

# **Natural Language Processing**

---

## **Chapter 1 : Introduction**

# Natural Language Processing

- **Textbook:** Speech And Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. By Daniel Jurafsky and James H. Martin, Prentice-Hall, 2000. Details about this book and its resources can be found at: <http://www.cs.colorado.edu/~martin/slp.html>
- **Course Objectives:**
  1. Gain an appreciation of the complexity of natural language.
  2. Survey some applications of natural language processing.
  3. Understand basic processes and representations used in syntax, semantics, and other components of natural language processing.
  4. Practice individual investigations in chosen topics

# Introduction to Natural Language Processing (NLP)

- A natural language is a human spoken language, such as English Arabic, French, ... etc.
- One of the aims of Artificial Intelligence (AI) is to build machines that can "understand" commands in natural language, written or spoken.
- A computer that can do this requires very powerful hardware and sophisticated software.
- At the present time, this is at the early stages of development.

# Introduction to Natural Language Processing (NLP)

- It is not an easy task to teach a person or computer a natural language.
- The main problems are syntax (the rules governing the way in which words are arranged), and understanding context to determine the meaning of a word.
- To interpret even simple phrases requires a vast amount of knowledge.
- The basic goal of Natural language Processing is to enable a person to communicate with a computer in a language that they use in their everyday life.

# Natural Language and Computer Language

- Natural language are those that we use for communicating with each other, eg. Arabic, English, French, Japanese, etc.
- Natural language are expressive and easy for us to use.
- Computer languages are those that we use for controlling the operations of a computer, eg. Prolog, C, C++, C#, Java, Python,..., etc.
- Computer languages are easy for a computer to understand, but they are not expressive.

# Applications for Natural Language Processing

- Machine Translation
- Fact Extraction.
- Information Retrieval / Search Engines:
- Retrieval, Categorization, Filtering, Summarization
- Question Answering Systems
- Speech Recognition & Spoken Language Understanding
- Intelligent Tutoring Systems
- Database Query Interfaces

# Major NLP Accomplishments

- Chomsky (1957) Syntactic Structures
- Weizenbaum (1966), ELIZA
- Woods (1967), Procedural semantics
- Thorne et al. and Woods (1968-70), ATNs
- Winograd (1970), Shrdlu
- Colby, Weber & Hilf, 1971; Colby, 1975, PARRY
- Wilks (1972), Preference semantics
- Woods et al. (1972), LSNLIS / Lunar
- Charniak (1972), Frames and demons
- Wilks (1973), Stanford machine translation project
- Montague (1973) IL semantics (Montague Grammar) in PTQ
- Grosz (1977), Focus in task-oriented dialogues

# Major NLP Accomplishments

- Marcus (1977), Deterministic parsing
- Cohen, Phil (1979), Planning speech acts
- Allen (1980), Understanding speech acts
- McDonald (1980), MUMBLE
- Heim/Kamp (1981) Discourse Representation Theory
- McKeown (1982), TEXT
- Appelt (1982), KAMP (Integration of Functional Grammar with Discourse Plans)
- Shieber (1984) Non context freeness of NL syntax proven
- Pollack (1986), Plan inference
- Mann & Thompson (1987), Rhetorical Structure Theory



# Five Processing Stages in a NLP System

- Phonological Analysis
- Morphological Analysis
- Syntactic Analysis
- Semantic Analysis
- Pragmatic Analysis

# Five Processing Stages in a NLP System

## (1) Phonological Analysis

- Phonetics: deals with the physical building blocks of a language sound system.  
eg. sounds of 'k', 't' and 'e' in 'kite'
- Phonology: organization of speech sounds within a language.
- eg. (1) different 'k' sounds in 'kite' vs 'coat'
- (2) different 't' and 'p' sounds in 'top' vs 'pot'

## Five Processing Stages in a NLP System

### (2) Morphological Analysis

- Morphology is the structure of words.
- It is concerned with **inflection**. (i.e *The various forms of the same basic word. eg. run-ran, dog-dogs, etc*)
- It is also concerned with **derivation** of new words from existing ones, eg. lighthouse (formed from light & house).
- In NLP, words are also known as lexicon items and a set of words form a lexicon.

## Five Processing Stages in a NLP System

### (2) Morphological Analysis

- Any NL analysis system needs a lexicon {a module that tells what words there are and what properties they have}.
- Simplest model is a full form dictionary that lists every word explicitly.
- Simply expanding the dictionary fails to take advantages of the regularities.
- No dictionary contains all the words one is likely to encounter in real input.
- Languages with highly productive morphology (e.g. Finnish, where a verb can have many thousands of forms.)
- Noun combination.

## Five Processing Stages in a NLP System

### (3) Syntactic Analysis

- Syntactic analysis is concerned with the construction of sentences.
- Syntactic structure indicates how the words are related to each other.
- Syntax tree is assigned by a grammar and a lexicon.
- Lexicon indicates syntactic category of words.
- Grammar (typically Context Free Grammar) specifies legitimate concatenations of constituents.

## Five Processing Stages in a NLP System

### (4) Semantic Analysis

- Semantic analysis is concerned with the meaning of the language.
- This stage uses the meanings of the word to extend and perhaps disambiguate the result returned by the syntactic parse.
- The first step in any semantic processing system is to look up the individual words in a dictionary (or lexicon) and extract their meanings.
- Unfortunately, many words have several meanings, for example, the word 'diamond' might have the following set of meanings:
  - (1) a geometrical shape with four equal sides.
  - (2) a baseball field
  - (3) an extremely hard and valuable gemstone
- To select the correct meaning for the word 'diamond' in the sentence *Joan saw Susan's diamond shimmering from across the room.*

## Five Processing Stages in a NLP System

### (4) Semantic Analysis

- It is necessary to know that neither geometrical shapes nor baseball fields shimmer, whereas gemstones do (process of elimination).
- The process of determining the correct meaning of an individual word is called word sense disambiguation or lexical disambiguation.
- It is done by associating, with each word in the lexicon, information about the contexts in which each of the word's senses may appear.
- Each of the words in a sentence can serve as part of the context in which the meanings of the other words must be determined.

## Five Processing Stages in a NLP System

### (5) Pragmatic Analysis

- This is an additional stage of analysis concerned with the pragmatic use of the language.
- This is important in the understanding of texts and dialogues.
- There are many important relationships that may hold between phrases and parts of their discourse context, as outlined below.

Identical entities. Consider :

- Bill had a red balloon.
- John wanted it.
- The word 'it' should be identified as referring to the red balloon. References such as this are call *anaphoric* or *anaphora*.