Unit-8: Natural Language **Processing**

Computer Engineering Department

Outline

- Introduction
- Syntactic Processing
- Semantic Analysis
- Discourse And Pragmatic Processing
- Applications of NLP
- Spell Checking

Introduction

▶ Natural language processing (NLP) is a field of Artificial Intelligence in which computers analyze, understand, and derive meaning from human language.

▶ The field focuses on communication between computers and humans in natural language and is about making computers understand and generate human language.

▶ Natural Language Processing (NLP) refers to communicating with an intelligent systems using a natural language such as English.

▶ Processing of Natural Language is required when you want an intelligent system like robot to perform as per your instructions, when you want to hear decision from a dialogue based clinical expert system, etc.

▶ By utilizing NLP, developers can organize and structure knowledge to perform tasks such as automatic summarization, translation, named entity recognition, relationship extraction, sentiment analysis, speech recognition, topic segmentation, etc.

Introduction

▶ There are the following two components of NLP -

1. Natural Language Understanding (NLU): Natural Language Understanding (NLU) helps the machine to understand and analyze human language by extracting the metadata from content such as concepts, entities, keywords, emotion, relations, and semantic roles.
   - NLU mainly used in Business applications to understand the customer's problem in both spoken and written language.

2. Natural Language Generation (NLG): Natural Language Generation (NLG) acts as a translator that converts the computerized data into natural language representation. It mainly involves Text planning, Sentence planning, and Text Realization.

▶ Natural Language Processing (NLP) problem can be divided into two tasks:

1. Processing written text, using lexical, syntactic and semantic knowledge of the language as well as the required real world information.
2. Processing spoken language, using all the information needed above plus additional knowledge about phonology as well as enough added information to handle the further ambiguities that arise in speech.

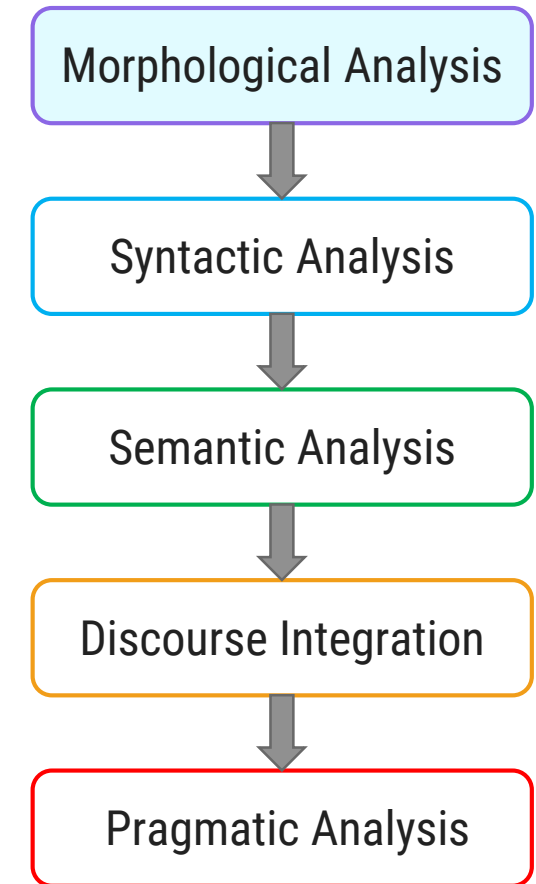# Steps in Natural Language Processing

1. Morphological Analysis:
   ➥ The morphological level of linguistic processing deals with the study of word structures and word formation, focusing on the analysis of the individual components of words.
   ➥ Lexicon of a language means the collection of words and phrases in a language. It involves identifying and analyzing the structure of words.
   ➥ Morphological analysis is dividing the whole chunk of text into paragraphs, sentences, and words.
   ➥ Suppose there is a sentences, "I want to print Bill's .init file."
   ➥ Morphological analysis must do the following things:
     ▪ Pull apart the word "Bill's" into proper noun "Bill" and the possessive suffix "'s"
     ▪ Recognize the sequence ".init" as a file extension that is functioning as an adjective in the sentence.
   ➥ This process will usually assign syntactic categories to all the words in the sentence.

Morphological Analysis

⬇

Syntactic Analysis

⬇

Semantic Analysis

⬇

Discourse Integration

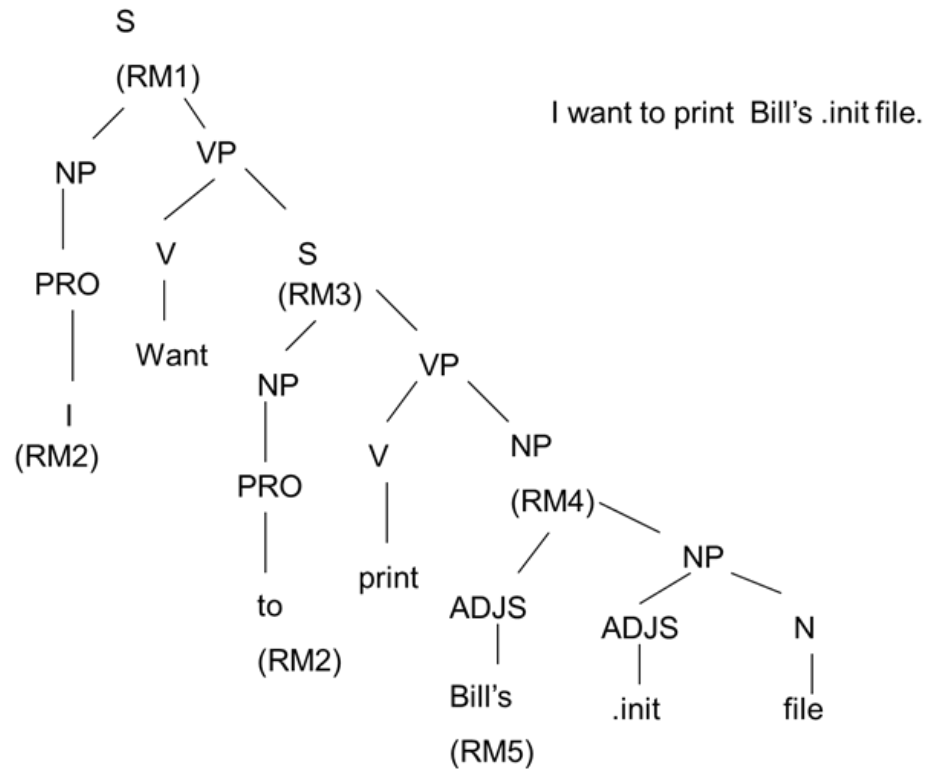⬇

Pragmatic Analysis
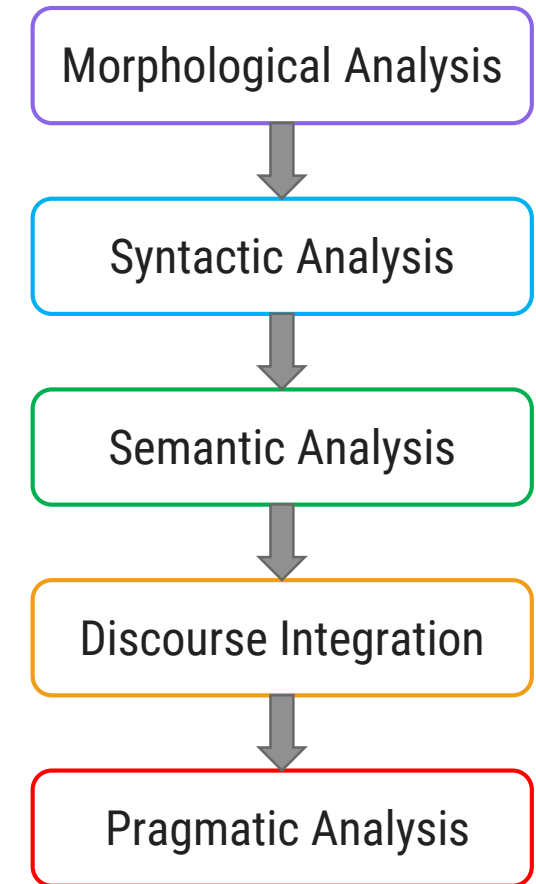
Steps in Natural Language Processing

2. Syntactic Analysis:
   ➥ Syntactic analysis must exploit the results of morphological analysis to build a structural description of the sentence.
   ➥ The goal of this process, called parsing, is to convert the flat list of words that forms the sentence into a structure that defines the units that are represented by that flat list.
   ➥ The important thing here is that a flat sentence has been converted into a hierarchical structure and that the structure corresponds to meaning units when semantic analysis is performed.
   ➥ Reference markers (set of entities) are shown in the parenthesis in the parse tree.
   ➥ Each one corresponds to some entity that has been mentioned in the sentence.
   ➥ These reference markers are useful later since they provide a place in which to accumulate information about the entities as we get it.

Morphological Analysis

↓

Syntactic Analysis

↓

Semantic Analysis

↓

Discourse Integration

↓

Pragmatic Analysis

Steps in Natural Language Processing

2. Syntactic Analysis:

Steps in Natural Language Processing

3. Semantic Analysis:
   ➥ Semantic analysis must do two important things:
      a. It must map individual words into appropriate objects in the knowledge base or database.
      b. It must create the correct structures to correspond to the way the meanings of the individual words combine with each other.

   ➥ The semantic level of linguistic processing deals with the determination of what a sentence really means by relating syntactic features and disambiguating words with multiple definitions to the given context.
   ➥ This level entails the appropriate interpretation of the meaning of sentences, rather than the analysis at the level of individual words or phrases.
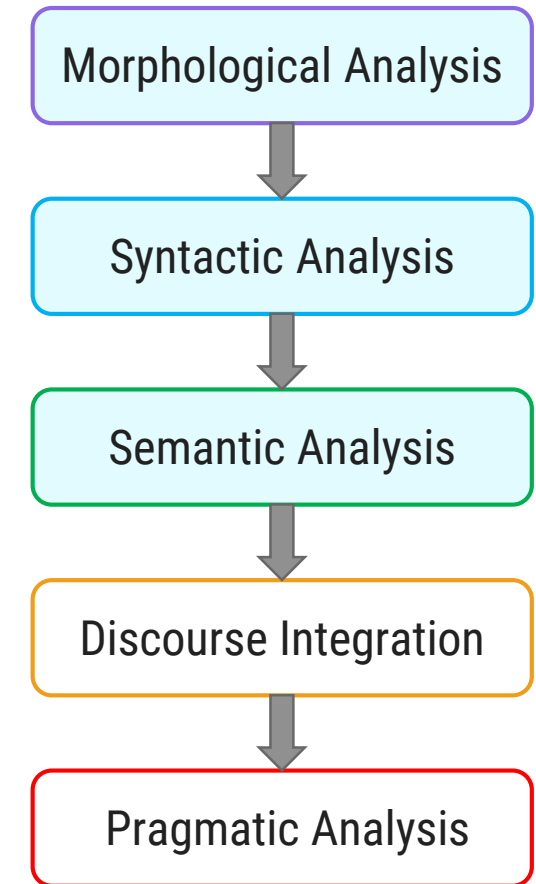
Morphological Analysis

↓

Syntactic Analysis

↓

Semantic Analysis

↓

Discourse Integration

↓

Pragmatic Analysis
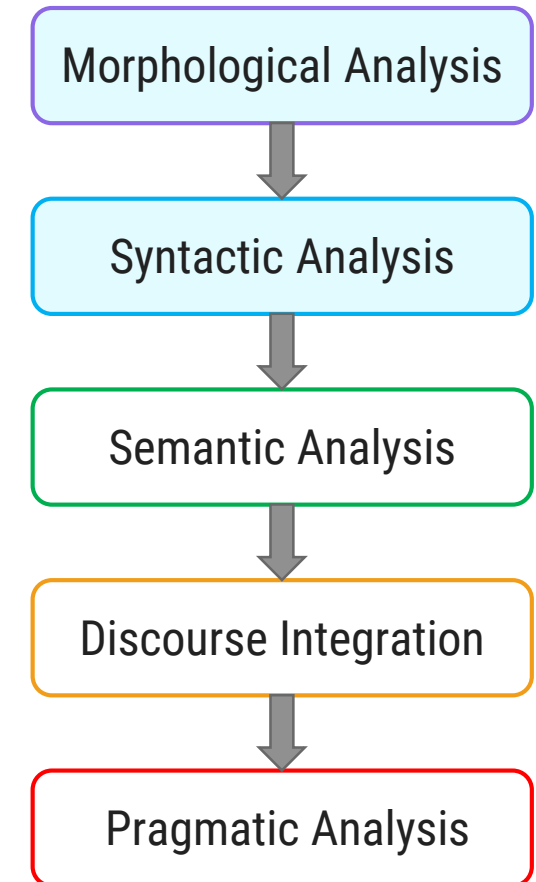
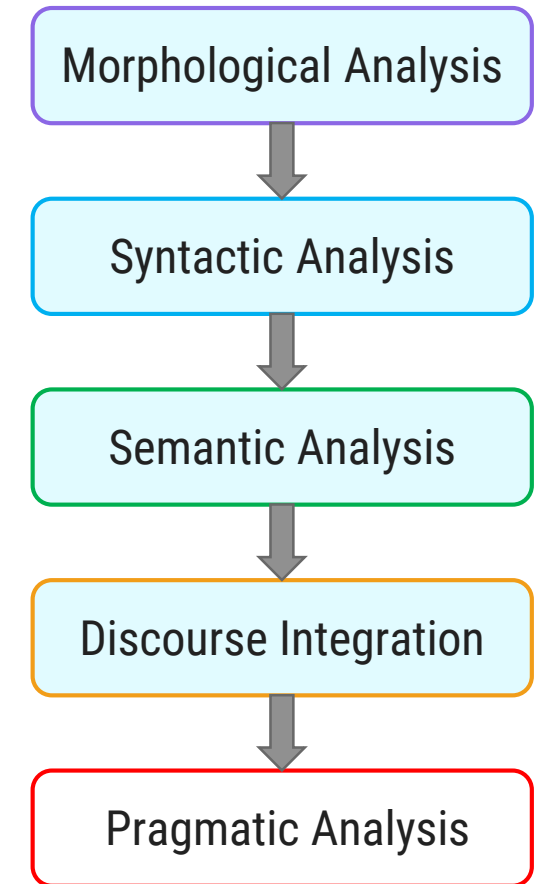# Steps in Natural Language Processing

4. Discourse Integration:
   ➥ The discourse level of linguistic processing deals with the analysis of structure and meaning of text beyond a single sentence, making connections between words and sentences.
   ➥ At this level, Anaphora Resolution is also achieved by identifying the entity referenced by an anaphor (most commonly in the form of, but not limited to, a pronoun).
   ➥ An example is shown below.
     ▪ "I voted for Obama because he was most aligned with my values," she said.
   ➥ With the capability to recognize and resolve anaphora relationships, document and query representations are improved.
   ➥ Structured documents also benefit from the analysis at the discourse level since sections can be broken down into (1) title, (2) abstract, (3) introduction, (4) body, (5) results, (6) analysis, (7) conclusion, and (8) references.

Morphological Analysis

↓

Syntactic Analysis

↓

Semantic Analysis

↓

Discourse Integration

↓

Pragmatic Analysis

Steps in Natural Language Processing

5. Pragmatic Analysis:
   ➥ The final step toward effective understanding is to decide what to do as a result.
   ➥ One possible thing to do is to record what was said as a fact and be done with it.
   ➥ For some sentences, whose intended effect is clearly declarative, this is the precisely correct thing to do.
   ➥ But for other sentences, including this one, the intended effect is different.
   ➥ We can discover this intended effect by applying a set of rules that characterize cooperative dialogues.
   ➥ The final step in pragmatic processing is to translate, from the knowledge based representation to a command to be executed by the system.
   ➥ The pragmatic level of linguistic processing deals with the use of real-world knowledge and understanding of how this impacts the meaning of what is being communicated.
   ➥ By analyzing the contextual dimension of the documents and queries, a more detailed representation is derived.

Morphological Analysis

Syntactic Analysis

Semantic Analysis

Discourse Integration

Pragmatic Analysis

Applications of NLP

▶ Sentiment Analysis :

→ Sentiment Analysis is also known as **opinion mining**. It is used on the web to analyze the attitude, behavior, and emotional state of the sender.

→ This application is implemented through a combination of NLP (Natural Language Processing) and statistics by assigning the values to the text (positive, negative, or natural), identify the mood of the context (happy, sad, angry, etc.)

→ Beyond determining simple polarity, sentiment analysis understands sentiment in context to help better understand what's behind an expressed opinion, which can be extremely relevant in understanding and driving purchasing decisions.

▶ Text Classification :

→ Text clarification is the process of categorizing the text into a group of words.

→ By using NLP, text classification can automatically analyze text and then assign a set of predefined tags or categories based on its context.

→ For e.g., Spam Detection is used to detect unwanted e-mails getting to a user's inbox.

Applications of NLP

▶ **Chat bots & Virtual Assistants :**
- ➥ Implementing the Chat bot is one of the important applications of NLP. It is used by many companies to provide the customer's chat services.
- ➥ A virtual assistant is a software that uses speech recognition, natural language understanding, and natural language processing to understand the verbal commands of a user and perform actions accordingly.

▶ **Information extraction :**
- ➥ Information extraction is one of the most important applications of NLP.
- ➥ It is used for extracting essential information from unstructured or semi-structured machine-readable documents.

▶ **Machine Translation :**
- ➥ Machine translation is used to translate text or speech from one natural language to another natural language while keeping the meaning intact. For e.g., Google Translate can easily convert text from one language to another language. These tools are helping numerous people and businesses in breaking the language barrier and becoming successful.

Applications of NLP

▶ Text Summarization :
  ➥ Summarization is the task of condensing a piece of text to a shorter version, reducing the size of the initial text while at the same time preserving key informational elements and the meaning of content.
  ➥ There are important applications for text summarization in various NLP related tasks such as text classification, question answering, legal texts summarization, news summarization, and headline generation.

▶ Auto-Correct :
  ➥ Microsoft Corporation provides word processor software like MS-word, PowerPoint for the spelling correction.
  ➥ Tools like Grammarly provide so many features in helping a person write better content. It is one of the most widely used applications of NLP that helps professionals in all job domains create better content.

▶ Speech Recognition :
  ➥ From smart home devices and appliances that take instructions, and can be switched on and off remotely, digital assistants that can set reminders, schedule meetings, play songs, to search engines that respond with relevant search results to user queries, speech recognition has become an indispensable part of our lives.

Spell Checking

▶ Spell Check is a process of detecting and sometimes providing suggestions for incorrectly spelled words in a text.

▶ In computing, Spell Checker is an application program that flags words in a document that may not be spelled correctly.

▶ Spell Checker may be stand-alone capable of operating on a block a text such as word-processor, electronic dictionary.

▶ A basic spell checker carries out the following processes:

➥ It scans the text and extracts the words contained in it.

➥ It then compares each word with a known list of correctly spelled words (i.e. a dictionary).

➥ An additional step is a language-dependent algorithm for handling morphology.

Spell Checking

▶ Spelling errors can be divided as:

➡ Non-word errors : These are the most common type of errors. You either miss a few keystrokes or let your fingers hurtle a bit longer. These are those error words that cannot be found in the dictionary. This words are complex to provide the suggestion, so this might not be suggested.

➡ Real-word errors : Sometimes instead of creating a non-word, you end up creating a real word, but one you didn't intend to do so. E.g., typing flower when you meant flour. These are those error words that are acceptable words in the dictionary.

➡ Cognitive Errors : The previous two types of errors result not from ignorance of a word or its correct spelling. Cognitive errors can occur due to those factors. The words piece and peace are homophones (sound the same). So you are not sure which one is which. Sometimes your damn sure about your spellings.

➡ Short forms/Slang/Lingo : These are possibly not even spelling errors. you are trying hard to fit in everything within a text message or a tweet.

▶ Error Detection

➡ Dictionary Lookup Technique : In this, Dictionary lookup technique is used which checks every word of input text for its presence in dictionary. If that word present in the dictionary, then it is a correct word. Otherwise it is put into the list of error words.

Thank You!