

**CENG 463**  
**INTRODUCTION TO NATURAL LANGUAGE PROCESSING**

**2023-2024 Fall**

Instructor: Ayşenur Birtürk

TA: Atakan Garipler

**Description:** This is a foundational course on Natural Language Processing. You will learn fundamental algorithms and mathematical models for processing natural language, and how these can be used to solve practical problems.

**Objective:** To form a basis for Natural Language Processing and Computational Linguistics. To present applicable linguistic theory by turning theories into practical techniques with emphasis on problems for which there are widely accepted solutions. To familiarize the student with the latest advances in Natural Language Technology and applications like information extraction, machine translation, natural language interfaces, etc.

**Text Book:**

- Speech and Language Processing, Jurafsky and Martin, Prentice-Hall, 2008, 2.edition.
- 3.edition draft available at:

<https://web.stanford.edu/~jurafsky/slp3>

**Reference Books:**

- Foundations of Statistical Natural Language Processing, Manning and Schütze MIT Press. Cambridge, MA: May 1999.
- Introduction to Information Retrieval, Manning, Raghavan, and Schütze, Cambridge University Press, 2008.
- Natural Language Understanding, J. Allen, 2.ed, Benjamin-Cummings, 1995. Prolog and Natural Language Analysis, F.C.N. Pereira, S.M. Shieber, CSLI, 1987.
- Natural Language Processing with Python—Analyzing Text with the Natural Language Toolkit, Bird, Klein, and Loper, O'Reilly, 2009.
- Natural Language Processing for Prolog Programmers, Michael A. Covington, Prentice-Hall, 1994.

## **Syllabus** (subject to change)

### **PART A** (2<sup>nd</sup> edition)

- Introduction
- Basic Text Processing
- Minimum Edit Distance
- Language Modeling
- Spelling Correction
- Text Classification
- Sentiment Analysis
- Maximum Entropy Classifiers
- Information Extraction and Named Entity Recognition
- Relation Extraction
- Advanced Maximum Entropy Models
- POS Tagging, Sequence Labeling for Parts of Speech and Named Entities
- Parsing Introduction
- Probabilistic Parsing
- Lexicalized Parsing
- Dependency Parsing
- Information Retrieval
- Ranked Information Retrieval
- Semantics
- Question Answering
- Summarization

### **PART B** (new in 3<sup>rd</sup> edition)

- Naive Bayes and Sentiment Classification
- Logistic Regression
- Vector Semantics and Embeddings
- Neural Networks and Neural Language Models
- RNNs and LSTMs
- Transformers and Pretrained Language Models
- Fine-tuning and Masked Language Models
- Lexicons for Sentiment, Affect, and Connotation

### **PART C**

- Some NLP Applications (Machine Translation, Question Answering and Information Retrieval, Chatbots and Dialogue Systems, Automatic Speech Recognition and Text-to-Speech)
- Turkish Morphology

**Lecture Slides:**

<https://web.stanford.edu/~jurafsky/NLPCourseraSlides.html>

<https://web.stanford.edu/~jurafsky/slp3>

Plus additional slides

**Grading (tentative):**

Midterm Exam: 30%

Final Exam: 40%

Take Home Assignments (Programming/Written HWs): 30%

Bonus: Attendance %5

**Exam Conduct (subject to change)**

- Midterm and Final exams will be closed book and face to face (F2F).

**Policy**

- The take-home exam assignments that are designated as individual assignments must be completed individually. Copying from others, either from other students or off the internet is strictly forbidden and will surely constitute grounds for failure.
- In case of cheating in one of the take-home exams, all of the take-home exams will be graded as 0 and furthermore the case will be passed to the Student Discipline Committee.

**Communication**

All announcements will be made in ODTU Class.