



# Kursplan

för forskarkurs

**Metoder för hantering av textdata**  
**Methods for Handling of Text Data**

**7.5 Högskolepoäng**  
**7.5 ECTS credits**

**Kurskod:** LI102FU  
**Gäller från:** VT 2023  
**Institution** Institutionen för lingvistik

## Förkunskapskrav och andra villkor för tillträde till programmet

Behörig är den som är antagen till utbildning på forskarnivå vid humanistisk eller annan fakultet vid Stockholms universitet, eller motsvarande vid annat lärosäte.

## Lärandemål

För godkänt resultat på kursen ska studenten kunna:

- beskriva och tillämpa metoder för urval, insamling och förbehandling av (digitaliserat) material till textkorpusar
- beskriva och tillämpa digitala metoder för annoteringar och analys av textdata för att besvara en given uppsättning forskningsfrågor

## Innehåll

Syftet med kursen är att introducera grundläggande metodologiska färdigheter för urval, insamling, preparering, annotering och analys av textdata inom humaniora och samhällsvetenskap. Studenterna lär sig grundläggande teoretiska begrepp och får praktisk erfarenhet av verktyg och metoder, inklusive följande: Relationen mellan forskningsfrågor, datainsamling och typer av korpusar. Textbehandling i Unix-skalet. Reguljära uttryck. Kvantitativa egenskaper i språket. Frekvenser, förekomster och samförekomster. Manuell och datadriven annotering. Automatisk annotering och analys av text med färdiga verktyg.

## Obligatoriska moment

Närvaro om minst 90 % på alla föreläsningar och laborationer är obligatorisk.

Om särskilda skäl föreligger kan kan examinator efter samråd med ansvarig lärare medge den studerande befrielse från skyldigheten att delta i viss obligatorisk undervisning eller fullgörande av obligatoriska uppgift. Studenten kan då åläggas en kompensationsuppgift.

## Examinationsformer

Kursen examineras genom skriftliga laborationsrapporter.

## Arbetsform

Undervisningen ges i form av föreläsningar och laborationer.

Kursen ges på campus eller distans enligt angivelse för respektive kurstillfälle. Undervisningen bedrivs på engelska eller svenska enligt angivelse för respektive kurstillfälle.

# Syllabus

for research course

**Methods for Handling of Text Data**  
**Metoder för hantering av textdata**

**7.5 Higher Education**  
**Credits**  
**7.5 ECTS credits**

**Course code:** LI102FU  
**Valid from:** Spring 2023  
**Department:** Department of Linguistics

## Prerequisites and special admittance requirements

### Learning outcomes

In order to pass the course students are expected to be able to:

- describe and apply methods for sampling, collecting and pre-processing of (digitized) material for text corpora
- describe and apply digital methods for annotation and analysis of textual data, in order to answer a given set of research questions

### Content

The goal of the course is to introduce fundamental methodological skills for sampling, collecting, preparing, annotating and analyzing textual data in social science and the humanities. The students learn fundamental theoretical concepts and obtain practical experience of tools and methods, including the following: The relation between research questions, data collection and types of corpora. Text processing in the Unix shell. Regular expressions. Quantitative properties of language. Frequencies, occurrences and co-occurrences. Manual and data-driven annotation. Automatic annotation and analysis of text using existing tools.

### Mandatory exams

Attendance of at least 90% of all lectures and lab sessions is mandatory.

Under special circumstances the examiner, after consulting the responsible teacher, can relieve the student from the responsibility to participate in some mandatory teaching or completion of mandatory assignments. The student may then be given a compensatory assignment.

### Forms of examination

The course will be assessed through written lab reports.

### Form of instruction

The course consists of lectures and lab sessions.

The course is given on campus or online, according to the stated schedule in the course description.

The language of instruction is English, or Swedish as indicated for each course instance.