

# **SLS 775: Seminar in SL Research Methods**

## **Natural Language Processing in Corpus Research**

**Fall 2026**

Time W 3:00-5:50 pm

Location Moore 102

### **Instructor**

- Name: Hakyung Sung
- Contact: [sungh@hawaii.edu](mailto:sungh@hawaii.edu)
- Office Hours: M 1:00-2:00 pm (in-person, Moore 554) & By appointment

### **Course description**

Corpus linguistics is a data-driven approach to language research that investigates patterns of language use in large collections of texts. Increasingly, corpus linguistics intersects with natural language processing (NLP), as computational methods make it possible to annotate and analyze increasingly large and complex language datasets. In this course, students will engage with key concepts and methods in corpus linguistics and explore how NLP approaches can be applied to linguistic research questions. Through hands-on lab sessions, students will gain practical experience using Python to process, annotate, and analyze corpus data. Students will also be introduced to reproducible research practices, including the systematic documentation of analytical procedures and the sharing of research workflows and materials. Previous knowledge of Python is not required. However, students are expected to invest the necessary time and effort throughout the semester to develop the computational skills needed for the course.

### **Course objectives**

By the end of this course, students will be able to

- Explain assumptions, strengths, and limitations of NLP approaches in corpus research.
- Use Python to apply NLP methods to corpus data.
- Critically engage with and lead discussions of selected research papers.
- Design and conduct corpus-based research using NLP approaches.

### **Grading components** [*number of items x points per item*]

- (Paper) Presentation [ $1 \times 10$ ] 10%
- Assignments [ $4 \times 10$ ] 40%
- Project 40%
  - Project proposal [ $1 \times 3$ ] 3%
  - Midway presentation [ $1 \times 5$ ] 5%
  - Midway report [ $1 \times 7$ ] 7%
  - Final presentation [ $1 \times 10$ ] 10%
  - Final paper [ $1 \times 15$ ] 15%
- Participation 10%
  - Completing ad-hoc events that the instructor announces in class 2%
  - Active participation in Presentation 4%
  - Active participation in midway/final presentation 4%

- **Extra credit 2%**  
You may participate in the Linguistics Beyond the Classroom (LBC) program for an extra 2% credit towards your final course grade. The purpose of this program is for students to get first-hand experience of language research by participating in an on-going research project conducted here at UH. For more information, see <http://ling.lil.hawaii.edu/sites/lbc/>. You may complete either Option 1 (research participation) or Option 2 (500-word event reflection) as outlined on the LBC website. I must receive your completed questionnaire (option 1) or event response (option 2) by December 4<sup>th</sup>, 2026 for you to receive this credit.

### Grading scale

100-97%=A+; 96-93%=A; 92-90%=A-; 89-87%=B+; 86-83%=B; 82-80%=B-; 79-77%=C+; 76-73%=C; 72-70%=C-; 69-67%=D+; 66-63%=D; 62-60%=D-; < 60%=F

### Class attendance policy

This course takes place in person. While attendance is not formally tracked, 10% of the final grade is based on class participation. Participation may take various forms, including asking questions, contributing discussions, and engaging in group activities.

### Late policy

- **2-hr grading window:** All assignments (submitted via Lamakū) include an automatic two-hour grace period to accommodate technical issues. This window is applied by the system, and no action is required.
- **Late penalty:** Assignments submitted after the grace period will incur a 10% deduction per day, for up to five days (e.g., one day late results in a 10% reduction). Submissions more than five days late will receive a grade of zero.
- **Extenuating circumstances:** Whenever possible, documentation verifying extenuating circumstances may be requested to ensure equitable accommodation while respecting student privacy. If documentation cannot be provided, students are encouraged to contact the instructor as soon as possible. Extensions are generally not granted retroactively.

### Use of AIs

Students may use generative AI tools as collaborative aids, similar to consulting a human collaborator. However, using these tools to simply generate answers and submitting them as one's own work is not permitted. Students are responsible for understanding, evaluating, and, when necessary, revising any AI-generated content they use. All use of generative AI must be clearly acknowledged.

### Course materials

All course materials and resources will be shared/updated through the course website: [https://hksung.github.io/Fall26\\_SLS775/](https://hksung.github.io/Fall26_SLS775/)

## Tentative course schedule

| Week | Date  | Topic                                                                                                                                                                                                       | Readings                                                             | Assignment<br>Due Friday,<br>11:59 pm |
|------|-------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------|---------------------------------------|
| 1    | 8/26  | Intro to NLP for corpus linguistics<br><b>Lab:</b> Intro to Python 1                                                                                                                                        | Dunn (2022)<br>Gries (2026)<br>Meurers (2020)                        | Survey                                |
| 2    | 9/2   | NLP in learner corpus research<br><a href="#">Presentation 1 (Kyle &amp; Eguchi, 2024)</a><br><b>Lab:</b> Intro to Python 2                                                                                 | Granger (2026)<br>Kyle (2021)<br>Kyle & Sung (2026)                  |                                       |
| 3    | 9/9   | Corpus design and selection<br><a href="#">Presentation 2 (Hanks et al., 2024)</a><br><b>Lab:</b> A survey of corpora                                                                                       | Gries (2009)<br>Egbert et al. (2022), Ch. 3                          | Assignment 1                          |
| 4    | 9/16  | Unit of analysis, Tokenization<br><a href="#">Presentation 3 (Jarvis &amp; Hashimoto, 2021)</a><br><a href="#">Presentation 4 (Beinborn &amp; Pinter, 2023)</a><br><b>Lab:</b> Tokenization, Lemmatization  | Webb (2021)<br>Jurafsky & Martin (2024), Ch.2                        | Project<br>proposal                   |
| 5    | 9/23  | Word vectors<br><a href="#">Presentation 5 (Lu &amp; Hu, 2022)</a><br><b>Lab:</b> Word embeddings, similarity                                                                                               | Lenci (2018)<br>Jurafsky & Martin (2024), Ch. 5                      |                                       |
| 6    | 9/30  | Neural network<br><a href="#">Presentation 6 (Hemilton et al. 2016)</a><br><b>Lab:</b> POS tagging                                                                                                          | Jurafsky & Martin (2024), Ch. 6                                      | Assignment 2                          |
| 7    | 10/7  | Dependency parsing and tagging<br><a href="#">Presentation 7 (Berzak et al., 2016)</a><br><b>Lab:</b> Dependency parsing and tagging                                                                        | de Marneffe et al. (2021)<br>Zeman (2025)                            |                                       |
| 8    | 10/14 | Measuring linguistic complexity<br><a href="#">Presentation 8 (Kim &amp; Lu, 2024)</a><br><a href="#">Presentation 9 (Biber &amp; Larson, 2026)</a><br><b>Lab:</b> Python libraries for linguistic research | Bulté et al. (2025)                                                  | Assignment 3                          |
| 9    | 10/21 | Linguistic annotations<br><a href="#">Presentation 10 (Sung &amp; Kyle, 2024)</a><br><b>Lab:</b> Treebank constructions                                                                                     | <a href="#">Adding linguistic annotation</a><br>Leech (2004)         |                                       |
| 10   | 10/28 | Midway presentation                                                                                                                                                                                         |                                                                      | Midway<br>report                      |
| 11   | 11/4  | Semantic role labeling<br><a href="#">Presentation 11 (Ettinger et al., 2023)</a><br><b>Lab:</b> Training vs. prompting LLMs                                                                                | Jurafsky & Martin (2024), Ch. 21                                     |                                       |
| 12   | 11/11 | Veterans Day ( <i>No Class</i> )                                                                                                                                                                            |                                                                      |                                       |
| 13   | 11/18 | Sentiment/discourse analysis<br><a href="#">Presentation 12 (Taboada, 2016)</a><br><a href="#">Presentation 13 (Zeldes et al., 2025)</a><br><b>Lab:</b> Related packages/tools                              | Jurafsky & Martin (2024), Ch. 22<br>Jurafsky & Martin (2024), Ch. 24 |                                       |
| 14   | 11/25 | Multilingual/Multimodal corpora<br><a href="#">Presentation 14 (Kuperman et al., 2025)</a><br><a href="#">Presentation 15 (Tian et al., 2026)</a><br><b>Lab:</b> Intro to GitHub                            | Baltrušaitis et al. (2018)                                           | Assignment 4                          |
| 15   | 12/2  | Reproducibility<br><a href="#">Presentation 16 (Imperial et al., 2025)</a><br><b>Lab:</b> Intro to HuggingFace                                                                                              | In'nami et al. (2022)<br>McManus (2021)                              |                                       |
| 16   | 12/9  | Final presentation                                                                                                                                                                                          | <b>*Final Paper</b><br>12/14, 11:59 pm                               |                                       |

## UH policies and resources

### Attendance Policy

Students who are enrolled in this course but never attend will be flagged by the course instructor for non-participation before the last day to add/drop for a 100% tuition refund. Flagged students will be administratively dropped by the Office of the Registrar.

Any changes to a student's enrollment status may affect financial aid eligibility and can result in the return of some or all federal student financial aid.

### Statement on Disability: KOKUA Program

If you have a disability and related access needs, please contact the KOKUA Program (Office for Students with Disabilities). Please know that I will work with you and KOKUA to meet your access needs based on disability documentation. KOKUA's services are confidential and offered free of charge.

**Phone:** (808) 956-7511

**Email:** [KOKUA@hawaii.edu](mailto:KOKUA@hawaii.edu)

**Location:** Room 013, Queen Lili'uokalani Center for Student Services

### Academic Integrity and Ethical Behavior

Cheating, plagiarism, or other forms of academic dishonesty are not permitted in this course and are prohibited under the System-wide Student Conduct Code (EP 7.208). Examples include fabrication, falsification, cheating, plagiarism, and the use of improper materials.

Any incident of suspected academic dishonesty will be reported to the Office of Student Conduct for review and possible adjudication. Additionally, the instructor may take action regarding the grade for the deliverable or course as they see fit.

**Resource:** Office of Student Conduct

### Department of Public Safety

**Emergency:** (808) 956-6911

**Non-Emergency:** (808) 956-8211

**Website:** <http://manoa.hawaii.edu/dps/>

### UH System Basic Needs

Basic needs include food and housing, childcare, mental health, financial resources, and transportation, among others. Student basic-needs security is critical for ensuring strong academic performance, persistence and graduation, and overall student well-being.

If you or someone you know is experiencing basic-needs insecurity, please see the following resource:

**Resource:** UH System Basic Needs

### Student Success Resources

The Office of the Vice Provost for Student Success provides student support services to promote success both inside and outside the classroom.

Resources include learning assistance, academic advising, career resources, counseling, family and relationship support, affinity and involvement resources, health and wellness services, opportunities for leadership growth, and community engagement.

**Phone:** (808) 956-3290

**Website:** <https://manoa.hawaii.edu/studentssuccess/departments/>

### Office of Title IX

UHM is committed to providing a learning, working, and living environment that promotes personal integrity, civility, and mutual respect and is free of all forms of sex discrimination and gender-based violence, including sexual assault, sexual harassment, gender-based harassment, domestic violence, dating violence, and stalking.

If you or someone you know experiences any of these, UHM has staff and resources on campus to provide support and assistance. Staff can also direct you to resources in the community.

**Phone:** (808) 956-2299  
**Email:** [t9uhm@hawaii.edu](mailto:t9uhm@hawaii.edu)  
**Website:** <https://manoa.hawaii.edu/titleix/>

---

*Confidential Resources*

If you wish to remain anonymous, speak with someone confidentially, or receive information and support in a confidential setting, contact one of the following resources:

**Counseling & Student Development Center**

**Phone:** (808) 956-7927  
**Website:** [manoa.hawaii.edu/counseling/](http://manoa.hawaii.edu/counseling/)

**Respondent Support**

**Phone:** (808) 956-4392  
**Email:** [PAUrs@hawaii.edu](mailto:PAUrs@hawaii.edu)

**Student Parents at Mānoa (SP@M)**

**Phone:** (808) 956-8059  
**Website:** [manoa.hawaii.edu/studentparents/](http://manoa.hawaii.edu/studentparents/)

**UH Confidential Advocacy**

**Email:** [advocate@hawaii.edu](mailto:advocate@hawaii.edu)

**University Health Services Mānoa**

**Phone:** (808) 956-8965  
**Website:** [hawaii.edu/shs/](http://hawaii.edu/shs/)

Additional information about confidential resources is available at:

<http://www.manoa.hawaii.edu/titleix/resources.html#confidential>

---

*Reporting an Incident*

If you wish to report an incident of sex discrimination or gender-based violence—including sexual assault, sexual harassment, gender-based harassment, domestic violence, dating violence, or stalking—as well as receive information and support, contact:

**Director and Title IX Coordinator**

Hawai'i Hall 124  
2500 Campus Road  
Honolulu, HI 96822  
**Phone:** (808) 956-2299  
**Email:** [t9uhm@hawaii.edu](mailto:t9uhm@hawaii.edu)

---

As a member of the University faculty, I am required to immediately report any incident of sex discrimination or gender-based violence to the campus Title IX Coordinator. Although the Title IX Coordinator and I cannot guarantee confidentiality, you will still have options regarding how your case will be handled. My goal is to make sure that you are aware of the range of options available to you and have access to the resources and support you need.