



PROGRAM DEN 18 JUNI

- 13:15 **Praktisk information**, Konferenscentrum Wallenberg
Moderatorerna och vicektör **hälsar alla välkomna!**
Susanna Karlsson, prefekt vid Institutionen för svenska, flerspråkighet och språkteknologi, Andrea Castro, vicedekan vid Humanistiska fakulteten, Max Petzold, vicektör Göteborgs universitet
- 13.30 **Språkbanken från lexikografi till AI.**
Lars Borin, föreståndare Språkbanken
- 14.00 **Språkbanken och Svenska Akademien – en lång och utvecklande relation.** Mats Malm, ständig sekreterare för Svenska Akademien
- 14.30 **SAOL vs SO: ett mångordigt battle.** Louise Holmer, huvudredaktör för SAOL och Emma Sköldberg, huvudredaktör för Svensk ordbok (SO)
- 15.00 **Paus med fika och tid att se utställningen**
- 15.30 **Svenskt teckenspråk på SU från 1972.** Johanna Mesch, professor i teckenspråk vid Stockholms universitet
- 16:00 **På Tal om Språkbanken.** Jens Edlund, föreståndare Språkbanken Tal
- 16:30 **Förnuft och känsla. Hur kan man använda språkteknologiska verktyg för att göra bruklighetsbedömningar?** Susanna Karlsson, prefekt vid Institutionen för svenska, flerspråkighet och språkteknologi
- 17.00 **Avslutning i Wallenbergssalen**
- 17:05 **Mingel i foajén och poängpromenad**
- 18:00– **Trerätersmiddag.** Välkommen till restaurangen på plan 2!

Tack till:



SVENSKA
AKADEMIEN

SPRÅKBANKEN
En forskningsinfrastruktur för språkliga data



Vi dokumenterar!



Vi fotograferar under dagen och kvällen för publicering på Språkbankens och Göteborgs universitets mediekanaler.

Om du inte vill vara med på bild så respekterar vi självklart det! Meddela oss på sb-info@svenska.gu.se.



SPRÅKBANKENTEXT
En forskningsinfrastruktur för språkliga data
och en språkteknologisk forskningsenhet





NÅGRA TIDIGA ÅRTAL

1969

Språkdata, "ett system för språklig databehandling" (med programbibliotek, grunderna till ett textarkiv, en lexikonserie och en ordbank).

1970

Språkdata, "forskningsgruppens programsystem".

Datatext, "ett arkiv innehållande texter i maskinläsbar form".

Logotek, "ett organ som tillvaratar och tillhandahåller maskinläsbar text".

1971

Logoteket. Består av ett textarkiv, en ordbank och språkvetenskapliga bearbetningar.

1973

Ett förslag om inrättandet av "ett organ för lagring och tillhandahållande av datamaskinellt läsbara texter, benämnt logotek" skickas in "Till Konungen, Utbildningsdepartementet".

1975

Logoteket inrättas.

1983

Namnbyte till Språkbanken då arbetet vidgades och fördjupades.

Källa: Rolf Gavare



Vi ger forskare nya möjligheter med hjälp av språkteknologi

SPRÅKBANKENTEXT

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

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

SPRÅKBANKENSAM

Språkbanken Sam arbetar för att göra Isofs språkliga material mer tillgängligt för forskare och allmänhet med hjälp av verktyg och metoder från språkteknologi och digital humaniora.

Plats: Institutet för språk och folkminnen, Isof, Göteborg, Stockholm och Uppsala

SPRÅKBANKENTAL

Språkbanken Tal utvecklar, förvaltar och distribuerar talteknologiska resurser för svensk talforskning och svensk talteknologisk forskning. Språkbanken Tal är en forskningsenhet vid KTH.

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

SPRÅKBANKENCLARIN

Språkbanken CLARIN är en nod i CLARIN – Common Language Resources and Technology Infrastructure. Språkbanken CLARIN koordineras av Uppsala universitet, och vänder sig till forskare och andra intresserade av digitala metoder och material inom humaniora och samhällsvetenskap.

Plats: Uppsala universitet

SPRÅKBANKEN
En forskningsinfrastruktur för språkliga data



PROJEKT OCH SAMARBETEN

Språkbanken driver egna projekt och samarbetar med andra aktörer för att utveckla språkteknologi. Här visar vi några exempel.

Braxen. Uttalexikonet Braxen har utvecklades under nästan två decennier av Myndigheten för tillgängliga medier, MTM för internt bruk. Tillsammans med Språkbanken Tal har MTM nu släppt det fritt för nedladdning av forskare, talsyntesföretag, förlag och andra. Lexikonet utökas kontinuerligt med nya nyhetsord och facktermer. Det kan användas för talforskning och för att hjälpa syntetiska röster att uttala saker rätt, till exempel latinska termer och utländska namn. I lexikonet finns idag cirka 850 000 ord och namn.

KB-Whisper. KB-Whisper är en tal-till-textmodell som har tränats på 50 000 timmar tal för att lära sig omvandla talat språk till text. KB-Whisper har tagits fram av Kungliga Biblioteket som bland annat har tränat AI-modellen på äldre dialektinspelningar från Institutet för språk och fornminnen som digitaliseras av Språkbanken Sam.

Mormor Karl är 27 år. Projektet Mormor Karl är 27 år syftar till att skapa språktechnologiska algoritmer som kan upptäcka personuppgifter och känslig information i stora textmassor och automatiskt ersätta detta med pseudonymer. På så sätt kan personuppgifter skyddas och alla texter användas i olika slags forskning.

Ryktesminering. Projektet Ryktesminering, syftar till att öka förståelsen för attityder och värderingar till vaccin och undersöka hur rykten om vaccin etableras och sprids på sociala medier. Projektet är ett samarbete med Lunds universitet.

Språkteknologi för nationella minoritetsspråk.

Språkbanken Sam arbetar med att utveckla språkteknologi som rättstavningsprogram och tangentbord för de nationella minoritetsspråken. Arbetet görs i samarbete med Giellatekno vid Universitetet i Tromsö.

SWENER-1800. Ett digitalt textmaterial med en halv miljon ord från 1700-talet till år 1900. Materialet kan användas för att träna datorprogram att känna igen historiska namn på personer och platser i en text och för att utvärdera hur väl språkmodeller hanterar historisk svenska. SWENER-1800 är ett samarbete inom Språkbanken CLARIN mellan Uppsala universitet, Riksarkivet och Språkbanken Text.

Svensk diakronisk korpus. En digital textsamling med cirka 16 miljarder ord som sträcker sig från 1200-talet till nutid. Korpusen ger möjlighet att göra storskaliga studier av svenska språkets utveckling genom historien och kan också användas för språkhistoriska jämförelser mellan svenska och andra språk. Svensk diakronisk korpus är ett samarbete inom Språkbanken CLARIN mellan Uppsala universitet och Språkbanken Text.

Svensk teckenspråkskorpus. Består just nu av 24 timmar inspelat material och över 200 000 tecken med annoteringar om ordklass och detaljerad information om hur ett tecken utförs. Korpusen bidrar med nya tecken till Svenskt teckenspråkslexikon, kan användas språkundervisning och har även stor betydelse för framtida teckenspråkforskning. Svensk teckenspråkskorpus är ett samarbete mellan Språkbanken CLARIN och Stockholms universitet.

SuperLim 2.0. En plattform för att testa och analysera svenska språkmodeller. Projektet är ett samarbete mellan Språkbanken Text, KB-labb, forskningsinstitutet RISE och AI Sweden.

The screenshot shows the dataset page for 'SuperLim 2.0' on the Språkbanken TEXT website. The page includes a brief description, a table of contents, and detailed metadata such as file sizes, formats, and download links. It also features a 'DOI' link and a 'Last updated' timestamp.

Swedia 2000. Mer än 100 talare av olika dialekter spelades mellan 1998 och 2000 in i Swedia 2000-projektet, som var ett omfattande samarbete mellan flera universitet. Ljudfilerna innehöll persondata som var avsedda att klippas bort och tillvaratas i efter processning, men detta skedde endast i vissa fall då projektet tog slut. Materialet kom till Språkbanken Tal 2024, och personuppgifterna är nu identifierade och borttagna, så att materialet kan göras tillgängligt för forskning både hos Språkbanken Tal och Språkbanken Sam.

SPRÅKBANKEN
En forskningsinfrastruktur för språkliga data



DETTA ÄR SPRÅKTEKNOLOGI

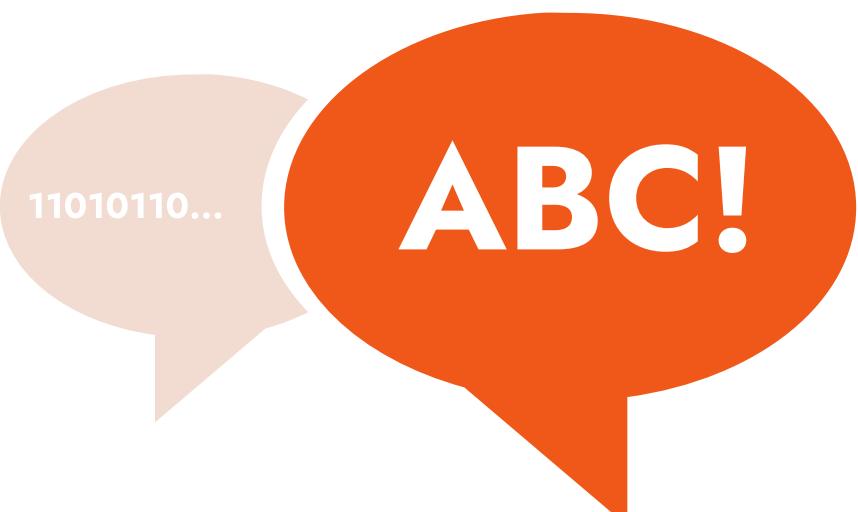
Språkteknologi handlar om hur människans språk är uppbyggt och hur en dator kan programmeras för att hantera mänskligt språk.

Språkteknologi är ett centralt område för vår språkanvändning och utvecklingen av svenska. Språktechniska produkter genomsyrar stora delar av samhället. Det handlar bland annat om dikteringsprogram, datoruppläst e-post, söktjänster på nätet, telefontaltjänster och automatisk textsammanfattning.

Målet med språkteknologi är att förenkla och förbättra kommunikationen mellan människor och mellan människor och datorer. Det kan handla om att du tack vare en automatiskt uppläst översättning kan prata svenska i telefonen med någon som pratar arabiska, att du kan prata in en text i datorn i stället för att använda tangentbordet eller att din text blir automatiskt språkrättad av datorn.

Exempel på språktechnologiska områden

- Maskinöversättning
- Informationshantering
- Skrivverktyg
- Elektronisk språkinlärning
- Dialogsystem
- Taligenkänning
- Talsyntes
- Textanalys
- Elektroniska lexikon
- Tal- och textkorpusar
- Teknik och språkresurser
- Teknik och språkresurser



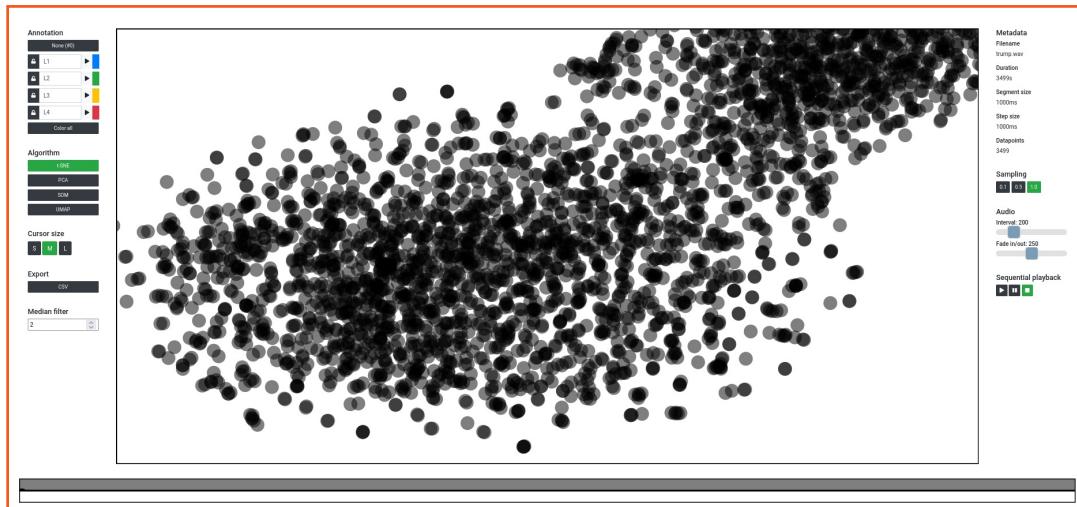
Språkteknologi utgörs av både tekniker och språkresurser (lexikon, korpusar etcetera). Teknikerna tar form som programvara, medan resurserna är den information som programvaran utnyttjar.



VERKTYG FRÅN SPRÅKBANKEN TAL

Språkbanken arbetar aktivt för att utveckla en språkteknologisk infrastruktur. Vi skapar, harmonisera och standardisera fria språkliga resurser och tar fram verktyg för att utforska resurserna och göra dem tillgängliga. Här visar vi några exempel från **Språkbanken Tal**.

Edyson. Ett verktyg som gör det möjligt att snabbt få en överblick av stora ljudmaterial utan att behöva lyssna igenom hundratals timmar eller träna modeller specifikt för uppgiften. Edyson klipper upp materialet i korta ljudsegment och bildar sedan kluster med datapunkter utifrån hur de låter. På så sätt går det att skilja ut tal från applåder, hitta pauser, perioder av tystnad, brus, musik, och många andra ljudtyper. Edyson har tagits fram av Språkbanken Tal.



TubeN. Ett lättviktsverktyg för talforskare och studenter som beräknar formantfrekvenser givet en beskrivning av en munhåla, och genererar statiska vokaler från detta. Verktyget implementerar de algoritmer Liljencrantz och Fant beskrev 1975, och kan användas både till att testa hypoteser och till att ge stöd i träning och undervisning. Verktyget är utvecklat av Språkbanken Tal.

Fonaydyn. Ett verktyg för att undersöka fonation samt för att utforska och validera så kallad röstmappning. Rösten betraktas här mer som en sammansatt helhet än på det traditionella reduktionistiska viset, så att kopplingar mellan till exempel grundton och styrka kan fångas. Verktyget utvecklas hos Språkbanken Tal.



VERKTYG FRÅN SPRÅKBANKEN SAM

Språkbanken arbetar aktivt för att utveckla en språkteknologisk infrastruktur. Vi skapar, harmonisera och standardisera fria språkliga resurser och tar fram verktyg för att utforska resurserna och göra dem tillgängliga. Här visar vi några exempel från **Språkbanken Sam**.

Folke. En digital arkivtjänst med ett omfattande folkminnesmaterial som belyser människors liv och arbete, vardag och fest, erfarenheter och minnen från slutet av 1800-talet och framåt. På Folke kan alla som vill bidra till att göra samlingarna mer tillgängliga genom att transkribera handskrivna texter. Folkminnesmaterialet har samlats in av Institutet för språk och folkminnen, Isof, och digitaliseras av Språkbanken Sam.

The screenshot shows the Folke digital archive interface. On the left, there are statistics: 126 avskrivena uppteckningar i april, 23 793 avskrivena uppteckningar totalt, 208 avskrivena sidor i april, 36 963 avskrivena sidor totalt, and 8 användare som har skrivit av uppteckningar i april. The main area displays a transcription of a handwritten document from 1925, with a map of Stockholm and surrounding areas on the right.

Lexin. Ett nätbaserat lexikon på nitton minoritetsspråk samt svenska. Lexin är speciellt anpassat för att användas i undervisningen i svenska som andraspråk. Lexin innehåller även bilder, animationer och dialoger i videoformat. Lexikonet har tagits fram av Institutet för språk och folkminnen, Isof, och digitaliseras av Språkbanken Sam.

Ordregn. Ordregn är en utveckling av det klassiska ordmolnet men placerar till skillnad från ordmolnet ord som har en liknande betydelse nära varandra och ger därför en överblick av vilka kategorier av ord som är viktiga i texten. Ett ordregn visar också mer exakt hur viktiga olika ord är. Ordregn har tagits fram av Språkbanken Sam.

Topics2Themes. Ett exempel på topic modelling, en språkteknologisk metod som gör det enklare att hitta, sortera och analysera stora mängder text. Verktyget har tagits fram av Språkbanken Sam.

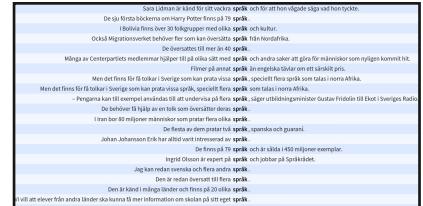


PLATTFORMAR FRÅN SPRÅKBANKEN TEXT

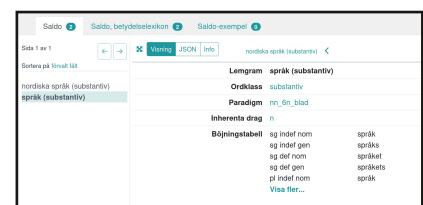
Vi utvecklar en språkteknologisk forskningsinfrastruktur genom att fritt tillgängliggöra språkdata och skapa fria plattformar för att utforska dem.



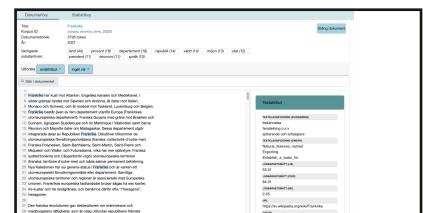
Korp är Språkbankens Textsordforskningsplattform där du kan söka i stora mängder text från bland annat dagstidningar, skönlitteratur och sociala medier. Textsamlingarna, korpusarna, är annoterade med språkliga och andra attribut.



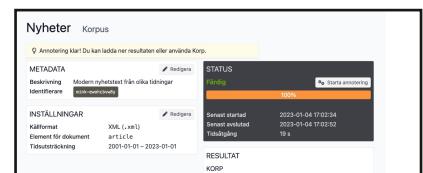
Karp är Språkbanken Texts dataredigeringsplattform där du kan redigera strukturerade data, exempelvis lexikon. Karp tillåter sökningar baserade på ordformer, böjningar samt komplexa frågor som involverar flera kriterier. I Karp finner du ett flertal lexikon, framför allt på svenska.



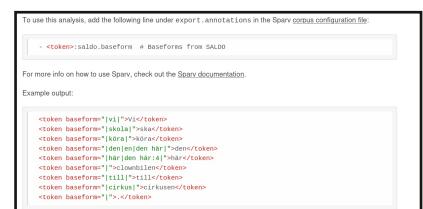
Strix är Språkbanken Texts textforskningsplattform där du kan utforska stora mängder språkteknologiskt förädlade texter på dokumentnivå. Med Strix kan du söka i en bred samling av korpusar, ta fram statistik och utforska liknande dokument genom vektorsökning.



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 Spary.



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



SPRÅKBANKENTEXT

En forskningsinfrastruktur för språkliga data
och en språkteknologisk forskningsenhet



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

2	mars	RESOURCERFUL-2025
2	mars	NoDaLiDa/Baltic-HLT 2025

sprakbanken.se

SPRÅKBANKENTEXT
En forskningsinfrastruktur för språkliga data och en språkteknologisk forskningsenhet

Språkbanken Text är en avdelning inom Språkbanken.

- Aktuellt
- Forskning
- Data
- Analyser
- Plattformar
- Frågor och svar
- Om oss
- Kontakt

Forskare: 22 varav 6 doktorander • Forskningsingenjörer: 12 • Aktiva projekt: 10 • Datamängder: 1251 • Analyser: 51 • Datapunkter: >33.7 G

Veckans ord
Ett exempel på användning av Språkbanken Texts resurser och API'er.

abborre
abborre, tsv. abborre, 1508, äldre; agborre – lida, agborre, da, aborre, no, aborre); ett speciellt nordiskt ord; den dindebur, rotens äkta vana vess i agn 1, ax, egg; (il)akrobati, o, vidare borre, av germ. "buran" (se d. o., jfr berörelse, best med v, bärach (av bars), abborre; efter den skarpändade ryggenfanen; Jr. Ikn. betyd; hos gärs, — Ordets brata.

Korp Ordet i uträddat korpus Nyckelord i kontext

Plattformar



Korp
Språkbankens ordforskningsplattform



Karp
Språkbankens delarbetningsplattform



Sparv
Språkbankens analysplattform



Strix
Språkbankens textforskningsplattform



Mink
Språkbankens dataplattform

Språkbanken Texts plattformar, verktyg och API'er →

Nyheter

[Researchdata.se – en ny portal för forskningsdata](#) →
2025-02-25

[Höjdpunkter på NoDaLiDa](#) →
2025-03-25

Datamängder

Senast uppdaterade/skapade samlingar, korpusar, lexikon, träningsdata och modeller.

[Segregationstexter: Göteborgs stad: Nämnder](#) →
2025-04-07 Korpus

Projekt

Aktuella projekt med extern finansiering.

[Språkiga näten, inom och mellan språk](#) →
Riksbankens Jubileumsfond
2022-01-01 - 2025-06-30

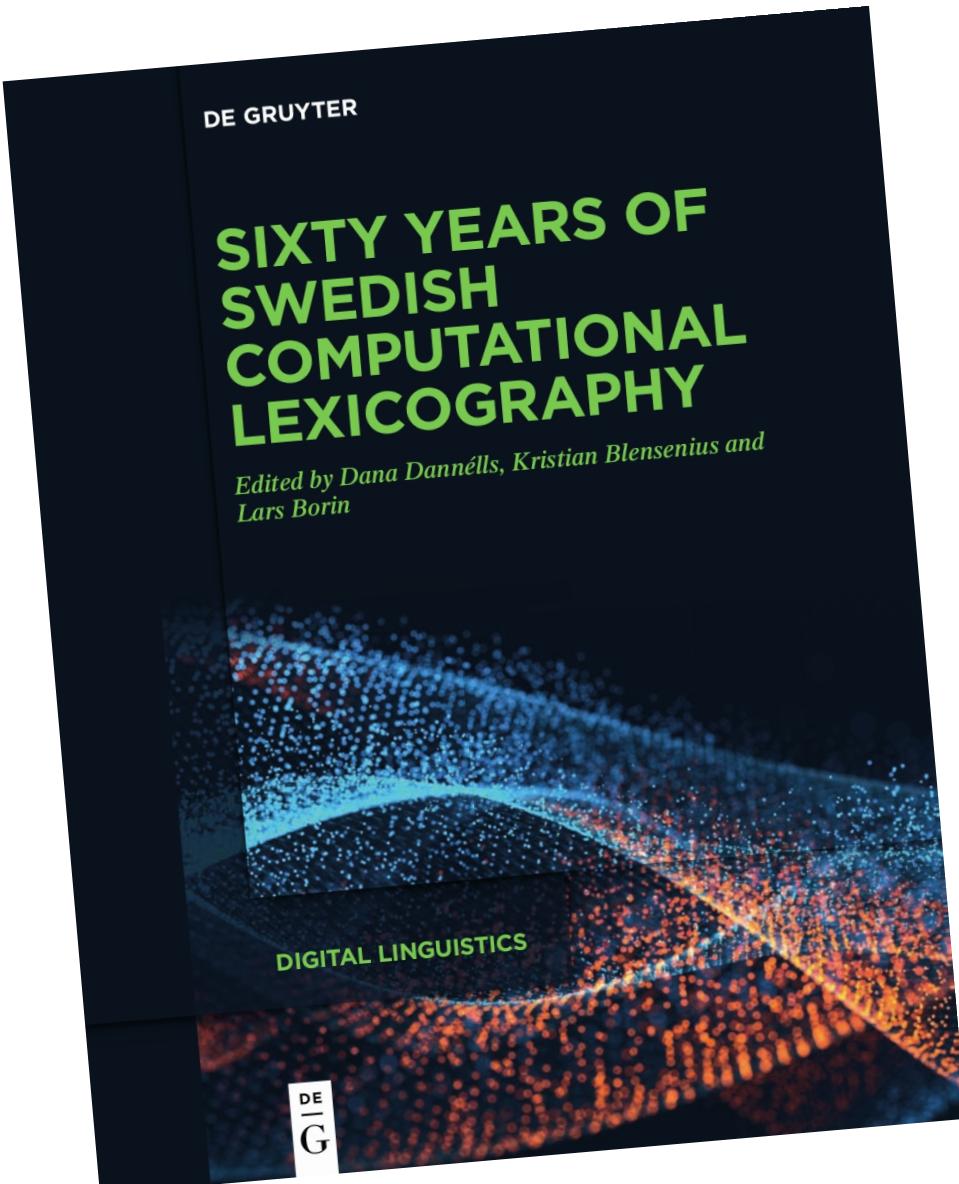
spraakbanken.gu.se

SPRÅKBANKEN
En forskningsinfrastruktur för språkliga data

SPRÅKBANKENTEXT
En forskningsinfrastruktur för språkliga data och en språkteknologisk forskningsenhet



Sixty years of Swedish computational lexicography



New and free
to download!



SPRÅKBANKENTEXT

En forskningsinfrastruktur för språkliga data
och en språkteknologisk forskningsenhet





Tack till Svenska Akademien som
har möjliggjort jubileumsfirandet
genom ett generöst bidrag.



SVENSKA
AKADEMIEN

SPRÅKBANKEN
En forskningsinfrastruktur för språkliga data