

Språkbanken

Dana Dannélls
Språkbanken Text

dana.dannells@svenska.gu.se

SND nätverksträff i Göteborg, 9 april 2025

Språkbanken

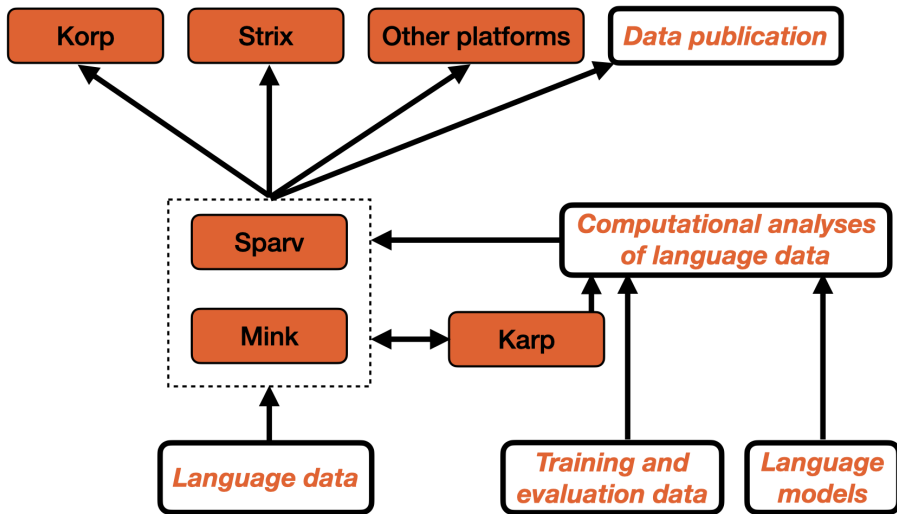
- Språkbanken är en nationell forskningsinfrastruktur baserad på språkteknologi för text och tal (och nu även teckenspråk)
- Språkbanken samlar in, bearbetar och tillhandahåller och gör därmed språkliga data mer forskningsbara på många olika sätt för många olika discipliner

Språkbanken består av fyra avdelningar

- Språkbanken Text: textbaserad språkteknologi (Göteborg)
- Språkbanken Tal: talteknologi (KTH, Stockholm)
- Språkbanken Sam: kulturarv och språkpolitik (Isof, Uppsala)
- Språkbanken CLARIN: koordinerar de svenska CLARIN-aktiviteterna (Uppsala)

<https://sprakbanken.se/>

Språkbanken Texts forskningsinfrastruktur



Var kommer datan ifrån

- Minnesinstitutioner, arkiv
- Forskare runt om i Sverige
- Forsknings(infrastruktur)projekt
- Samarbeten med andra forskningsinfrastrukturer
- ...

Datamängder och analyser

I nuläget tillgängliggör vi 1330 datamängder

- Korpusar: Stora textmängder
 - ▶ nyhetstexter, tidskrifter, myndighetstexter, skönlitteratur, sociala medier, ...
- Språkteknologiska lexikaliska resurser och annan strukturerad data
 - ▶ associationslexikon, ordnät, frasnät, attitydlexikon, ordlistor ...
- Tränings- och utvärderingsdata
 - ▶ språkrättning, OCR-transkribering, namnigenkänning, ...
- Analyser
 - ▶ attitydanalys, ordförutsägelse, OCR-korrigerig, ordklasstagging, ordvektorer (både moderna och historiska), betydelsedisambiguerig, ...

Det mesta är fritt tillgängligt (CC BY 4.0)

Datacitering

Kubhist 2: Nytt och Gammalt 1790-talet

Datacitering

Språkbanken Text (2021). *Kubhist 2: Nytt och Gammalt 1790-talet* (uppdaterad: 2021-11-24). [Data set]. Språkbanken Text. <https://doi.org/10.23695/sww8-rp15>

 Ytterligare sätt att citera datamängden.

Del av samlingen Kubhist 2

Tillgänglig via

Åtkomst	Plattform	Licens
https://spraakbanken.gu.se/korp/?mode=kubhist#?corpus=kubhist2-nyttochgammalt-1790	 Språkbankens ordforskningsplattform	CC BY 4.0

Ladda ned

Fil		Storlek	Modifierad	Licens
kubhist2-nyttochgammalt-1790.xml.bz2 corpus  (XML)		35.54 MB	2021-11-24	CC BY 4.0
stats_kubhist2-nyttochgammalt-1790.csv Ordstatistik:  (CSV)		9.4 MB	2021-11-25	CC BY 4.0

Del av samling

[Kubhist 2](#)

Typ

Korpus

Språk

svenska

Storlek

Meningar: 82 347

Token: 1 327 695

Uppdaterad

2021-11-24

Kontakt

sb-info@svenska.gu.se

DOI

<https://doi.org/10.23695/sww8-rp15>

Analyscitering

sbx-swe-namedentity-swener

Standardreferens ⓘ

Dimitrios Kokkinakis, Jyrki Niemi, Sam Hardwick, Krister Lindén, and Lars Borin. 2014. HFST-SweNER — A New NER Resource for Swedish. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 2537-2543. Reykjavik, Iceland. European Language Resources Association (ELRA).

Analyscitering ⓘ

Språkbanken Text (2020). *sbx-swe-namedentity-swener* (uppdaterad: 2020-05-13). [Analysis]. Språkbanken Text.



Namnigenkänning känner igen och förser namn och namnliknande uttryck (s.k. entiteter) i löpande text med fördefinierade etiketter, som organisation, person eller plats.

Namnigenkänning är en språkteknologisk teknik som automatiskt känner igen och förser namn och namnliknande uttryck (s.k. entiteter) i löpande text med fördefinierade etiketter, som t. ex. person eller organisationer, men, beroende på tillämpningsområdet, även numeriska uttryck och tidsuttryck. HFST-SweNER bygger på konvertering, modellering och anpassning av en tidigare svenskt NER-system till Helsinki Finite-State Transducer Technology (HFST)-plattformen. HFST-SweNER är en fullfjädrad implementering med öppen källkod som stöder en mängd olika generiska namngivna entitetstyper och består av flera lexikala resurslager såsom olika n-gram-baserade namngivna namnlistor (s.k. gazetteers).

Exempel

This analysis is used with Sparv. Check out [Sparv's quick start guide](#) to get started!

To use this analysis, add the following lines under `export . annotations` in the Sparv [corpus configuration file](#):

Del av samling

[sbx-swe-mink_analyses](#)

[sbx-swe-standard_analyses](#)

Typ

Analys

Uppgift

namnigenkänning

Beroenden

Externa verktyg

[HFST-SweNER](#)

CC-BY

Taggmängd

[Named entity tags from hfst-SweNER](#)

Skapad

2014-07-04

Uppdaterad

2020-05-13

Kontakt

sb-info@svenska.gu.se

Data plattformar

- Korp: Språkbankens ordforskningsplattform
- Strix: Språkbankens textforskningsplattform
- Karp: Språkbankens dataredigeringsplattform
- Sparv: Språkbankens analysplattform
- Mink: Språkbankens dataplattform

Korp korpuser

Moderna | Parallella | Fornsvenska | Litteraturbanken | Kubhist | Kubord | Fler ▾

Lo

Korp Språkbankens ordforskningsplattform

Enkel

Utökad

Avancerad

Jämförelse

1

Sök

☐ förlädd 

☐ mellanled 

☐ efterled 

Korp är Språkbankens

Här kan du söka i stora

dagstidningar, skönlitte

Textsamlingarna, korpu

språkliga och andra attr

[Användarhandledning](#)

Om du använder Korp i

oss. För frågor och kom

23 av 288 korpuser valda — 256,30M av 16,40G token



✓ Markera alla

✗ Avmarkera

- ☐ Akademiska texter (2)
- ☐ August Strindberg (2)
- ☐ Finlandssvenska texter (56)
- ☐ L2-Korp (8)
- ☐ Medicinska texter (13)
- ☐ Myndighetstexter (35)
- ☒ Nyhetstexter (59)
- ☒ Skyddade korpuser (18)
- ☐ Skönlitteratur (7)
- ☐ Sociala medier (68)
- ☐ Tidskrifter (3)
- ☐ Dramawebben (demo)
- ☐ Etnologiska frågelistor



























KWIC
Statistik
OrdBild

☐ Visa kontext
Träffar per sida: 25
Sortera varje korpus på: förekomst

Antal träffar: 100 696

1 2 3 4 5 6 7 8 9 10
Gå till sida: av 4 028

8 SIDOR (stödjer ej utökad kontext)

Daghem i Frankrike ska inte längre få ha trådlösa **nätverk** till sina datorer.
Nätverken i skolor för barn upp till sjätte klass ska stängas av när de inte används.
DAGENS ARENA
När den organiserade brottsligheten etableras i **nätverk** över hela landet måste också polisens och rättsväsendets arbete organiseras om.
Nätverket som är spridda över hela landet.
Det finns starka militär-industriella och politiska **nätverk** som arbetar hårt för att fixera tanken på ett nytt kallt krig.
Storstäder orienterar sig mot ett **nätverk** av andra större städer även om dessa är geografiskt avlägsna.
Och Socialistinternationalen är ett stort politisk **nätverk** med medlemspartier och associerade rörelser och partier över hela världen.
Igår basunerade femton tunga borgerliga politiker ut på DN-debatt att de startar **nätverk** mot FRA-lagen.
Ett antal kommuner och två regioner har gått ihop i **Nätverket** för Västkustbanan.
Jag ville för ett par veckor sedan få kontakt med någon på **nätverket**.
Bara ett pressmeddelande adresserat "Till media" och en hemsdeadress, med hälsningar från **Nätverket** för Västkustbanan.
Men det är märkligt att det **nätverk** som snabbt har bildats, Stoppa matchen, gör israeliska idrottsmäns deltagande i Davis Cup till en så stor symbolfråga.
De socialdemokrater som är engagerade i **nätverket** för Framtidens Energi låter sig användas av borgerligheten i ett utspel som inte nödvändigtvis kommer att ge nya reaktorer, ut.
Frågan är vilka svar som vinner: mer eller mindre demokrati, starkare institutioner eller lösare **nätverk**.
Ett **nätverk** som aktivt trycker på för att den svenska regeringen ska avbryta allt militärt samarbete med Israel vore mer adekvat och symb.
Ett **nätverk** som ifrågasätter de pengar som vi svenskar tjänar på Israels krigföring och de privilegier vi som medborgare får av de pengarna.
Sveriges försvar har i realiteten blivit en pusselbit i Natos militära **nätverk**.
Julian Assange må vara Wikileaks posterboy, men utan det **nätverk** som utför Wikileaks arbete skulle ingen ens veta vem han är.
Men de **nätverk** som möjliggör handelsrelationer mellan individer är i en betydligt mer utsatt position.
I Sverige återfinns ett antal debattörer, det (FP) - anslutna **nätverket** Liberal mångfalds sekreterare bland annat, som har uppgitt att det avgörande i en asylprocess ska vara den sökandes värderingar.
Vi i den Öppna Kris Kommissionen har alla vår bas i det progressiva **nätverket** NetRoots.
Ett företag som eBay, i egenskap av etablerat globalt **nätverk**, lär bli kvar.
Större problem är dock den "dolda lobbyismen", som bygger på en anda av samförstånd och täta **nätverk** mellan politiska makthavare, Försvarsmakten och vapenindustrin.
slut behöver fler egna experter, större förmåga att ta vara på andras kunskap och bredare personliga **nätverk**.
Böcker, cyklar och datorspel byts eller säljs via globala **nätverk** som fungerar som Myrorna.

1 2 3 4 5 6 7 8 9 10
Gå till sida: av 4 028

Ladda ner träffsida som...

Dagens Arena

TEXTATTRIBUT

Blingbring:

- otid
- ärefall
- otro

Svenskt Frasnät:

- Delimitation_of_diversity
- Resurrection
- Expressing_publicly

läsbarhetsindex: 42.45
ordvariationsindex: 78.40
nominalkvot: 1.22
datum: 2008-09-09 16:39:00
insamlingsdatum: 2022-01-11
författare: Per Wirtén
ID: coverage-satt-inte-bara-press-pa-ryssland
tags: /tom/
titel: Sätt inte bara press på Ryssland
URL: <https://www.dagensarena.se/opinion/satt-styckestyp: /tom/>

ORDATTRIBUT

betydelse:

- nätverk

sammansatta ordformer:

- nät+verk

sammansättningar:
nät (substantiv) + verk (substantiv)
universella features:
Case=Nom
Definite=ind
Gender=Neut
Number=Plur

dependensrelation: Egentligt subjekt
attityd: neutral
Blingbring:

- genomskinligghet
- korsning
- slingring

Svenskt Frasnät:

- Network

KWIC
Statistik
OrdBild

☐ Visa kontext
Träffar per sida: 25
Sortera varje korpus på: förekomst

Antal träffar: 100 696

1
2
3
4
5
6
7
8
9
10

Gå till sida: av 4 028

8 SIDOR (stödjer ej utökad kontext)

Daghem i Frankrike ska inte längre få ha trådlösa **nätverk** till sina datorer.
Nätverken i skolor för barn upp till sjätte klass ska stängas av när de inte används.
DAGENS ARENA
När den organiserade brottsligheten etableras i **nätverk** över hela landet måste också polisens och rättsväsendets arbete organiseras om.
Nätverket som är spridda över hela landet.
Det finns starka militär-industriella och politiska **nätverk** som arbetar hårt för att fixera tanken på ett nytt kallt krig.
Storstäder orienterar sig mot ett **nätverk** av andra större städer även om dessa är geografiskt avlägsna.
Och Socialistinternationalen är ett stort politisk **nätverk** med medlemspartier och associerade rörelser och partier över hela världen.
Igår basunerade femton tunga borgerliga politiker ut på DN-debatt att de startar **nätverk** mot FRA-lagen.
Ett antal kommuner och två regioner har gått ihop i **Nätverket** för Väst kustbanan.
Jag ville för ett par veckor sedan få kontakt med någon på **nätverket**.
Bara ett pressmeddelande adresserat "Till media" och en hemsdeadress, med hälsningar från **Nätverket** för Väst kustbanan.
Men det är märkligt att det **nätverk** som snabbt har bildats, Stoppa matchen, gör israeliska idrottsmäns deltagande i Davis Cup till en så stor symbolfråga.
De socialdemokrater som är engagerade i **Nätverket** för Framtidens Energi låter sig användas av borgerligheten i ett utspel som inte nödvändigtvis kommer att ge nya reaktorer, ut
Frågan är vilka svar som vinner: mer eller mindre demokrati, starkare institutioner eller lösare **nätverk**.
Ett **nätverk** som aktivt trycker på för att den svenska regeringen ska avbryta allt militärt samarbete med Israel vore mer adekvat och symboliskt.
Ett **nätverk** som ifrågasätter de pengar som vi svenskar tjänar på Israels krigföring och de privilegier vi som medborgare får av de pengarna.
Sveriges försvar har i realiteten blivit en pusselbit i Natos militära **nätverk**.
Julian Assange må vara Wikileaks posterboy, men utan det **nätverk** som utför Wikileaks arbete skulle ingen ens veta vem han är.
Men de **nätverk** som möjliggör handelsrelationer mellan individer är i en betydligt mer utsatt position.
I Sverige återfinns ett antal debattörer, det (FP) - anslutna **nätverket** Liberala mångfalds sekreterare bland annat, som har uppgitt att det avgörande i en asylprocess ska vara den sökandes värderingar.
Vi i den Öppna Kris Kommissionen har alla vår bas i det progressiva **nätverket** NetRoots.
Ett företag som eBay, i egenskap av etablerat globalt **nätverk**, lär bli kvar.
Större problem är dock den "dolda lobbyismen", som bygger på en anda av samförstånd och täta **nätverk** mellan politiska makthavare, Försvarsmakten och vapenindustrin.
slutet behöver fler egna experter, större förmåga att ta vara på andras kunskap och bredare personliga **nätverk**.
Böcker, cyklar och datorspel byts eller säljs via globala **nätverk** som fungerar som Myrorna.

1
2
3
4
5
6
7
8
9
10

Gå till sida: av 4 028

Ladda ner träffsida som...

Dagens Arena

TEXTATTRIBUT

Blingbring:

- otid
- äfterfall
- otro

Svenskt FrasNät:

- Delimitation_of_diversity
- Resurrection
- Expressing_publicly

läsbarhetsindex: 42.45
ordvariationsindex: 78.40
nominalkvot: 1.22
datum: 2008-09-09 16:39:00
insamlingsdatum: 2022-01-11
författare: Per Wirtén
ID: coverage-satt-inte-bara-press-par-ryssland
tags: /tom/
titel: Sätt inte bara press på Ryssland
URL: https://www.dagensarena.se/opinion/samfundet/2022-01-11/tydskestyp: /tom/

ORDATTRIBUT

betydelse:

- nätverk

sammansatta ordformer:

- nät+verk

sammansättningar:
nät (substantiv) + verk (substantiv)
universella features:
Case=Nom
Definite=ind
Gender=Neut
Number=Plur

dependensrelation: Egentligt subjekt
attityd: neutral
Blingbring:

- genomsiktighet
- korsning
- slingring

Svenskt FrasNät:

UNIVERSITY OF GOTHENBURG | SPRÅKBANKENTEXT

11

Title: Regeringskrisen i Sverige 2014

Corpus name ID: SWEDISH WIKIPEDIA

Document size: 1108 tokens

Year: 2022

Close Document

Most common nouns:

regering (16) december (15) Sverigedemokrat (10) sverigedemokrat (10) parti (10) allians (9) statsminister (8) budget (8)
budgetförslag (7) regeringskris (6)

Most common names:

Stefan Löfven (14) Sverigedemokraterna (9) Sverige (5) riksdagen (5) Miljöpartiet (4)

Document view

Statistics view

Explore

word attributes ▾

Sentiment label ▾

Välj från lista (3 alternativ)

Search the current document

Neutral (344)

Positive (36)

Negative (10)

1 Regeringskrisen i Sverige 2014

2

3 Regeringskrisen i Sverige 2014 inleddes den 2 december.

4 Sverigedemokraterna hade meddelat att partiet avsåg att rösta för

5 Alliansens budgetförslag. En sådan röstning skulle leda till att

6 regeringens budgetproposition skulle fällas den 3 december.

7 Statsminister Stefan Löfven **förklarade** att regeringen avsåg att den 298 december utlysa **extra** val den 22 mars 2015. Den 27 december träffade

9 dock regeringen och oppositionspartierna i Alliansen den så kallade

10 decemberöverenskommelsen, enligt vilken partierna bland annat **lovade**11 att **underlätta** för minoritetsregeringar att **få** igenom sin budget i

12 riksdagen. Samtidigt gav statsministern besked om att regeringen inte

13 hade för avsikt att utlysa **extra** val.

14

15 Detta var den första regeringskrisen i Sverige efter regeringen

16 Carlssons avgång 1990. Det i början av december **aviserade extra** valet,17 skulle ha blivit det första **extra** valet efter andrakammarvalet 1958

18 efter ATP-omröstningen. Att en budgetproposition röstades ned i sin

19 helhet i riksdagen hade aldrig hänt i **modern** tid.

20

21 Bakgrund

22

23 Resultatet av riksdagsvalet i september 2014 innebar en parlamentarisk

24 situation där Socialdemokraterna, Miljöpartiet och Vänsterpartiet

Text attributes

TEXT CLASSIFICATION (BLINGBRING)

kompromiss

endrätt

samverkan

TEXT CLASSIFICATION (SWEFN)

Alliance

Make possible to do

Win prize

READABILITY MEASURE (LIX)

52.10

READABILITY MEASURE (OVIX)

60.36

READABILITY MEASURE (NK)

2.04

URL

[https://sv.wikipedia.org/wiki/Regeringskrisen i Sverige 2014](https://sv.wikipedia.org/wiki/Regeringskrisen_i_Sverige_2014)

DATE

2022-02-13T05

Title: Regeringskrisen i Sverige 2014
Corpus name ID: SWEDISH WIKIPEDIA
Document size: 1108 tokens
Year: 2022

Close Document

Most common nouns: regering (16) december (15) Sverigedemokrat (10) sverigedemokrat (10) parti (10) allians (9) statsminister (8) budget (8)
budgetförslag (7) regeringskris (6)
Most common names: Stefan Löfven (14) Sverigedemokraterna (9) Sverige (5) riksdagen (5) Miljöpartiet (4)

Document view

Statistics view

Explore

word attributes

Sentiment label

Välj från lista (3 alternativ)

Search the current document

Neutral (344)

Positive (36)

Negative (10)

1 Regeringskrisen i Sverige 2014
2
3 Regeringskrisen i Sverige 2014 inleddes den 2 december.
4 Sverigedemokraterna hade meddelat att partiet avsåg att rösta för
5 Alliansens budgetförslag. En sådan röstning skulle leda till att
6 regeringens budgetproposition skulle fällas den 3 december.
7 Statsminister Stefan Löfven **förklarade** att regeringen avsåg att den 29
8 december utlysa **extra** val den 22 mars 2015. Den 27 december träffade
9 dock regeringen och oppositionspartierna i Alliansen den så kallade
10 decemberöverenskommelsen, enligt vilken partierna bland annat **lovade**
11 att **underlätta** för minoritetsregeringar att **få** igenom sin budget i
12 riksdagen. Samtidigt gav statsministern besked om att regeringen inte
13 hade för avsikt att utlysa **extra** val.
14
15 Detta var den första regeringskrisen i Sverige efter regeringen
16 Carlssons avgång 1990. Det i början av december aviserade **extra** valet,
17 skulle ha blivit det första **extra** valet efter andrakammarvalet 1958
18 efter ATP-omröstningen. Att en budgetproposition röstades ned i sin
19 helhet i riksdagen hade aldrig hänt i **modern** tid.
20
21 Bakgrund
22
23 Resultatet av riksdagsvalet i september 2014 innebar en parlamentarisk
24 situation där Socialdemokraterna, Miljöpartiet och Vänsterpartiet
25 tillsammans blev större än de fyra borgerliga allianspartierna, men
26 ändå var i minoritet. Sverigedemokraterna ökade sin representation och
27 blev det tredje största partiet i riksdagen, vilket innebar att partiet
28 förstärkte sin vägmästarroll. Samtliga övriga partier gjorde
29 utfästelser att inte **samarbeta** med Sverigedemokraterna.
30

Text attributes

TEXT CLASSIFICATION (BLINGBRING)

kompromiss
endräkt
samverkan

TEXT CLASSIFICATION (SWEFN)

Alliance
Make possible to do
Win prize

READABILITY MEASURE (LIX)

52.10

READABILITY MEASURE (OVIX)

60.36

READABILITY MEASURE (NK)

2.04

URL

https://sv.wikipedia.org/wiki/Regeringskrisen_i_Sverige_2014

DATE

2022-02-13T05

TITLE

Regeringskrisen i Sverige 2014

CATEGORIES

Artiklar som behöver förtydligas-samtliga
Artiklar som behöver förtydligas-

samtliga#Regeringskrisen%20i%20Sverige%202014

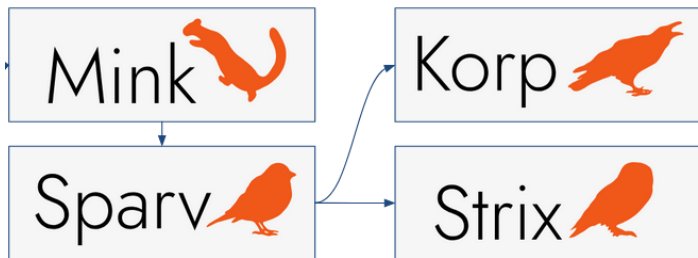
Korpusarbete på Språkbanken Text

Språkteknologisk analys via Sparv



Mink

Låter användare importera, processa och arbeta vidare med egna datamängder.



Mer om Mink i eftermiddag!

Slutord

Språkbanken Text stödjer forskning (inom, bland annat, språkteknologi, språkvetenskap, samt digital humaniora och samhällsvetenskap) genom att tillämpa öppenhetsprincipen på:

- datan och metadatan;
- analyserna;
- plattformarna; samt
- all programkod vi producerar.

<https://spraakbanken.gu.se/>

Tack för uppmärksamheten!