

Topics in Language Processing, CS 5319 (crosslisted as Special Topics in Computer Science, CS 4390.)

Introduction to Speech and Language Processing

Speech and language processing has recently seen tremendous advances, and the core technologies of speech recognition and speech synthesis are now mature, with readily-usable APIs available for developers and well-designed systems used by millions of people every day. At the same time, the range of viable applications is currently very limited, due to a large number of challenging open problems.

This class will provide a survey and sampling of the techniques and issues in speech and language processing.

Topics:

1. Basic properties of language, along with feature-based, vector-space and other representations of each level (acoustics, phonetics, prosody, morphology, syntax, meaning, pragmatics), with attention to similarities and differences across languages, genres, and speakers.
2. Standard and for-purpose techniques for speech and language processing components, including approaches to knowledge discovery, model building, and model training.
3. The components, engineering, and properties of large-scale systems for search, question-answering, conversational agents, and machine translation.
4. Other applications including language identification, emotion and speaker-state detection, language-proficiency assessment, various information retrieval problems, turn-end detection, response selection.
5. System design, integration, and testing (in linkage with the final project)

Class Format: lectures, reading-based discussions, in-class project time, etc.

Approximate Grading: tests 40%, assignments 25%, project 25%, participation 10%

Textbook: Speech and Language Processing, Daniel Jurafsky and James H. Martin, 2nd edition, Prentice-Hall, 2008.

Prerequisites: Linear Algebra (Math 3323 or equivalent), Probability and Statistics (Stat 3320 or equivalent), strong programming and systems-integration skills (CS 3331 or equivalent). Knowledge of Machine Learning techniques and of introductory linguistics concepts are helpful but not required.

Graduate students from other departments are welcome even with some incomplete prerequisites; see the instructor for permission.