



VÅR HISTORIA

En redogörelse för Språkbanken Texts historia måste börja med Sture Alléns (1928–2022) banbrytande insats. Han var en föregångare i att introducera korpuslingvistik i Sverige för svenska språket. Hans doktorsavhandling från 1965 publicerades i två delar, en där han beskrev den datorstödda metod som han hade använt – efter att först ha lärt sig programmera den själv i maskinkod – för att undersöka en textkorpus av 1600-talsbrev, den andra en vetenskaplig utgåva av dessa brev.

Efter att ha försvarat sin avhandling initierade Allén ett projekt som syftade till att bana väg för korpusbaserad lexikografi för svenska. Det mest omedelbara resultatet av detta projekt var en textkorpus med en miljon ord svenska nyhetstexter, som tillhandahöll råmaterialet för en rad svenska ordböcker.

Som professor och vetenskaplig ledare för Enheten för datorlingvistik, som etablerades 1972 vid Göteborgs universitet, tog Allén initiativ till ett grundutbildningsprogram i datorlingvistik som startade vid universitetet 1984. Hans eget huvudfokus förblev dock utvecklingen av korpusar och korpusverktyg till stöd för svensk lexikografi, och han initierade ett systematiskt arbete för att bygga en datorstödd forskningsinfrastruktur som kunde främja detta mål.

Planerna för Språkbanken drogs upp i en debattartikel skriven av Allén för den svenska dagstidningen Dagens Nyheter i september 1970. År 1973 lämnade Enheten för datorlingvistik in en formell anhållan till Utbildningsdepartementet, där de begärde öronmärkta medel för det som skulle bli Språkbanken. Två år senare blev denna forskningsinfrastruktur verklighet, när Logoteket (som den kallades i början) etablerades med nationell finansiering 1975. Fokus för Språkbanken förändrades märkbart runt

sekelskiftet, när de lexikografiska och språktekhnologiska aktiviteterna av olika skäl skildes åt organisoriskt. De förra kom att bedrivas vid Lexikaliska institutet som etablerades i samband med detta, medan Språkbanken utvidgade sin språktekhnologiska verksamhet långt bortom det lexikografiska.

Sedan dess har Språkbanken Text vuxit till en nationellt och internationellt erkänd forsknings- och utvecklingsenhet för svensk språktekhnologi och språkresurser. Den koordinerade de svenska aktiviteterna inom den europeiska forskningsinfrastrukturen CLARIN ERIC 2014–2024, och är den koordinerande noden för den nationella forskningsinfrastrukturen Språkbanken. Språkbanken Text är en av dess fyra nationellt distribuerade avdelningar, där de andra tre är: talteknologiavdelningen (Språkbanken Tal) vid Kungliga Tekniska högskolan (KTH) i Stockholm, avdelningen för kulturarv och språkpolitik (Språkbanken Sam) vid Institutet för språk och folkmitten i Uppsala, Stockholm och Göteborg, och avdelningen som koordinerar de svenska CLARIN-aktiviteterna (Språkbanken CLARIN) vid Uppsala universitet.

Som forskningsinfrastruktur är Språkbanken relativt unik då många av forskningsresultaten som kommer ut ur den forskning den stödjer i stor utsträckning bidrar till den fortsatta utvecklingen av infrastrukturen själv. Språkbanken stödjer forskning inom språktekhnologi (text, tal och tecken) med en infrastruktur som i sig själv är byggd på språktekhnologi (text, tal och tecken), mycket likt den antika mytologins Ouroborosorm.

Lars Borin



DET HÄR ÄR SPRÅKBANKEN TEXT

På Göteborgs universitet finns **Språkbanken Text**. Vi finns på Institutionen för svenska, flerspråkighet och språkteknologi.

Språkbanken Text är en **forskningsinfrastruktur** för språkliga data och en språkteknologisk forskningsenhet. Vi är en del av Språkbanken, en nationell infrastruktur till stöd för forskning baserad på språkliga data.

Vi utvecklar, förädlar och tillgängliggör fria språkliga **forskningsdata** enligt FAIR-principerna, med ett särskilt fokus på svenska språket genom tiderna.

Vi utvecklar och tillgängliggör fria digitala **forskningsplattformar** (Korp, Karp, Mink, Sparv och Strix), där vi strävar efter att stödja alla typer av forskning där språkliga data är centrala.

Vi bedriver egen **forskning** i språkteknologi, inklusive språkbaserad AI, samt deltar i projekt inom andra discipliner. Våra samarbeten sker både nationellt och internationellt och inkluderar minnes- och kulturarvsinstitutioner, skolor och företag.

Språkbanken inrättades redan 1975, vilket gör oss till en av världens äldsta forsknings- och utvecklingsenheter inom språkteknologi.

Läs mer om oss på vår hemsida:

www.spraakbanken.gu.se



SPRÅKBANKENTEXT
En forskningsinfrastruktur för språkliga data
och en språkteknologisk forskningsenhet



STUDERA SPRÅKTEKNOLOGI

Språkbankens personal undervisar aktivt på flera kurser inom **Masterprogrammet i språkteknologi**, med kursansvar för kurserna

- Introduktion till programmering
- Grundläggande färdigheter för språkteknologi
- Teman i datalingvistik och språkteknologi
- Språkteknologiresurser
- Masterprojekt 15hp/30hp

Vi undervisar också i **kurser i lingvistik** vid Institutionen för filosofi, lingvistik och vetenskapsteori, samt **handleder uppsatser** på kandidat- och magisternivå, till exempel inom programmen för Språkteknologi, Applied data science, Datavetenskap och Svenska som andraspråk.

Vi har även undervisat inom Masterprogrammet i digital humaniora.

Vi har två representanter i programkommittén för Masterprogrammet i språkteknologi, där vi är delaktiga i programmets utveckling och antagning av studenter.

Doktorander

På Språkbanken Text arbetar också flera doktorander från olika delar av världen.. De forskar alla inom språkteknologi.





LÄS MER OM SPRÅKBANKEN

SPRÅKBANKEN

- AKTUELLT
- VERKTYG OCH RESURSER
- VAD ÄR SPRÅKTEKNOLOGI?
- MÅNADENS PROFIL
- HÖSTWORKSHOP
- OM OSS
- BLOGG
- NYHETSREV
- Q
- EN

Vi ger forskare nya möjligheter med hjälp av språkteknologi.

Språkbanken skapar möjligheter att forska i digitala text- och talmaterial med hjälp av verktyg och metoder i gränslandet mellan språkteknologi och AI.

Nyheter



27 februari 2025
Ny professor vill arbeta för en säker och etisk användning av AI

Den 13 februari utnämndes Elena Volodina till professor i språkteknologi vid Göteborgs universitet. Hennes forskning fokuserar främst på hur stora tex...

Kalender

<div style="border: 1px solid #ccc; border-radius: 50%; width: 30px; height: 30px; display: flex; align-items: center; justify-content: center;">2 mars</div> <p>RESOURCEFUL-2025</p> <p>Den 2 mars 2025 arrangeras den tredje upplagan av den internationella workshoppen RESOURCEFUL som fokuserar på språk med begränsade resurser.</p>	<div style="text-align: center; border: 1px solid #ccc; border-radius: 50%; width: 30px; height: 30px; margin-bottom: 5px;"></div> <p>NoDaLiDa/Baltic-HLT 2025</p> <p>2-5 mars 2025 sammankomma NoDaLiDa och Baltic-HLT - den 35:e nordiska forskarsymposionen.</p>
--	---

sprakbanken.se

SPRÅKBANKENTEXT

En forskningsinfrastruktur för språkliga data och en språkteknologisk forskningsenhet

English

Språkbanken Text är en avdelning inom Språkbanken.

Aktuell • Forskning • Data • Analyser • Plattformar • FAQ • Om oss • Kontakt

Forska 23 varor 7 döksträvar • Forskningsprojekter: 13 • Aktiva projekt: 9 • Datamängder: 1249 • Analyser: 51 • Datapunkter: >33.7G

Det här är Språkbanken Text. Vi utvecklar, fridrar och tillgängliggör fria språkliga forskningsdata, med ett särskilt fokus på svenska språket genom lidens. Vi utvecklar och tillgängliggör via digitala forskningsplattformar, där vi strävar efter att slösa alla typer av fordon för språkliga data är centrala. Vi bedriver egen forskning i språkteknologi, inklusive språkbaserad AI, samt deltar i projekt inom andra discipliner.

Plattformar



Korp
Språkbankens ordforskningsplattform



Karp
Språkbankens datorledringsplattform



Sparv
Språkbankens analysplattform

Nyheter

Ny professor vill arbeta för en säker och etisk användning av AI →
2025-02-19

Månadens profil: Fredrik Nylen →
2025-02-19

Hur fungerar Mink? Välkommen på en virtuell rundtur! →
2025-02-11

Alla aktuella nyheter

Blogg

Datamängder

Senast uppdaterade datamängder, körningar, lexikon, träningsdata och modeller

Fölk →
2025-01-10 Korpus

Learnert Language →
2025-01-10 Korpus, Samling

MultIGEC →
2025-01-10 Korpus, Tränings- och utvärderingsdata

MuCoGED →
2025-01-10 Korpus, Tränings- och utvärderingsdata

DaLAJ-GED-Superlim 2.0 →
2025-01-19 Korpus, Tränings- och utvärderingsdata

Projekt

Aktuella projekt med extern finansiering

Språkliga närläv, inom och mellan språk →
Riksbankens Jubileumstond (M21-021)

Change is Key! →
Riksbankens Jubileumstond (M21-021)

Mormor Karl →
Venerisupplägg (2022-02311)

HUMINFRÅ →
Venerisupplägg (2021-00176, 2023-00171)

www.spraakbanken.gu.se



VÅRA PLATTFORMAR

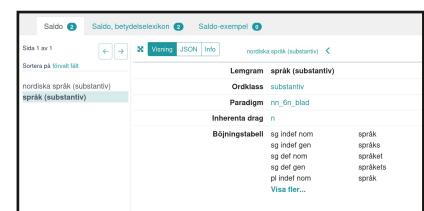
Vi utvecklar en språkteknologisk forskningsinfrastruktur genom att fritt tillgängliggöra språkdata och skapa fria plattformar för att utforska dem.



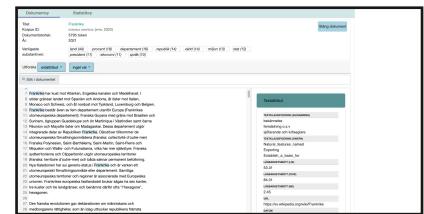
Korp är Språkbankens Textsordforskningsplattform där du kan söka i stora mängder text från bland annat dagstidningar, skönlitteratur och sociala medier. Textsamlingarna, korpusarna, är annoterade med språkliga och andra attribut.



Karp är Språkbanken Texts dataredigeringsplattform där du kan redigera strukturerade data, exempelvis lexikon. Karp tillåter sökningar baserade på ordformer, böjningar samt komplexa frågor som involverar flera kriterier. I Karp finner du ett flertal lexikon, framför allt på svenska.



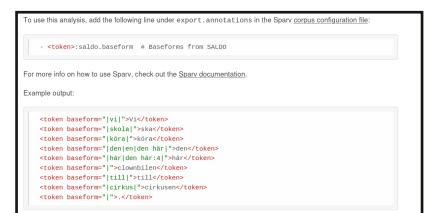
Strix är Språkbanken Texts textforskningsplattform där du kan utforska stora mängder språkteknologiskt förädlade texter på dokumentnivå. Med Strix kan du söka i en bred samling av korpusar, ta fram statistik och utforska liknande dokument genom vektorsökning.



Mink är Språkbanken Texts dataplattform där du importerar egna textsamlingar, gör analyser och skapar korpusar som du kan undersöka med våra andra verktyg. Mink fungerar på många sätt som ett webbgränssnitt för Sparv.



Sparv är Språkbanken Texts analysplattform och består av en pipeline för korpusannotering, ett webb-API samt ett webbgränssnitt. Våra korpusar är mestadels automatiskt uppmärksat med Sparv.



SPRÅKBANKENTEXT

En forskningsinfrastruktur för språkliga data
och en språkteknologisk forskningsenhet



VI ÄR FYRA AVDELNINGAR

Språkbanken Text är en forskningsinfrastruktur för språkliga data och en språktekologisk forskningsenhet. Vi utvecklar, förädlar och tillgängliggör fria språkliga forskningsdata enligt FAIR-principerna, med ett särskilt fokus på svenska språket genom tiderna.

Plats: Institutionen för svenska, flerspråkighet och språktekologi, Göteborgs universitet

Språkbanken Sam gör Institutet för språk och folkminnens arkivmaterial mer tillgängligt för både forskare och allmänhet. Det sker med hjälp av verktyg och metoder från språktekologi och digital humaniora.

Plats: Institutet för språk och folkminnen, Uppsala

Språkbanken Tal utvecklar, förvaltar och distribuerar talteknologiska resurser för svensk talforskning och svensk talteknologisk forskning.

Plats: Avdelningen för musik, tal och hörsel, Kungliga Tekniska Högskolan, Stockholm

Språkbanken-CLARIN utvecklar och tillhandahåller nationell och europeisk infrastruktur för tal- och textbaserad e-vetenskap,

Plats: Uppsala universitet



PROJEKT OCH SAMARBETEN

exempel på projekt och samarbeten

vad vill vi lyfta

Månadens projekt? (Mia). Jag kan höra om hon har fått
några tips.

Sasha och Dana?

intervjuerna i Månadens profil: <https://xn-sprkbanken-35a.se/manadens-profil>



DETTA ÄR SPRÅKTEKNOLOGI



SPRÅKBANKENTEXT
En forskningsinfrastruktur för språkliga data
och en språkteknologisk forskningsenhet



FLER VERKTYG FRÅN OSS

Vi utvecklar en språkteknologisk forskningsinfrastruktur genom att fritt tillgängliggöra språkdata och skapa fria plattformar för att utforska dem.

några av de andra avdelningarnas verktyg (Mia kanske kan välja ut vilka):
<https://språkbanken.se/verktyg-och-resurser>

Folke

Sägenkartan

Dialektkartan