

[Jump to Today](#)

Discovering Language Data

LING 4130/5400 | Dr. Sadaf Munshi | University of North Texas | Spring 2026 |

Course meets Tue/Thu 4:00pm-5:20pm in **Matt 102**

Instructor: Dr. Sadaf Munshi

Email: sadaf.munshi@unt.edu

OFFICE: B201E at DP.

Course Description

The course is aimed at learning the application of technical and computational tools to produce, manage, curate, analyze, and disseminate a variety of research products from text, video, and audio language data. The students will work on assignments focusing on different aspects of language documentation, archiving and dissemination of language data with a focus on the creation and management of lexical databases. We will learn to investigate the phonetic, morphological and syntactic aspects of a language and get hands on training in methods of language transcription, data analysis, and presentation formats. We will also discuss issues in orthography development and dissemination.

Prerequisites: LING 3070 or approval from the instructor.

Course Requirements and Grading

You will be evaluated on the basis of what you have learned in/outside the class in terms of your ability to use various tools in analyzing a language that you do not speak or have not learned before. You'll be working on a set of assignments using tools/applications such as Audacity, PRAAT, Fieldworks Language Explorer (FLEX), Saymore and more. The course topics, assignments and due dates will be posted on **canvas** (unt.instructure.edu).

Students will **work in teams** on a project aimed at creating a lexical corpus using various tools and methods that we will learn and implement throughout the semester. These include: tools and techniques for data collection, recording, transcription, analysis,

translation, annotation as well as data management, preservation, presentation, dissemination and archiving.

Grading will reflect the different aspects of what you learned in this course and how you contributed to the projects and assignments, your performance, and your participation in class. Course requirements, due dates (tentative) and grading criteria are as follows”

1. Lexical database: 40% (total)
 1. Part I: 10% (due February 20)
 2. Part 2: 10% (due March 6)
 3. Part 3: 10% (due April 6)
 4. Part 4: Combined database (Part 1 + 2 + 3): 10% (due April 24)

Using Excel, students will create and populate a lexical database for Part 1-3 using a template provided in class (This will be shared in canvas). For Part 4, students will combine the three parts in Excel and import them into FLEEx. Thus, Part 4 will have two components: an Excel sheet and a FLEEx database in Dictionary view. Instructions will be provided in class/canvas.

The link for our master class spreadsheet will be provided to you. Please refer to it often for any questions on your assigned word list. Each of you is to input 150 entries which you will be assigned from a larger database into your lexical database.

2. Data creation, management and metadata: 20% (total)
 - a. Recording and creation of audio data in soundproof booth (NTDP): 10% (due February 13)
 - b. Data curation, management, and metadata creation: 10% (due April 10)

Students will create sound files in a soundproof booth and analyze data in the Language Documentation Lab located at the Discovery Park. For Part 2 of this assignment, create 10-12 entries in different file formats and different genre types. Each team will work together and coordinate data creation, curation and management (Learn to use Saymore for data curation and management). The meetings will have to be scheduled and/or facilitated by the course TA in coordination with the Instructor.

3. Transcription and analysis of an assigned text: 10% (due April 17)

Using Saymore and FLEx, students will analyze a previously assigned text morphemically. The completed assignment will be submitted as a PDF file in canvas. The analyzed lexical items from this text will eventually feed into the FLEx database. Instructions for the latter part will be provided in class/canvas.

4. Review of a Webonary dictionary and/or CoRSAL language collections (20%) (due April 16)

Students must demonstrate knowledge of the various aspects of lexicography commenting on assigned dictionary/lexical databases. Issues discussed will cover the following topics: orthography issues, etymology, language contact, genetic classification, sound changes and/or contact-induced changes in the language under consideration, presentation and organization of the data, etc.

Graduate students will in addition work on reviewing assigned language collections from CoRSAL archives (UNT Digital libraries Collections).

5. Attendance, participation and classroom performance: 10%

Grading Rubric for assignments: You will be graded for the following aspects of your work:

- Quality of work (20%)
- Presentation of the data/assignments (20%)
- Completion of the assignments (20%)
- Organization of your data/assignments/files (20%), and
- Ability to demonstrate the successful use of the tools/methods introduced in this course. (20%)

Attendance Policy

We have a strict attendance policy. If you want full credit, full attendance is an absolute must; only legitimate absences (e.g. an illness, accident, or a death in family, etc.) will be excused. Unexcused absences will impact your performance and reflect negatively towards your grade.

Important: Meeting and communicating with your instructor and/or her TA at regular intervals constitute part of your participation. So, feel free to get in touch.

Course Resources

There is NO required textbook for this course.

We will be using the training materials available on the UNT CORSAL website: <https://ci.unt.edu/corsal/trainings/index.html>[Links to an external site.](#)

Additional readings and other resources will be available in Canvas.

Plagiarism Policy

For information on the University's policies regarding academic integrity and dishonesty, see the UNT Center for Student Rights and Responsibilities, <http://www.unt.edu/csrr/Links>[to an external site.](#) For more information on how to avoid plagiarism and how to use citations, see <https://guides.library.stonybrook.edu/citations/avoid-plagiarism>[Links to an external site.](#)