

# Introduction to Machine Learning Applications

Spring 2023

Natural Language Processing (NLP) overview

**Minor Gordon**

[gordom6@rpi.edu](mailto:gordom6@rpi.edu)



**Rensselaer**

# Language is challenging

## Verb ambiguity:

- I looked straight at an eclipse of the sun.
- I looked after my little sister.
- I looked at the catalogue and saw the dress that I wanted to order.

## Noun ambiguity:

- River bank / financial bank
- Mountain bike / tandem bike
- Wine glass / sea glass

# Subfields of linguistics

Subfield	Description
Phonetics	The study of the sounds of human language
Phonology	The study of sound systems in human language
Morphology	The study of formation and internal structure of words
Syntax	The study of formation and internal structure of sentences
Semantics	The study of the meaning of sentences
Pragmatics	The study of the way sentences with their semantic meanings are used for particular communicative goals.

Table from: Linguistic Fundamentals for Natural Language Processing, June 2013, Emily M. Bender

# Different Components

- Language models
- Part Of Speech tagging
- Parsing
- Named Entity Recognition
- Co-reference
- Automatic Speech Recognition
- Text To Speech
- Semantic roles

# Applications

- Text classification
  - Spam detection
  - Sentiment analysis
- Question answering systems
- Machine translation
- Spoken dialogue systems
  - Siri, Alexa

# Types of text

- Subjective text:
  - Blogs, online product reviews, movie reviews, social media posts
- Objective text:
  - Wikipedia articles, NYT news articles, WSJ articles

# Sentiment analysis

Positive	Negative
My experience so far has been fantastic!	Your support team is useless
Package is nice.	Absolutely horrible!
This product is second to none.	This is better than nothing.
I will absolutely recommend it.	Yeah sure. So smooth!

# Sentiment analysis: practical issues

- Cleaning the text
- Tokenization
- POS tagging
- Semantics
- Named entity recognition
- Coreference resolution



# Sentiment analysis: approaches

- Unsupervised classification
  - Using POS tagging, bigrams and calculating Semantic Orientation (SO)
- Supervised classification
  - Naïve Bayes
  - Maximum entropy
  - Support vector machines