What is Natural Language Processing (NLP)?

The process of building computational models for understanding natural language.

**INPUT**: natural language text.

**OUTPUT**: representation of the text’s meaning.

Sometimes called *natural language understanding* (NLU) or *computational linguistics*.

Course Goals

- To study algorithms and methods for building computational models of natural language understanding, such as: parsing techniques, semantic representations, discourse analysis, and statistical and corpus-based methods for text analysis and knowledge acquisition.

- To study issues involved in understanding natural languages, such as cognitive and linguistic phenomena.

- To learn about applications that can benefit from NLP.

By the end of this course, everyone will have the skills to build NLP tools and an understanding of (and appreciation for!) the issues involved in analyzing natural language.

Related Disciplines

**Computer Science**: artificial intelligence, machine learning

**Linguistics**: computational linguistics

**Psychology**: cognitive psychology, psycholinguistics

**Statistics**: probabilistic methods, information theory

Applications for NLP

- Information Retrieval / Search Engines: Document Categorization, Routing, Filtering
- Question Answering, Summarization
- Fact Extraction
- Sentiment Analysis
- Machine Translation
- Speech Recognition, Spoken Language Understanding
- Database Query Interfaces
- Intelligent Tutoring Systems
Levels of Analysis and Knowledge

**Morphology:** how words are constructed; prefixes & suffixes

**Syntax:** structural relationships between words

**Semantics:** meanings of words, phrases, and expressions

**Discourse:** relationships across different sentences or thoughts; contextual effects

**Pragmatic:** the purpose of a statement; how we use language to communicate

**World Knowledge:** facts about the world; common sense

Morphology

- kick, kicks, kicked, kicking
- sit, sits, sat, sitting
- murder, murders

But it’s not as simple as naively adding and deleting endings...

- gorge, gorgeous
- glass, glasses
- arm, army

Morphology Humor

Syntax: Part-of-Speech Tagging

*The boy threw a ball to the brown dog.*

*The armed man saw a bear.*
Syntax: structural ambiguity

Time flies like an arrow.

Metaphor:
Time/NOUN flies/VERB like/PREP an/ART arrow/NOUN

New Fly Species:
Time/NOUN flies/NOUN like/VERB an/ART arrow/NOUN

Stopwatch Imperative:
Time/VERB flies/NOUN like/PREP an/ART arrow/NOUN

Syntax: structural ambiguity
(attachment)

• I saw the Grand Canyon flying to New York.

• I watered the plant with yellow leaves.

• I saw the man on the hill with the telescope.

But syntax doesn’t tell us much about meaning…

• Colorless green ideas sleep furiously. [Chomsky]

• fire … match … arson … hotel

• “dog collar” vs. “flea collar”

• plastic cat food can cover
Semantics: Lexical Ambiguity

• I walked to the bank ... of the river. to get money.

• The bug in the room ... was planted by spies. flew out the window.

• I work for John Hancock ... and he is a good boss. which is a good company.

Word Sense Ambiguity

Syntactic or Semantic Ambiguity?
(These are real headlines!)

Discourse: coreference

A Short Story

Squad Helps Dog Bite Victim
Safety Experts Say School Bus Passengers Should Be Belted
Drunk Gets Nine Months in Violin Case
Lung Cancer in Women Mushrooms
Iraqi Head Seeks Arms
Eye Drops Off Shelf
Teacher Strikes Idle Kids
Stud Tires Out
Farmer Bill Dies in House
Reagan Wins on Budget, But More Lies Ahead

President John F. Kennedy was assassinated.
The president was shot yesterday.
Relatives said that John was a good father.
JFK was the youngest president in history.
His family will bury him tomorrow.
Friends of the Massachusetts native will hold a candlelight service in Mr. Kennedy’s home town.
Discourse: coreference

A Short Story

President John F. Kennedy was assassinated. The president was shot yesterday. Relatives said that John was a good father. JFK was the youngest president in history. His family will bury him tomorrow. Friends of the Massachusetts native will hold a candlelight service in Mr. Kennedy's home town.

Pragmatics

Rules of Conversation
- Can you tell me what time it is?
- I would like the salt.

Speech Acts
- I bet you $50 that the Jazz will win tonight.
- Will you marry me?

Pragmatics Humor

World Knowledge

John went to the diner. He ordered a steak. He left a tip and went home.

John wanted to commit suicide. He got a rope.
John got up one morning and discovered his power was out. Unable to shave, he called his next door neighbor and asked if he could come over to borrow the bathroom. But everyone on the street was out. So John called FG&E and drove to work hoping no one would see him before he found a bathroom with hot water. Unfortunately, he ran into his boss on the elevator. He explained his predicament, but did not feel reassured by Mr. Carver’s silence. John stumbled through the rest of the week half-expecting to find a pink slip in his mailbox.

What’s the State-of-the-Art?

- statistical and machine learning methods are dominant
- training with huge text collections
- emphasis on empirical methods
- real-world texts, in many forms
- quantitative evaluations
- automated knowledge acquisition
- but deep semantics only in limited domains
- some areas starting to break through commercially