

Natural Language Processing

DSECL ZG565

Dr. Ashish Kulkarni

BITS Pilani

March 27, 2021

Session Content

1. Course Overview
2. Natural Language Processing - Introduction

Objective of the Course

- To learn the fundamental concepts and techniques of natural language processing (NLP)
- To learn computational properties of natural languages and the commonly used algorithms for processing linguistic information
- To apply NLP techniques to state-of-the-art applications
- To learn implementation of NLP algorithms and techniques

Books

T1	Speech and Language processing: An introduction to Natural Language Processing, Computational Linguistics and speech Recognition by Daniel Jurafsky and James H. Martin[3rd edition]
T2	Natural language understanding[2nd edition] by James Allen
R1	Handbook of Natural Language Processing, Second Edition—Nitin Indurkha, Fred J. Damerau, Fred J. Damerau
R2	Natural Language Processing with Python by Steven Bird, Ewan Klein, Edward Loper

Evaluation Plan

Name	Type	Duration	Weight
Assignment-1	Take home		15%
Quiz	Online		5%
Mid-Semester Exam (after 7 sessions)	Closed Book	1.5 hrs	30%
Comprehensive Exam	Open Book	2.5 hrs	50%

Language

One of the fundamental aspects of human behavior

- in written form it serves as a **long-term record of knowledge**
- in spoken form it serves as our primary **means of coordinating** our day-to-day behaviors with others

Natural Language Processing (NLP)

Research of language comprehension and language production

The ultimate goal is to be able to specify models that approach human performance in the linguistic tasks of reading, writing, hearing, and speaking.

The Major Disciplines Studying Language

Discipline	Typical Problems	Tools
Linguists	How do words form phrases and sentences? What constrains the possible meanings for a sentence?	Intuitions about well-formedness and meaning; mathematical models of structure (for example, formal language theory, model theoretic semantics)
Psycholinguists	How do people identify the structure of sentences? How are word meanings identified? When does understanding take place?	Experimental techniques based on measuring human performance; statistical analysis of observations
Philosophers	What is meaning and how do words and sentences acquire it? How do words identify objects in the world?	Natural language argumentation using intuition about counter examples; mathematical models (e.g. logic and model theory)
Computational Linguists	How is the structure of sentences identified? How can knowledge and reasoning be modeled? How can language be used to accomplish certain tasks?	Algorithms, Data Structures, formal models of representation and reasoning, AI techniques

Motivations for Computational Models

Scientific Motivation

Obtain a better understanding of how language works

Practical Motivation

Revolutionize the way computers are used

Applications of Natural Language Understanding

Text-based

- Information retrieval
- Information extraction
- Machine translation
- Document summarizing

Dialogue-based

- Question-answering
- Automated customer service
- Tutoring systems
- Voice commands-based systems
- General cooperative problem solving systems (e.g. plan and schedule freight shipments)

Evaluating Language Understanding Systems

Black box evaluation: evaluate system performance without looking inside to see how it works

Glass box evaluation: identify various sub-components of a system and then evaluate each one with appropriate tests

Men are all alike.
IN WHAT WAY
They're always bugging us about something or other.
CAN YOU THINK OF A SPECIFIC EXAMPLE
Well my boyfriend made me come here.
YOUR BOYFRIEND MADE YOU COME HERE
He says I'm depressed much of the time.
I AM SORRY TO HEAR YOU ARE DEPRESSED
..
WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU?
My father
YOUR FATHER
You are like my father in some ways.
WHAT RESEMBLANCE DO YOU SEE
You are not very aggressive but I think you don't want me to notice that.
WHAT MAKES YOU THINK I AM NOT AGGRESSIVE
You don't argue with me.
WHY DO YOU THINK I DON'T ARGUE WITH YOU
You are afraid of me.
DOES IT PLEASE YOU TO BELIEVE I'M AFRAID OF YOU
My father is afraid of everybody.
WHAT ELSE COMES TO MIND WHEN YOU THINK OF YOUR FATHER
Bullies.
DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR BOYFRIEND MADE YOU COME HERE

Figure 26.4 Sample ELIZA dialogue from Weizenbaum (1966).

Levels of Language Understanding

- **Phonetic and phonological knowledge:** how words are related to the sounds that realize them
- **Morphological knowledge:** how words are constructed from more basic meaning units called morphemes
- **Syntactic knowledge:** how words can be put together to form correct sentences
- **Semantic knowledge:** study of context-independent meaning
- **Pragmatic knowledge:** how sentences are used in different situations
- **Discourse knowledge:** how the immediately preceding sentences affect the interpretation of the next sentence
- **World knowledge:** general knowledge about the structure of the world

1. Language is one of the fundamental aspects of human behavior and is a crucial component of our lives.
2. Green frogs have large noses.
3. Green ideas have large noses.
4. Large have green ideas nose.

Representation and Understanding

Words have multiple senses (*e.g.* “cook”). This inherent ambiguity poses challenges in building computational models for language understanding.

This calls for representation using a more precise language, such that,

- the representation must be precise and unambiguous
- the representation should capture the intuitive structure of the natural language sentences that it represents

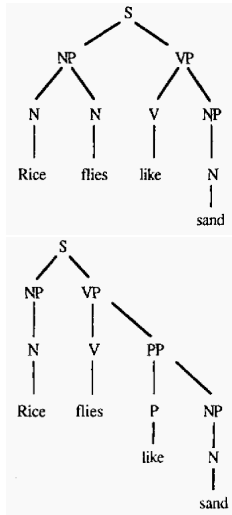
The tools to do this come from mathematical and logical formalism

Language Representation

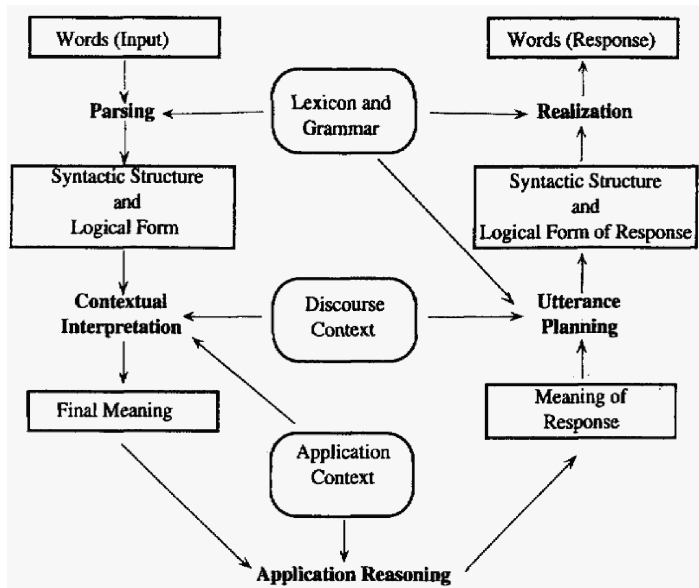
Syntax: Representing sentence structure - the way that words in a sentence are related to each other. Most syntactic representations of language are based on the notion of **context-free grammars**.

The logical form: Representation of the context-independent meaning of a sentence. *E.g.* “the catch” can have different meanings depending on whether the speaker is talking about a baseball game or a fishing expedition.

General **knowledge representation** (KR): This is the language in which all the specific knowledge based on the application is represented. Typically uses **first-order predicate calculus** (FOPC).



The Organization of Natural Language Understanding Systems



The End