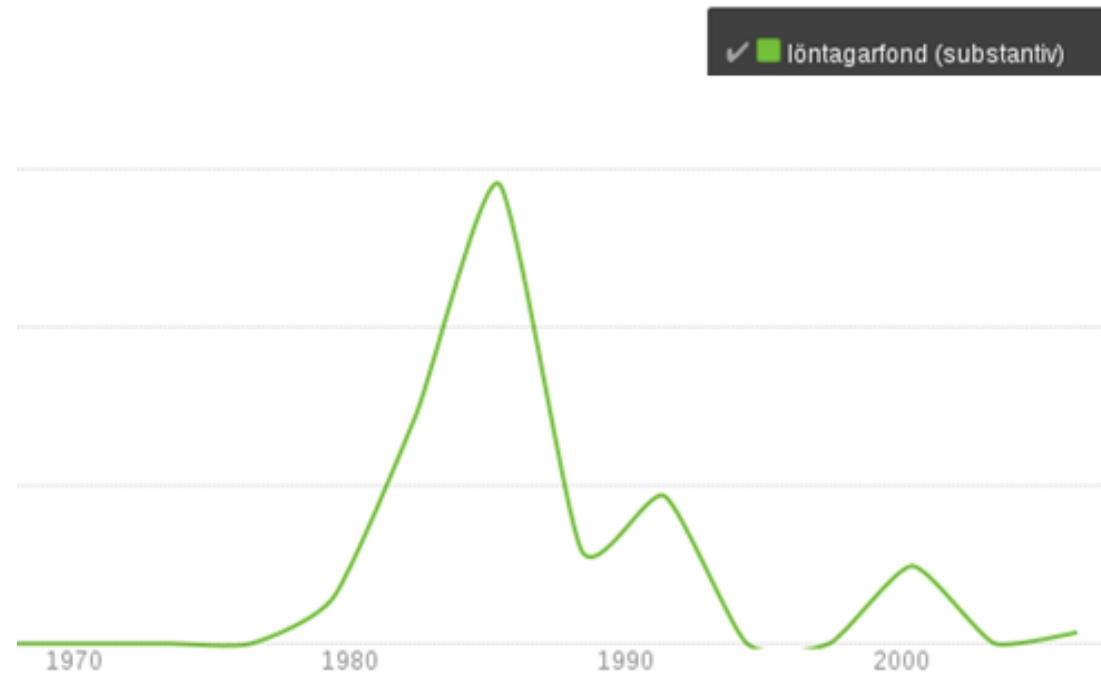
De röda symbolerna är [Språkbanken](#) , de blå träffar i Dagens Nyheter på Mediearkivet. korpusen Press65, som finns på [Språkbanken](#) , så finns där 46 förekomster av äta men ingen alls av käka, och talar är mer än tatar och talar i Mediearkivet och [Språkbanken](#) , resultaten syns i diagrammet nedan.

BLOGGMIX 2011 (stödjer ej utökad kontext)

har därför skapat en gemensam [språkbank](#) med frivilliga medlemmar och förtroendevalda.

WHAT IS SPRÅKBANKEN TEXT?

- A **bank** creates profit
- Språk**banken** creates knowledge



WHO NEEDS SPRÅKBANKEN TEXT?

Any researcher in the field of human communication:

- computer scientists
- philologists
- gender scientists
- historians
- health researchers
- cultural scientists
- linguists
- literary scholars
- media scientists
- education researchers
- psychologists
- social anthropologists
- language technologists
- political scientists
- and more!

THERE ARE MORE THAN ONE SPRÅKBANK

- **Språkbanken** is a research infrastructure financed jointly by the Swedish Research Council and 10 universities and government organisations.
- It consists of three divisions: **Språkbanken Sam**, **Språkbanken Tal** and **Språkbanken Text**.
- This presentation is about Språkbanken Text, which works with written language data.



OUR DATA

- More than 1000 corpora
- More than 50 lexicons

You can search for and download them on our webpage:
spraakbanken.gu.se/en/resources

Svenska

 UNIVERSITY OF GOTHENBURG

SPRÅKBANKENTEXT

A research infrastructure for language data and a language technology research unit

Språkbanken Text is a part of Språkbanken.

News and events Research Data Analyses Platforms FAQ About us Contact us

[Home](#) / Language resources

Language resources

On this page you can browse and search our datasets. Click on a row name to see what files are available for download. You can go directly to the search interface by clicking on the tool logo.

All (1330) Collections (31) Corpora (1201) Lexicons (66) Training and evaluation data (15) Models (48)

| Name or description | Language | | | |
|---|----------|---------|--|---|
| | - Any - | | | |
| 8 Sidor News articles from 8 SIDOR. | Corpus | Swedish | Dataset: attasidor.xml.bz2 2024-03-07 – 150.01 MB – CC BY 4.0 | Explore in:  |
| Academic texts: Humanities A corpus with academic texts | Corpus | Swedish | Dataset: sweachum.xml.bz2 2017-05-19 – 208.67 MB – CC BY 4.0 | Explore in:  |

WHAT IS A CORPUS?

- A large searchable collection of texts.
- To improve searchability, corpora are often annotated, that is, linguistic information has been added.
- The annotation is often done by a computer, and may contain errors.

Annotation gives information:

- **Morphological**: *Boken* is a noun, singular definite form, the same word as *bok*, *böcker* and *böckerna*
- **Syntactic**: *Jag läser en bok*: *bok* är the object
- **Semantic**: Vilken sense of *bok* – "book" or "beech"?
- **Metadata**: In what context is the word used? Who wrote the text? When?

SOME OF SPRÅKBANKEN'S CORPORA

- fiction
- newspapers
- magazines
- social media
- blogs
- webforums
- Twitter
- government documents
- Wikipedia
- Finland Swedish
- historical texts
- Old Swedish
- modern Swedish
- parallel texts
- the same texts in 25+ languages
- the Gigaword corpus
- a large, balanced corpus of different genres of modern Swedish from 1950 forward

SOME OF SPRÅKBANKEN'S LEXICONS

- SALDO: a semantisc and morphological lexicon
- Hellquist's Swedish etymological dictionary
- Bliss symbol lexicon
- Schlyter's and Söderwall's dictionaries of medieval Swedish
- attitude lexicon
- lexicon of borrowed words
- Swesaurus (a Swedish word net, a Swedish conceptual lexicon)

Our lexicons are made for automatic text analysis.

SPRÅKBANKEN TEXT

SPRÅKBANKEN'S MOST IMPORTANT PLATFORMS



- **Korp:** a search engine which gives you access to 30 000 million words in Språkbanken's corpora.
- **Karp:** a platform for working with our lexicons.
- **Sparv:** an annotation platform for adding morphological, syntactic and semantic information to text.
- **Strix:** a search engine which is document centered and can focus more on the semantic content of the document.
- **Mink:** a platform where you can apply our language technology methods to your own texts.

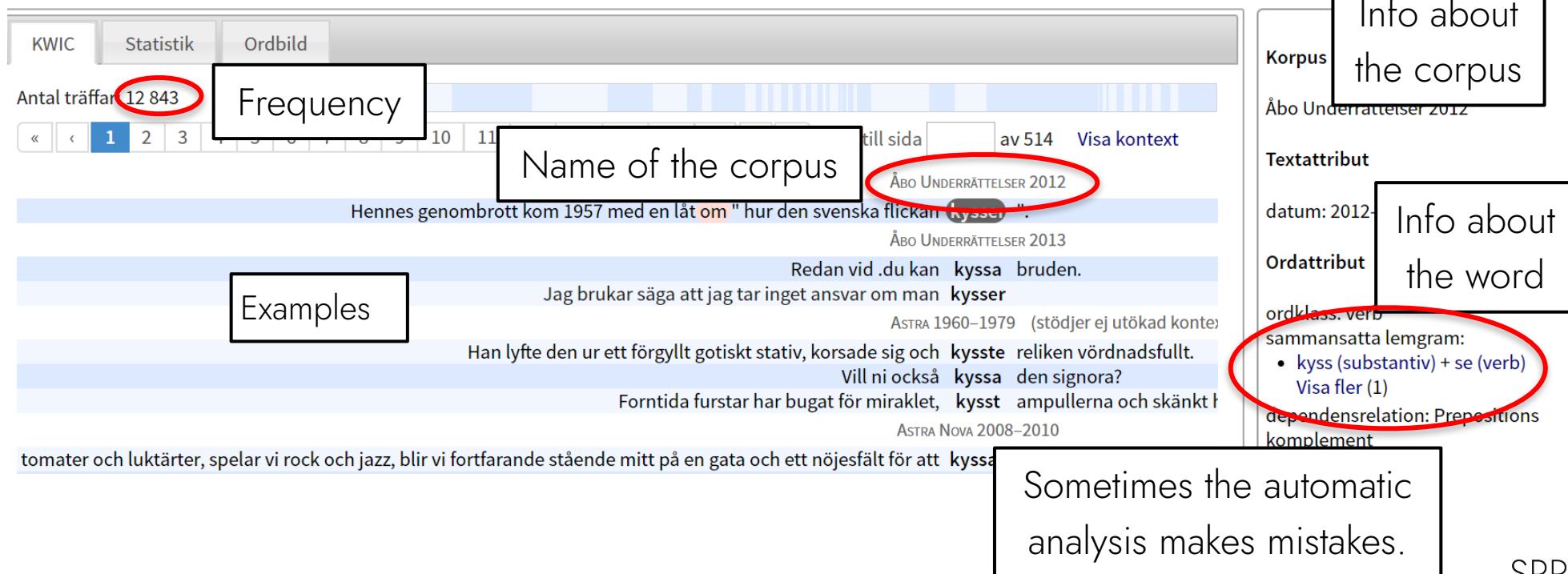
WHAT IS KORP?



- a search engine
- made for investigating linguistic phenomena

WORD SEARCH

- Gives information about the frequency of a word and shows all its occurrences in context. The word *kyssa* "kiss":

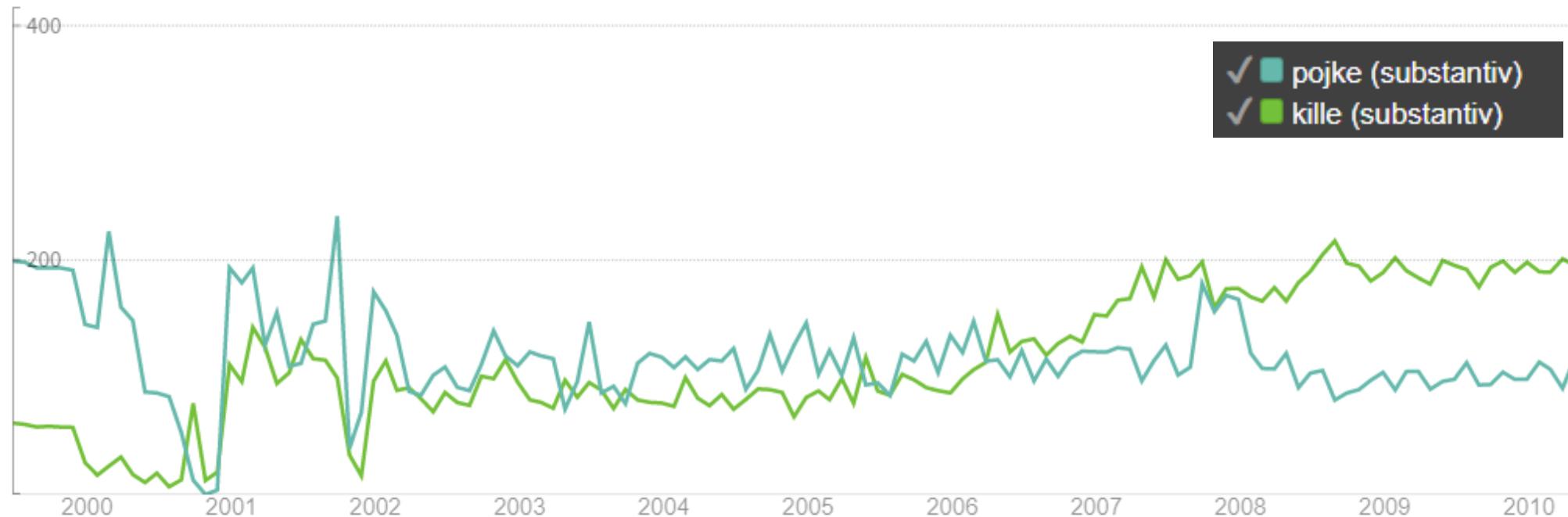


The screenshot shows the Korp search interface with the following highlights:

- Frequency:** The total number of hits (12 843) is circled in red.
- Name of the corpus:** The corpus name (ÅBO UNDERRÄTTELSE 2012) is circled in red.
- Examples:** Several examples of the word in context are shown.
- Info about the word:** A box on the right contains the following information:
 - Korpus: Åbo Underrättelser 2012
 - Textattribut: datum: 2012-
 - Ordattribut: ordklass. verb
sammansatta lemmgram:
 - kyss (substantiv) + se (verb)
 - Dependensrelation: Prepositions komplement
- Sometimes the automatic analysis makes mistakes.**: A note in a box pointing to the lemma analysis.

TREND DIAGRAMS

- Language change:



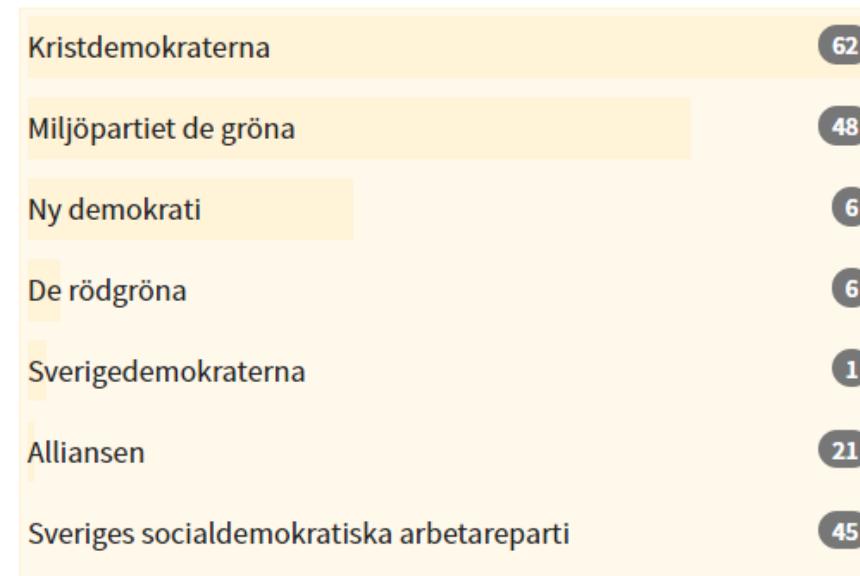
COMPARISONS BASED ON PROBABILITY

- Which political parties focus more on education, and which focus more on healthcare?

Utmärkande för *utbildning*



Utmärkande för *sjukvård*



WORD PICTURES

■ Who kisses whom/what, and in what way?

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

WHAT IS KARP?



- A tool for working with Språkbanken's lexicons.

filtrera

| | | | | | | |
|--|--|--|-----------------------------------|---|--|---|
| ACADEMIC WORDLIST Vanligt förekommande ord i en korpus på drygt 25 miljoner ord av avhandlingar och publicerade artiklar | BLINGBRING En SALDO-länkad och moderniserad version av Brings "Svenskt ordföråd ordnat i begreppsklasser" (1930) | DALIN Ordbok över 1800-talsspråket, utifrån den tryckta ordboken: Dalin, Anders Fredrik Ordbok Öfver svenska språket. Vol I-II. (Stockholm 1850–1853). | FORSVENSK MORFOLOGI TBA | FULA ORDBOKEN Ordbok med fokus på ämnen sex, snusk, kön och svordomar | HELLQUISTS SVENSK ETYMOLOGISK ORDBOK Svensk etymologisk ordbok av Elof Hellquist, digitalisering från första utgåvan från 1922 | KELLY Keywords for Language Learning for Young and adults alike |
|--|--|--|-----------------------------------|---|--|---|

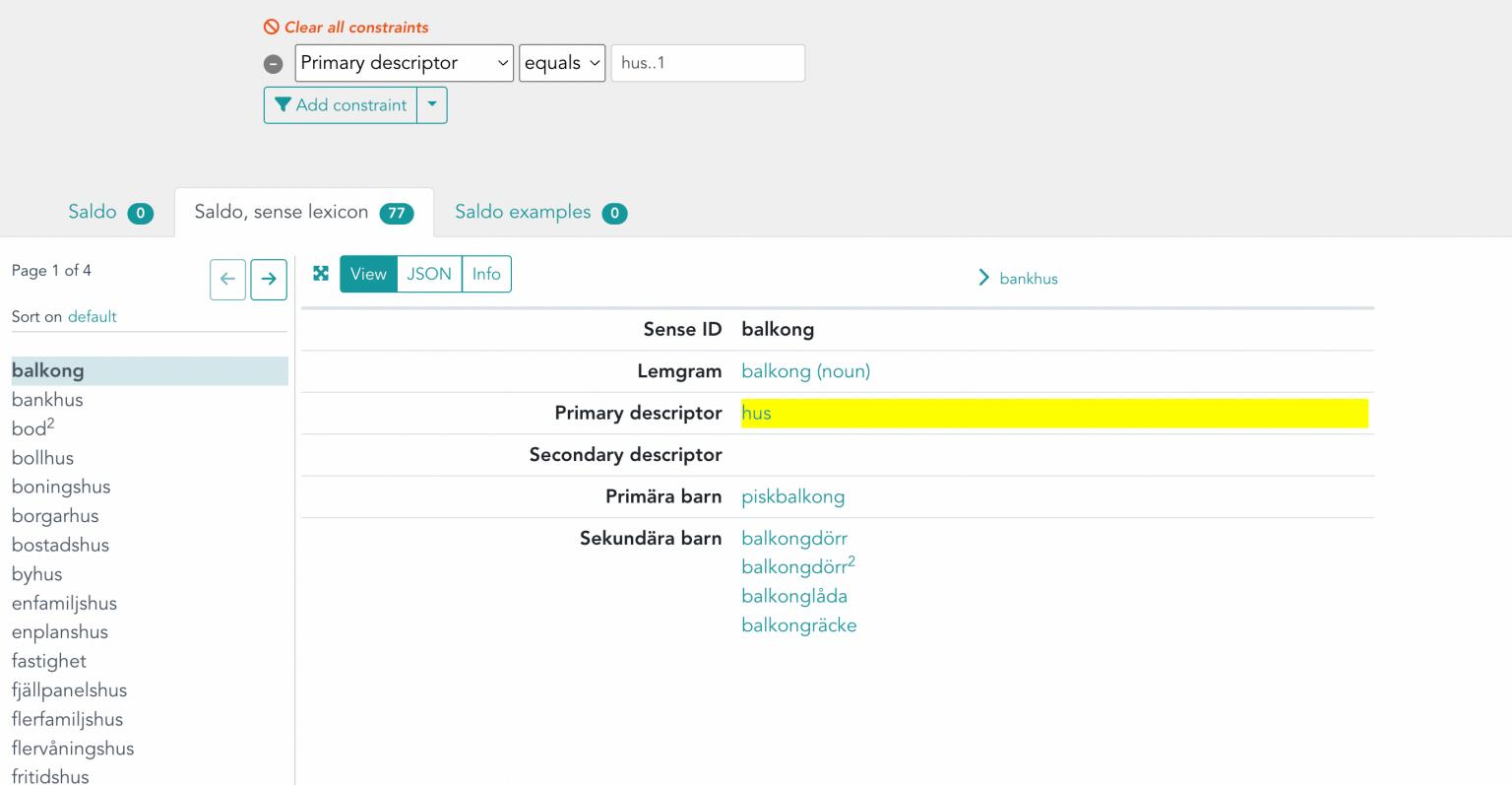
Språkbanken Text is a research infrastructure for language data. We create and maintain language data analysed with state-of-the-art language technology methods for researchers everywhere, and take pride in providing free and open access where possible.

Karp is our lexical infrastructure for searching and editing lexicons.

Choose a **LEXICON** above to get started

SPRÅKBANKENTEXT

■ What is the semantic neighbourhood for *hus* "house"?



The screenshot shows the SALDO interface for exploring the semantic neighbourhood of the word 'hus' (house). At the top, there is a search bar with the query 'Primary descriptor equals hus..1'. Below the search bar, there are tabs for 'Saldo' (0), 'Saldo, sense lexicon' (77), and 'Saldo examples' (0). The main content area displays a list of words related to 'hus', with 'balkong' selected. The right side shows detailed information for the selected sense:

| Sense ID | balkong |
|----------------------|--|
| Lemgram | balkong (noun) |
| Primary descriptor | hus |
| Secondary descriptor | Primära barn: piskbalkong |
| Sekundära barn | balkongdörr balkongdörr ² balkongläda balkongräcke |

On the left, a sidebar lists other words: bankhus, bod², bollhus, boningshus, borgarhus, bostadshus, byhus, enfamiljshus, enplanshus, fastighet, fjällpanelshus, flerfamiljshus, flerväningshus, fritidshus.

ETYMOLOGICAL DICTIONARY



- From which languages does Swedish have the most loan words?

The query: "find all entries where a source language is given, and that language is not Old Swedish"

Hellquists Svensk etymologisk ordbok 483

Sida 1 av 20 ← → Visning JSON Info

Sortera på förvalt fält

ackompanjera

adagio
ass
baisse
bajadär
bajonett
bakterie
balett
balja
balkong
ballad
ballong

Grundform ackompanjera
Ordklass vb
Definition ackompanjera, om musik i sv. från senare hälften av 1700-t., från fra.
accompagner, beledsaga; se f. ö. kompani.
Etyma Språk fra
Form accompagner
Beskrivning beledsaga
Faksimil

adagio

SPRÅKBANKENTEXT

MORE ABOUT KORP

- Let's search for the word *grym* "cruel, wicked":

| Different senses! | |
|--|---|
| GP 2013 | grym |
| Berättelserna om tortyr, övergrepp, skräck blir just så obegripligt. | grymma – som i verkligheten. |
| Vi tappade bollen – och de hade sådan | grym kvalitet i det som de gjorde. |
| Mika har gjort | grymt mycket poäng och Kari är en av allsvenskanas bästa vänsterbackar, säger han. |
| n, Vägen mot Bälberget, 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. |

AUTOMATIC TEXT ANALYS



■ Senses in Korp:

| 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 sådan | grym kvalitet i det som de gjorde. |
| Mika har gjort | grymt mycket poäng och Kari är en av allsvenskans bästa vänsterbackar, säger han. |
| an, Vägen mot Bälberget, ger sig Therése Söderlind i kast med detta | grymma och svårbegripligt kapitel i Sveriges historia. |
| Häcken är en | grym klubb, en skön klubb. |
| gen varje månad ser arrangörerna till att ge besökarna tung bas och | grym electro. |
| indratals 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 mäste 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. |
| indratals 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 längt kvar tills ridän 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. |

Gå till sida av 19

Korpus

GP 2013

Textattribut

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

Ordattribut

ordklass: adjektiv
sammansatta lemgam: *[tom]*
dependensrelation: Subjektspredikativ
(subjektiv predikatsfyllnad)

förled: *[tom]*

etterled: *[tom]*

betydelse:

- **grym**²
[Visa fler \(1\)](#)

msd: JJ.POS.UTR.SIN.IND.NOM [?](#)

sammansatta ordformer: *[tom]*

grundform:

grym

lemgam:

grym (adjektiv)

[Visa dependensträd](#)

AUTOMATIC TEXT ANALYS



- The senses in Korp come from the SALDO lexicon:

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 sådan 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)

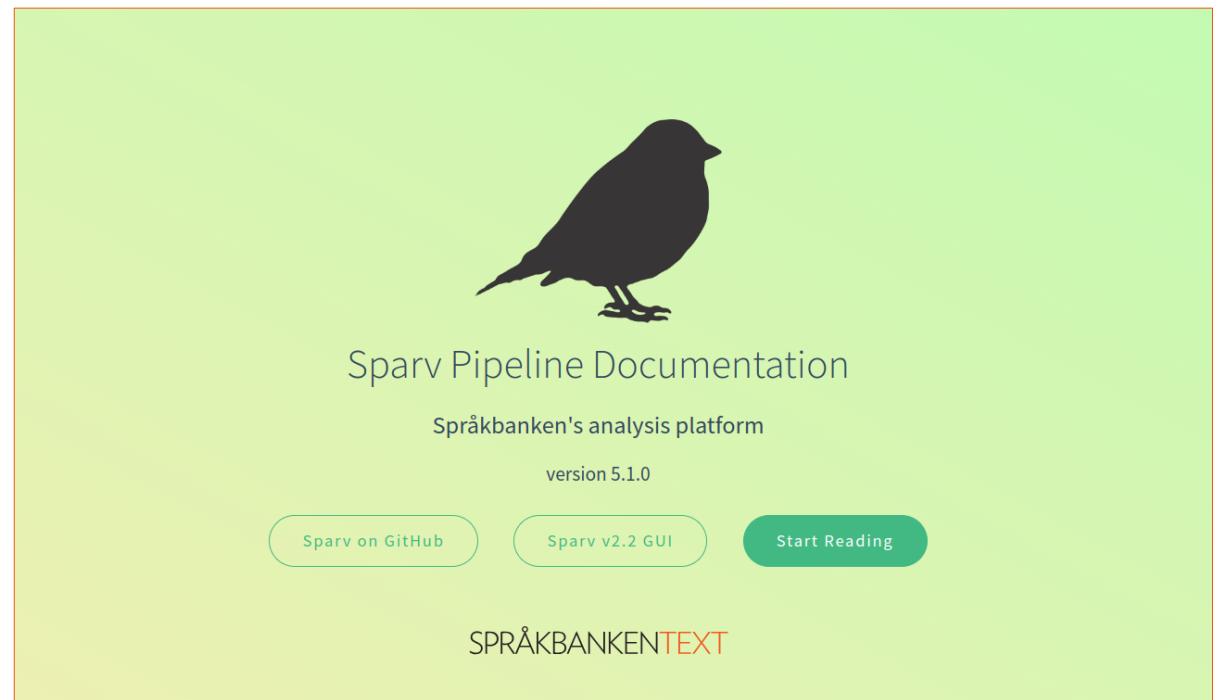
| BETYDELSE | LEMGÅRM | ORDKLASS | PRIMÄR | SEKUNDÄR | BARN (PRIMÄRA) | BARN (SEKUNDÄRA) |
|-------------------|----------------------------|----------|--------|----------|---|------------------|
| grym | grym (adjektiv) 1087340 | adjektiv | hård | | bestialisk grymhet kärlekslös | |
| grym ² | grym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

WHAT IS SPARV?



Sparv can annotate your texts, and add:

- lexical analysis
- syntactic analysis
- semantic analysis
- sentiment analysis
- readability analysis
- and more!

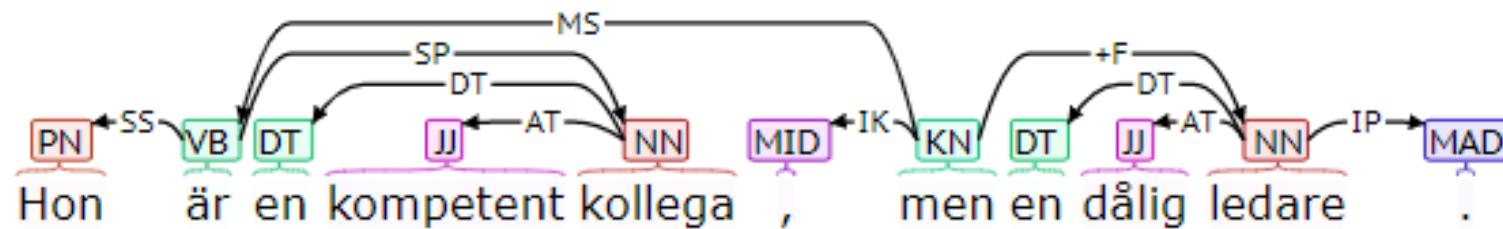


SPRÅKBANKEN TEXT

SYNTAKTIC ANALYSIS

Hon är en kompetent kollega, men en dålig ledare.

"She is a competent colleague, but a poor leader."



SEMANTISK ANALYS



Hon är en kompetent kollega, men en dålig ledare.

The word *ledare* has at least three senses:

| Sense | Probability in this sentence, according to Sparv |
|--------------------|---|
| leader | 0,65 |
| electric conductor | 0,21 |
| editorial | 0,14 |

SENTIMENT ANALYSIS



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 |

WHAT IS STRIX?



A search engine which:

- is focused on entire documents (unlike Korp, which focuses on words)
- analyses the semantic contents of the document

WHAT CAN STRIX DO?



- text analysis
- find keywords
- find names
- sentiment analysis
- topic classification
- show entire annotated documents
- work with collections of documents
- show metadata

The screenshot shows a document titled "Sveriges liberala partis valprogram 1924" (Swedish Liberal Party election manifesto 1924) from the year 1924. The document has 1599 tokens. The most common nouns include parti (12), land (11), fråga (6), folk (6), fred (6), riksdag (5), frihet (5), hänsyn (4), läge (4), utveckling (4), Sveriges (9), Nationernas förbund (4), riksdagen (1), Gent (1), and Sunda (1). The document view shows the text of the manifesto, which discusses issues like the economy, social welfare, and peace. The sidebar on the right provides various annotations for the word 'liberal': Text attributes, Structural attributes, Word attributes, LEMORAM (with entries for liberal¹ (adjective) and liberal² (adjective)), SENSE (with entries for liberal¹ and liberal²), COMPOUND WORD FORMS, COMPOUND LEMRAMS, DEPENDENCY RELATION (with AT and DEPHEAD), and DEPHEAD.

Title: Sveriges liberala partis valprogram 1924
Corpus name ID: SWEDISH PARTY PROGRAMS AND ELECTION MANIFESTOS
Document size: 1599 tokens
Year: 1924

Most common nouns: parti (12), land (11), fråga (6), folk (6), fred (6), riksdag (5), frihet (5), hänsyn (4), läge (4), utveckling (4)

Most common names: Sveriges (9), Nationernas förbund (4), riksdagen (1), Gent (1), Sunda (1)

Document view Statistics view

Explore Name tag Type Välj från lista (4 alternativ)

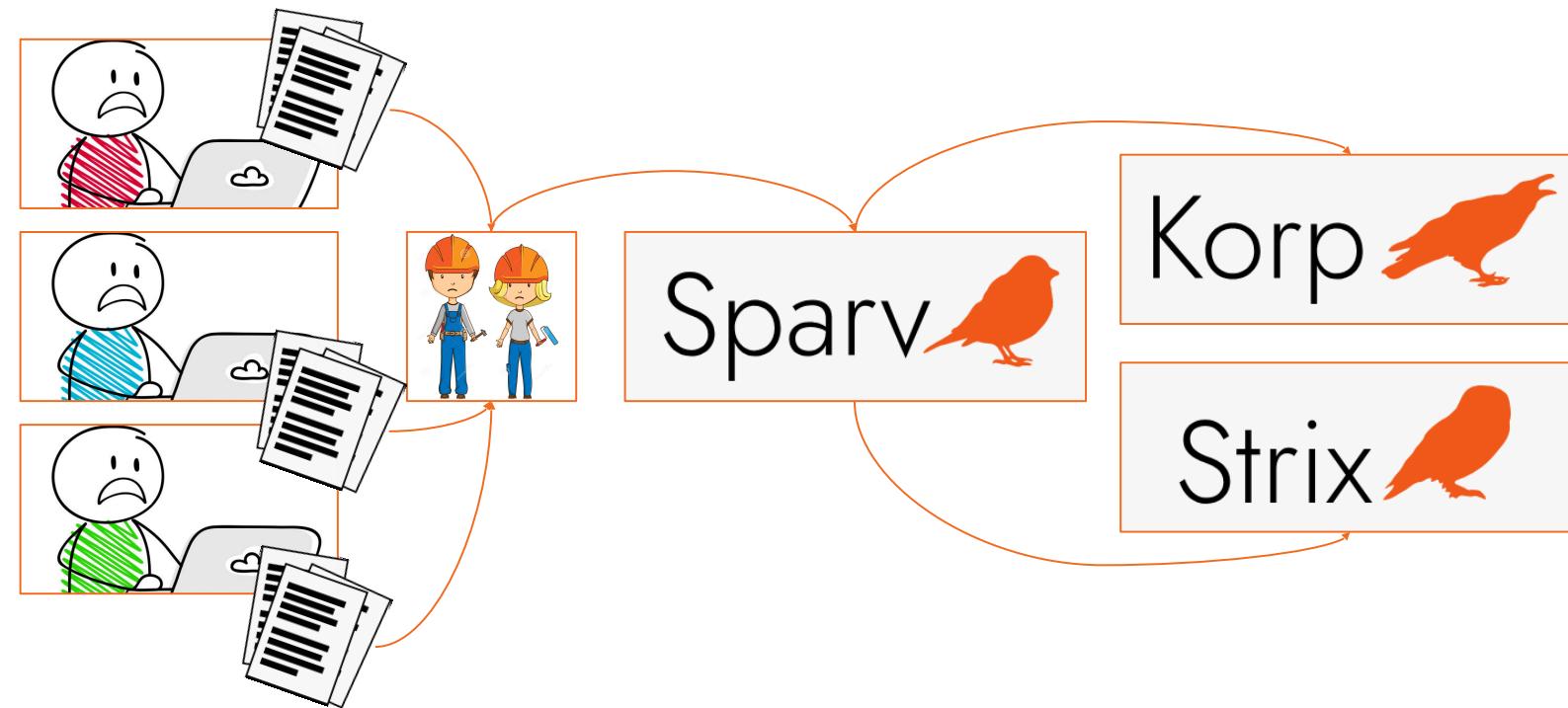
Search the current document liberal

1 Sveriges liberala partis valprogram.
2
3 De val till andra kammar, som nu förestå, är mer än vanligt betydelsefulla. De gälla främst försvarsfrågan som vid årets riksdag lämnats olöst. Men de gälla också andra frågor av utomordentlig vikt. Sveriges folk skall i dessa val för hela fyra år framåt sätta sitt bestämmande ord om de stora stridsfrågorna i vårt ekonomiska, sociala och andliga liv.
4
5 Splittringen mellan parterna har under den gångna riksdagen ökats och motsatserna skärpts. Alla bekänna sig till samma mål: att avhjälpa ytterligare otryghet och ekonomiskt betryck. Men vägarna till målet går ohjälpligt isär. Ytterlighetsritningar åt båda sidor rusta sig till en oförsiktig strid.
6
7 Sveriges liberala parti vägrar att gå med i denna kamppolitik. Den kan icke gagna land och folk. Fredens sak främjas icke därav, att vårt land brådstörtat slår in på avväpningens väg trots oron och osäkerheten i världen. Sveriges fred kan ej heller tryggas allenaest genom militära rustningar utan tillräcklig hänsyn till det ekonomiska läget och folkets vilja. Landets välfärd skulle icke befördas utan undergrävas genom socialisering av enskilda företag och genom kroppsarbartarnas herravälde över produktionen. Det har lika litet något att vinna genom höga tollmurar eller sociala tvångsåtgärder i arbetsfredens namn.
8
9 Sveriges liberala parti är förvissat däröm, att fred och välfärd måste sökas på andra vägar. Dess politik bygger på personlighetens värde och rätt. Den hävdar alias frihet, samhörighet och lika rätt i samhället. Den kräver ett aktivt fredsarbete, men anser att landet alltidjämt behöver ett ändamålsenligt ordnat försvarsväsen. Den vill utjämna de skärpta motsatserna mellan olika meningsskriftningar, och utan snäva partifabian framhäva vidgad samverkan för allmänt väl.
10
11 Med hävning till sitt den 7 oktober 1923 antagna allmänna program vill Sveriges liberala parti här angiva sin ställning till vissa särskilt aktuella frågor.
12
13 AKTIV FREDSPOLITIK. Det liberala partiet vill medverka i det mellanfolkliga samarbetet för en varaktig fred. Målet för liberalismens strävan den i alla länder är att låta rätten träda i väldets ställe och att grunda förhållandet mellan staterna på frihet och likställdhet.
14

WHAT IS MINK?



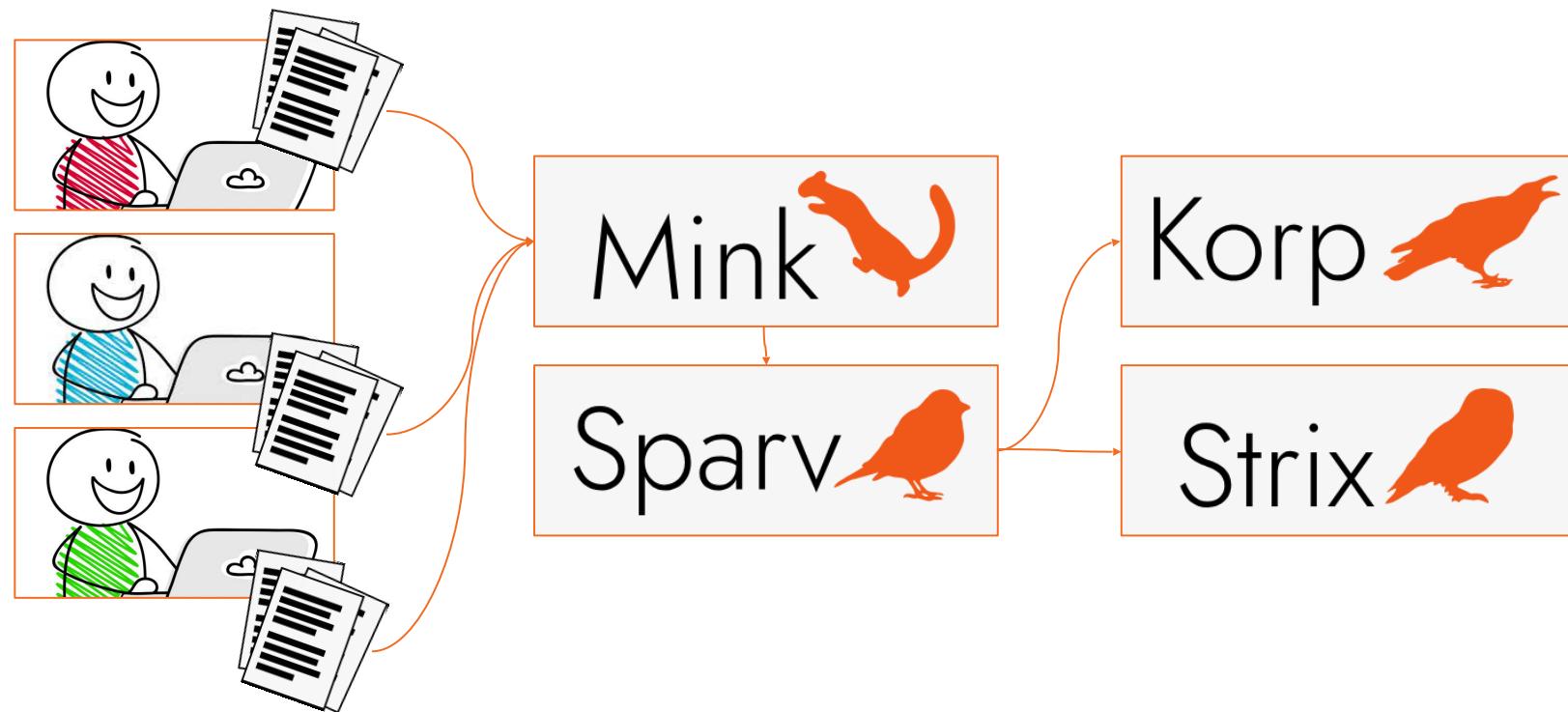
Adding new corpora involves Språkbanken Text.



WHAT IS MINK?



With Mink, you can add and analyse your own texts!



WHAT IS MINK?



Then Spary runs in the background and annotates your text.

ORDATTRIBU

grundform:

- Blingbring:

 - filantropi
 - ansvarspåföld
 - religionsurkund

Visa färre (2)

- Svenskt FrasNat:
 - Languages
 - Publishing

• Content

Läsbartextindow E1-E4

ordvariationsindex: 62

nominalkvot: 1.92

text source: 2021-09-16-nytt-l
om-nationella-minoritetssprå

teckens

| lemgram:

- teckenspråk

betydelse:

- teckenspråk
sammansatta ord

- lecker+sprak

Sammansättningar.
tecken (substantiv) + språk
(substantiv)

dependensrelation: Samordnat led

msd: NN.NEU.SIN.IND.NOM

ordklass: substantiv

attltyd: [tə]

BlingBling:

betydelse

meddelande

Speaking

SPRÅKBANKEN TEXT

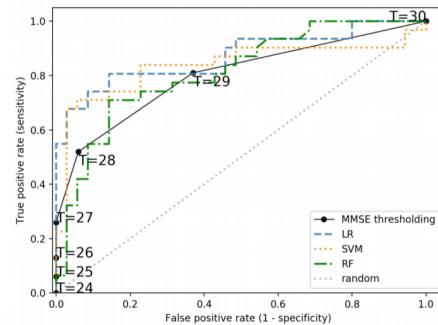
TECHNICAL USE OF OUR PLATFORMS

- Most of our tools are open source.
- Korp and Spary can be used through an API.
- Korp can be adapted to other corpora and other languages, and is used in Italy, Estonia, Finland, Iceland, Denmark och Norway.

EXAMPLES OF APPLICATIONS

Language technology supports medical diagnostics:

- A common test used to diagnose mild cognitive impairment (MCI) has an accuracy ("AUC score") of 68%.
- If we supplement the test with automatic analysis of the patient's language, the accuracy is improved to 87%.

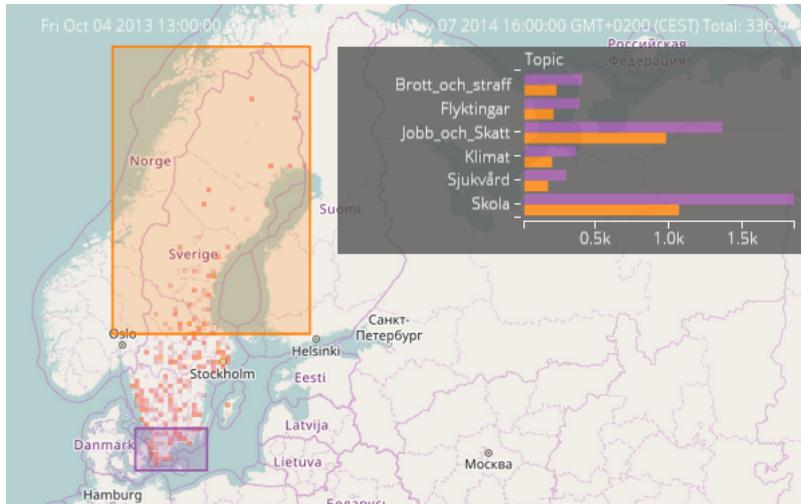


Fraser, K., Lundholm Fors, K., Eckerström, M., Themistocleous, C., & Kokkinakis, D. (2018). Improving the sensitivity and specificity of MCI screening with linguistic information.

EXAMPLES OF APPLICATIONS

What are Swedes discussing on X (Twitter) right now?

- Orange versus purple: How much is the topic discussed in northern versus southern Sverige?
- We see no big differences in the distribution.



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

EXAMPLES OF APPLICATIONS

Simplifying the reading of historic texts.

The screenshot shows a comparison between a raw text passage and a lexical entry. On the left, a block of Old Swedish text from the 'Younger Västgötalagen' is displayed. Several words are highlighted in yellow, including 'och' (and), 'ok' (and), 'kn' (conjunction), and 'oc' (and). On the right, a detailed lexical entry for the word 'ok' is shown. The entry title is 'ok, 14 träffar (ok/uk/iak/oker)'. Below the title, a bullet point lists 'ok' with a cursor icon, followed by a list of 14 entries from various sources: Schlyter, Ok (hoc, Sk.*), Söderwall, and others. Ellipses indicate more entries.

I.

Crister ær först i laghūm 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**
guthfather **oc** guthmother a halda **oc** til nampn sigiae.
præster skall biuda huru længe father **oc** mothor skal
uarhueta. Hænder soot a vægh **oc** ma ey til kirkyu koma.
tha skal döpa i uatne. i nampn 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 nampn
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.

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

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*

EXAMPLES OF APPLICATIONS

More examples of research where language technology plays a part:

- How can we see the growth of consumerism in older Swedish novels (1830–1860)?
- Grammar books contain valuable information about thousands of languages. Can we extract that information automatically?
- How can we use "big data" and language technology to analyse changes in language and society over time?

spraakbanken.gu.se/en/research

SPRÅKBANKEN TEXT

OTHER THINGS WE DO

- Teach language technology
- Supervise doctoral students
- Inform about our resources, our research, and language in general
- Answer questions from researchers and the public
- Arrange workshops
- Arrange user days
- Blog

WHAT CAN WE HELP YOU WITH

Other things we can do:

- Arrange a workshop to help you or your students get started with Språkbanken Text and language technology.
- Help you use Språkbanken Text for your research question.
- Cooperate to find new research questions.

SPRÅKBANKENTEXT

A research infrastructure for language data
and a language technology research unit



GÖTEBORGS
UNIVERSITET



spraakbanken.gu.se