CS/Ling 581 Computational Linguistics
Spring 2025
Schedule # 6637 / 5265

SDSU

COURSE INFORMATION

Time:	TTh 11:00am–12:15pm	Email:	rmalouf@sdsu.edu
Place:	HH-130	Office hours:	TTh 1:00pm–2:00pm
Mode:	In person	Real office:	SHW 201
Instructor:	Prof. Rob Malouf	Zoom office:	see Canvas

This course will serve as an introduction to the field of computational linguistics, which includes aspects of speech recognition, natural language processing, information retrieval, and information extraction. The course begins with an introduction to finite-state automata and some basic natural language applications; this is extended to finite-state transducers with applications in text processing. Other topics covered: n-gram language models, classifiers (logistic regression and multilevel perceptrons), sentiment analysis, dependency parsing, distributional semantics, and 'deep' learning.

ESSENTIAL STUDENT INFORMATION

For essential information about student academic success, please see the SDSU Student Academic Success Handbook.

- SDSU provides disability-related accommodations via Student Disability Services (sds@sdsu.edu |https://sds.sdsu.edu/). Please allow 10-14 business days for this process.
- Class rosters are provided to the instructor with the student's legal name. Please let me know if you would prefer an alternate name and/or gender pronoun.

COURSE MATERIALS

The textbook for this course is:

Jurafsky, Daniel and Martin, James H. 2023. *Speech and Language Processing*. Third edition draft (January 12, 2025). web.stanford.edu/~jurafsky/slp3/

This book is only available as a downloadable PDF. There are hard copy versions of the 1st and 2nd editions for sale on, e.g., Amazon, but those are the wrong books. **Do not buy any books for this class!**

All other course information, additional readings, assignments, slides, etc. will be available on Canvas.

COURSE DESIGN: MAJOR ASSIGNMENTS AND ASSESSMENTS

The grade for the course will be based on readings and homework assignments (60%), an in class midterm exam (20%), and an in class final exam (20%). Late homework will be accepted with a grade penalty. Programming assignments should be completed in Python. Students with no programming background will probably find this course extremely challenging.

COURSE SCHEDULE

Proposed course outline:

Week	Topic	Reading
1	Introduction	
2—3	Finite state machines	Chapter 2
4—5	Language models	Chapter 3
6—7	Classifiers	Chapter 4, 5
8	Midterm	
9	Word meanings	Chapter 6
10–11	Sequence labeling	Chapter 8
12–13	Deep learning	Chapter 7, 9
14	Dependency parsing	Chapter 18
15	Future prospects	

AI SYLLABUS STATEMENT

Students should not use generative AI applications in this course except as approved by the instructor. Any use of generative AI outside of instructor-approved guidelines constitutes misuse. Misuse of generative AI is a violation of the course policy on academic honesty and will be reported to the Center for Student Rights and Responsibilities.

STUDENT LEARNING OUTCOMES

Upon successful completion, students will have the knowledge and skills to:

- Implement and evaluate finite-state automata and transducers for basic natural language processing tasks such as text normalization and pattern matching.
- Design and train n-gram language models to predict word sequences and evaluate their performance using standard metrics like perplexity.
- Develop supervised machine learning models, including logistic regression and multilayer perceptrons, to solve text classification problems such as sentiment analysis.
- Analyze syntactic relationships in text using dependency parsing techniques and evaluate parsing accuracy.
- Apply distributional semantic methods to represent word meanings and measure semantic similarity between words and phrases.
- Compare and contrast traditional machine learning approaches with deep learning architectures for various natural language processing tasks.