

Ling 780
Interdisciplinary Seminar in Linguistics

Quantitative Methods for Linguistics

Winter 2021

Overview

Vitals

Lectures: T/Th, 1:00–2:30, Remote
<https://umich.zoom.us/j/96671125114> (password: 6q541o)

Instructor: Steven Abney (abney@umich.edu)
Office hours: T/Th 2:30–3:30
<https://umich.zoom.us/j/7966461021>

Instructor: Jonathan Brennan (jobrenn@umich.edu)
Office hours: Th 9:00–10:00
<https://jobrenn.youcanbook.me>
Open “coffee hour”: Tu 9:00–10:00
<https://umich.zoom.us/j/99616733417> (password: 959731)

Description

This seminar explores quantitative and computational methods for the analysis of language data. The emphasis is on the development of practical skills through the use of existing tools, though that goes hand in hand with theoretical understanding: one cannot use a tool effectively unless one understands what it does.

Exactly what we cover will depend in part on student interests—bring your datasets and questions! A provisional schedule is given below.

Traditionally, a seminar revolves around reading papers and discussing them. In our case, the reading will be in service of learning how to do things with language data, but students will have a great deal of initiative, and class time will be used for discussion.

This is an occasional topics course, not a course with a well-established curriculum. We expect to be learning as much as teaching—that, also, is in the spirit of a traditional seminar.

Requirements

In the course of the semester, students will identify one or more linguistic datasets and questions of interest, learn about methods for processing and an-

alyzing the data to answer those questions, and write a term paper clearly communicating what was learned.

Grades will be based on the final paper, on checkpoints along the way, and on contribution to discussion.

Materials

There is a Canvas site, where all materials will be posted.

Schedule

The topics that we tentatively plan to cover are as follows. This is a lot of material, so we will not go into everything in the same amount of detail.

- **Corpus processing**

- Programming in Python. It is not assumed that you already know how to program. On the other hand, Python is not the only thing that we want to cover, so the focus will largely be on getting you started and giving you resources so that you can teach yourself.
- Language corpora. What does a corpus look like? How are characters represented? How do I get files into Python and how do I deal with corpora in different scripts? What are the different file formats? Possible topics include the LDC corpora, Universal Dependencies treebanks, Unicode, XML, tabular data, social media data.
- Corpus processing. How can I process Praat/Elan/Flex data without being dependent on Praat/Elan/Flex? How does one preprocess data to get it into a useful form? How does one count things of interest?

- **Data analysis**

- Programming in R. R is introduced in comparison to Python with a focus on matching tools to problems.
- Organizing data. How should data be organized? We discuss how different kinds of tabular data be converted into a “tidy” format appropriate for analysis with a common set of statistical and visualization tools.
- Building statistical models for linguistic data. What kinds of statistical models are used for experimental and corpus data? How are these models specified and interpreted with tools like lme4 and brm (R)?
- Classification and clustering (Scikit-Learn)
- Designing simulations to test models and check assumptions. How do we ensure that our analyses, often made up of hundreds of lines of computer code, are doing what we think they are doing? By

simulating data with real-world properties, we can check that data pre-processing, visualization, and statistical analyses are working as intended.

- **Data visualization**

- Exploratory data analysis. Visualization is a primary method for understanding patterns in data. We look at tools such as GGplot (R) and Matplotlib (python) for summarizing both data and statistical models.
- Preparing visualizations for publication. How do I make professional looking figures? What strategies are available for making complex and high dimensional data quickly understandable and accessible to a broad audience?
- Dimensionality reduction techniques

Tentative dates:

- Feb 16 – First presentation. Identify some linguistic data of interest, create one or more data tables, identify one or more questions.
- Mar 16 – Second presentation. Have a clearly defined research question, build a predictive model or run a hypothesis test. Describe the significance of the numeric results.
- Apr 20 – Third presentation. A full draft of the paper must be complete. Present your results using good visualization methods. Get feedback for the final draft.
- Apr 29 – Paper is due.

General notices

Accommodation for disability

If you think you need an accommodation for a disability, please let me know at your earliest convenience. Some aspects of this course, the assignments, the in-class activities, and the way the course is usually taught may be modified to facilitate your participation and progress. As soon as you make me aware of your needs, we can work with the Office of Services for Students with Disabilities (SSD) to help us determine appropriate academic accommodations. SSD (734-763-3000; ssd.umich.edu) typically recommends accommodation through a Verified Individualized Services and Accommodations (VISA) form. Any information you provide is private and confidential and will be treated as such.

Academic integrity

All students are expected to be aware of the College of LSA's standards of academic integrity:

<https://lsa.umich.edu/lsa/academics/academic-integrity.html>
<https://www.lib.umich.edu/academic-integrity>

Student Mental Health and Wellbeing

University of Michigan is committed to advancing the mental health and well-being of its students. If you or someone you know is feeling overwhelmed, depressed, and/or in need of support, services are available. For help, contact Counseling and Psychological Services (CAPS) at (734) 764-8312 and

<https://caps.umich.edu/>

during and after hours, on weekends and holidays, or through its counselors physically located in schools on both North and Central Campus. You may also consult University Health Service (UHS) at (734) 764-8320 and

<https://www.uhs.umich.edu/mentalhealthsvcs>,

or for alcohol or drug concerns, see

www.uhs.umich.edu/aadresources.

For a listing of other mental health resources available on and off campus, visit: <http://umich.edu/~mhealth/>.