

Veranstaltungskommentar

DIGITAL: Data Analysis for Linguists

(ENG-4, ENG-4a, ENG-10, ENG11, AA7, AA9, LAA8, LAA11, IAA20, IAA23, IAA24, IAA25, ENGAGE-M01a, ENGAGE-M02)

2st. Di 16–18 ab 05.04.22

Prof. Dr. Robert Fuchs

Over the last three decades or so, linguistics as an academic discipline has come to rely on the analysis of ever-increasing amounts of data. Individual linguistic corpora, especially those involving data on English, now comprise more than 1 billion words of written data and hundreds of hours of speech. This transformation of the discipline of linguistics is part of the wider trend of the big data transformation of a range of academic fields.

Broadly speaking, more data equals better research, since more data may allow us to make claims or conclusions that are based on sounder empirical foundations. In technical terms, more data equals a greater sample size. But the allure of ever-increasing amounts of data can also bring about hidden and not-so-hidden biases, which can critically undermine the validity and representativeness of our results.

This class will introduce students to methods that allow linguists to handle large amounts of data in order to answer clearly defined research questions. Students should have a serious interest in linguistics and no deep-seated aversion to working with computers and numbers. This class will be particularly useful for students who aim to write their Bachelor's or Master's thesis in linguistics or who are working towards a PhD. All students are encouraged to contribute their own problems and data to the class, which can then be analysed jointly.

This class is open to all students from the language, literature and media division of the university (SLM I and II). While the main language of the class is English, students may make occasional contributions in German in order to facilitate communication.

Students who do not take this class as a regular part of an English studies or associated degree (BA/BEd/MEd Englisch, ENGAGE, MOTION, etc.) may submit a term paper in German. They may also choose to focus on data from a language other than English.