

SPRÅKBANKEN: EN HISTORIK

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 språkteknologiska aktiviteterna av olika skäl skildes åt organisatoriskt. De förra kom att bedrivas vid Lexikaliska institutet som etablerades i samband med detta, medan Språkbanken utvidgade sin språkteknologiska verksamhet långt bortom det lexikografiska.

Sedan dess har Språkbanken Text vuxit till en nationellt och internationellt erkänd forsknings- och utvecklingsenhet för svensk språkteknologi 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 folkminnen 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 språkteknologi (text, tal och tecken) med en infrastruktur som i sig själv är byggd på språkteknologi (text, tal och tecken), mycket likt den antika mytologins Ouroborosorm.

Lars Borin

SPRÅKBANKEN
En forskningsinfrastruktur för språkliga data

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



SPRÅKBANKEN TEXT
En forskningsinfrastruktur för språkliga data
och en språkteknologisk forskningsenhet

UNDERVISNING

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

- Introduktion till programmering
- Grundläggande färdigheter för språkteknologi
- Teman i datalingsvistik 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.

Studierektor i Språkteknologi är Gerlof Bouma på Språkbanken Text.

Doktorander

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



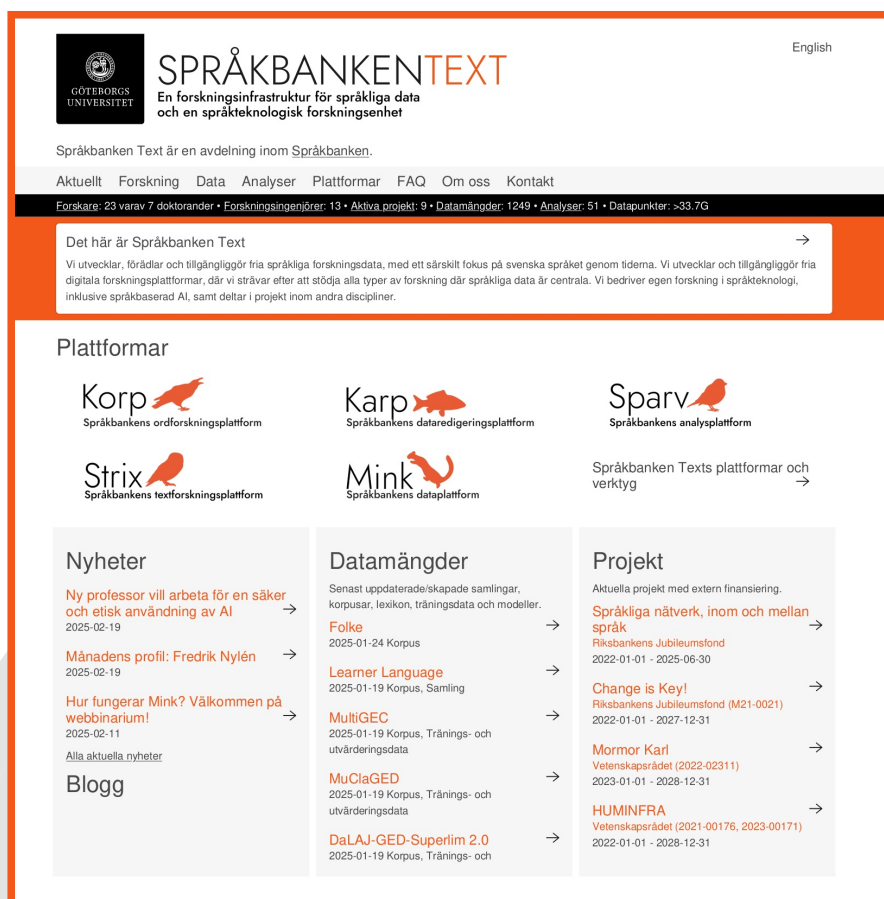
SPRÅKBANKENTEXT
En forskningsinfrastruktur för språkliga data
och en språkteknologisk forskningsenhet

LÄS MER OM SPRÅKBANKEN



The screenshot shows the homepage of Språkbanken. At the top is a navigation bar with links: SPRÅKBANKEN, AKTUELLT, VERKTYG OCH RESURSER, VAD ÄR SPRÅKTEKNOLOGI?, MÅNADENS PROFIL, HÖSTWORKSHOP, OM OSS, BLOGG, NYHETSREVIS, and a search icon. The main content area features a large orange banner with the headline "Vi ger forskare nya möjligheter med hjälp av språkteknologi." and an illustration of a person sitting at a desk with a speech bubble. Below the banner, there are two columns: "Nyheter" (News) and "Kalender" (Calendar). The "Nyheter" section includes a news item about a professor's appointment. The "Kalender" section lists upcoming events like "RESOURCEFUL 2025" and "NoDaLiDa/Baltic-HLT 2025".

sprakbanken.se



The screenshot shows the Språkbanken Text website. The header includes the Göteborgs Universitet logo, the title "SPRÅKBANKENTEXT", and the subtitle "En forskningsinfrastruktur för språkliga data och en språkteknologisk forskningsenhet". A navigation bar lists: Aktuellt, Forskning, Data, Analyser, Plattformer, FAQ, Om oss, and Kontakt. Below the navigation bar, there is a section "Det här är Språkbanken Text" with a description of the platform. The main content area is divided into three columns: "Plattformer" (Platforms) listing Korp, Sparv, Strix, and Mink; "Nyheter" (News) listing recent news items; and "Projekt" (Projects) listing ongoing projects. The footer contains the website address "www.spraakbanken.gu.se".

www.spraakbanken.gu.se

SPRÅKBANKEN
En forskningsinfrastruktur för språkliga data

SPRÅKBANKENTEXT
En forskningsinfrastruktur för språkliga data
och en språkteknologisk forskningsenhet