



## Language Technology research & research infrastructure

The research activities in Språkbanken focus on **creating, refining and making available linguistic resources**, with the aim of making these resources useful for research and higher education. We build and develop research infrastructure – **e-science** – for a number of disciplines, primarily **language technology and linguistics**, but also **humanities and social science** disciplines, as well as **medicine**. The work is primarily done for Swedish, but other languages are also represented.

### Research



#### VARIATION

- Big data methods in areal & genetic linguistics (South Asia, Himalayan microarea)
- Lexical resources: multilingual FrameNet, Blingbring, Swesaurus, Swedish wordnet, Constructicon, Saldo, Sentiment lexicon
- Historical data analysis
- Computational morphology: paradigm induction & paradigm prediction
- Korp, Karp & Sparv platforms – visualizing & embedding the results of the research

#### CONTENT

- Culturomics
- Word sense disambiguation/induction
- Word sense change
- Sense similarity
- Topic modeling
- Sentiment analysis
- Lexical class analysis on the document level
- Medicine: early prediction of cognitive impairment
- Korp, Sparv platforms – visualizing & embedding the results of the research

#### LEARNING

- ICALL – Intelligent Computer-Assisted Language Learning
- Readability and linguistic complexity analysis
- Learner corpus research
- Error annotation, detection & correction
- Learner text classification per level
- Learner modeling
- Exercise generation & assessment, automatic diagnostic testing
- Lärka platform – visualizing & embedding the results of the ICALL research

... and more!

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

### Research infrastructure



Tecknade symbolbilder ger ett instrument som helt ersätter det talade ordet .

Preposition	Attribut	grammatik	Efterställt Attribut	Preposition	Attribut	retorik	Efterställt Attribut
1. om	99	1. generativ	36	1. över språk	10	1. med	4196
2. med	178	2. engelsk	79	2. i sin	9	2. tom	802
3. inom	25	3. svensk	129	3. för språk	7	3. av	2659
4. av	125	4. latinsk	37	4. i skrivspråk	4	4. enligt	496
5. enligt	15	5. kontextfri	16	5. i tweet	7	5. i	4599
6. för	59	6. tysk	55	6. för mellanligg	4	6. trots	209
						1. politisk	1453
						2. från	606
						3. feministisk	288
						4. skolastisk	142
						5. hita-på	115
						6. socialdemokratisk	291
						1. används	212
						2. från	138
						3. från sida	110
						4. vara	298
						5. på	105
						6. på ut	56



- Corpus search interface
- Corpus statistics and analysis tools
- Corpus database/repository (various time frames, languages; monolingual and parallel)
- 455 corpora, 13.5 bln tokens, 900 mln sentences



- Lexicon search interface
- Corpus statistics and analysis tools
- Lexical database/repository (various time frames, languages; monolingual, bilingual, multilingual)
- Editing possibilities
- 32 lexica, 787 449 entries



- Language learning platform for Swedish
- Readability analysis
- Exercise generation
- L2 resources and L2 data
- Essay analysis
- (to develop into) L2 infrastructure SweLL



- Språkbanken's annotation tool
- Automatic linguistic annotation for multiple languages
- ...incl semantic disambiguation, compound analysis, named entity recognition, topic modeling, readability
- (coming soon) manual annotation possibilities

... and more!

<https://spraakbanken.gu.se/eng/research/infrastructure>

<https://sweclarin.se/>