



Computational Linguistic and Natural Language Processing

Course Outcomes

- . Understand the basic concepts of text classification and lexical analysis.
- . Understanding the basic concepts of Natural Language Processing.
- . Understanding the basic concepts of information retrieval

Credit subject (2)

2 Quiz

2 Assignments

2 Test

Computational Linguistic

- “Human knowledge is expressed in language. So computational linguistics is very important.”
- Computational linguistics (CL) is the application of computer science to the analysis, synthesis and comprehension of written and spoken language.
- Computational linguistics is used in instant machine translation, speech recognition (SR) systems, text-to-speech (TTS) synthesizers, interactive voice response (IVR) systems, search engines, text editors and language instruction materials. The interdisciplinary field of study requires expertise in machine learning (ML), deep learning (DL), artificial intelligence (AI), cognitive computing and neuroscience.

Text classification

- “Text classification is a machine learning technique that assigns a set of predefined categories to open-ended text”
- Text classifiers can be used to organize, structure, and categorize any kind of text.
- Text classification is one of the fundamental tasks in natural language processing with broad applications such as sentiment analysis, topic labeling, spam detection, and intent detection.

What is Natural Language Processing (NLP)

- The process of computer analysis of input provided in a human language (natural language), and conversion of this input into a useful form of representation.
- Natural language processing (NLP) is the ability of a computer program to understand human language as it is spoken and written -- referred to as natural language. It is a component of artificial intelligence.
- The field of NLP is primarily concerned with getting computers to perform useful and interesting tasks with human languages.
- The field of NLP is secondarily concerned with helping us come to a better understanding of human language.
- NLP is subfield of Artificial Intelligence. It is used to identify meaning of words and sentences in natural language. It also helps computers to understand human languages in better way.

Forms of Natural Language

The input/output of a NLP system can be:

- **written text**
- **speech**

We will mostly concerned with written text (not speech).

To process written text, we need:

- **lexical, syntactic, semantic knowledge about the language**
- **discourse information, real world knowledge**

To process spoken language, we need everything required to process written text, plus challenges of speech recognition and speech synthesis.

NLP Applications

Search Engine

Opinion Mining

Sentiment analysis

Chatbot

Document/text classification

Machine Translation

Recommendation

Social Network Analysis

Spelling and grammar check

Components of NLP

Natural Language Understanding

- Mapping the given input in the natural language into a useful representation. The meaning of spoken/typed language is found.
- Different level of analysis required:
 - syntactic analysis,*
 - semantic analysis*

Natural Language Generation

- Producing output in the natural language from some internal representation.
- Different level of synthesis required:
 - deep planning* (what to say),
 - syntactic generation*

NL Understanding is much harder than NL Generation.

Go beyond the keyword matching



Identify the **structure** and **meaning** of **words**, **sentences**, **texts** and **conversations**

Deep understanding of **broad** language

NLP is all around us

Phases of NLP

Tokenization – Create tokens from sentence

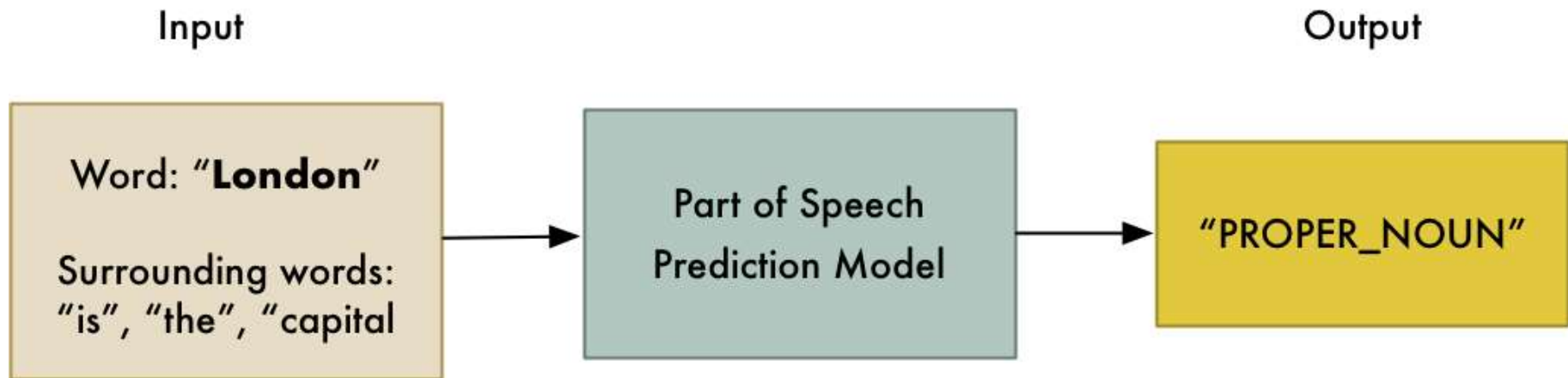
Stemming – Normalize words into base form

Lemmatization – Output different words into single word. E.g. went, gone into go, better, best----- good

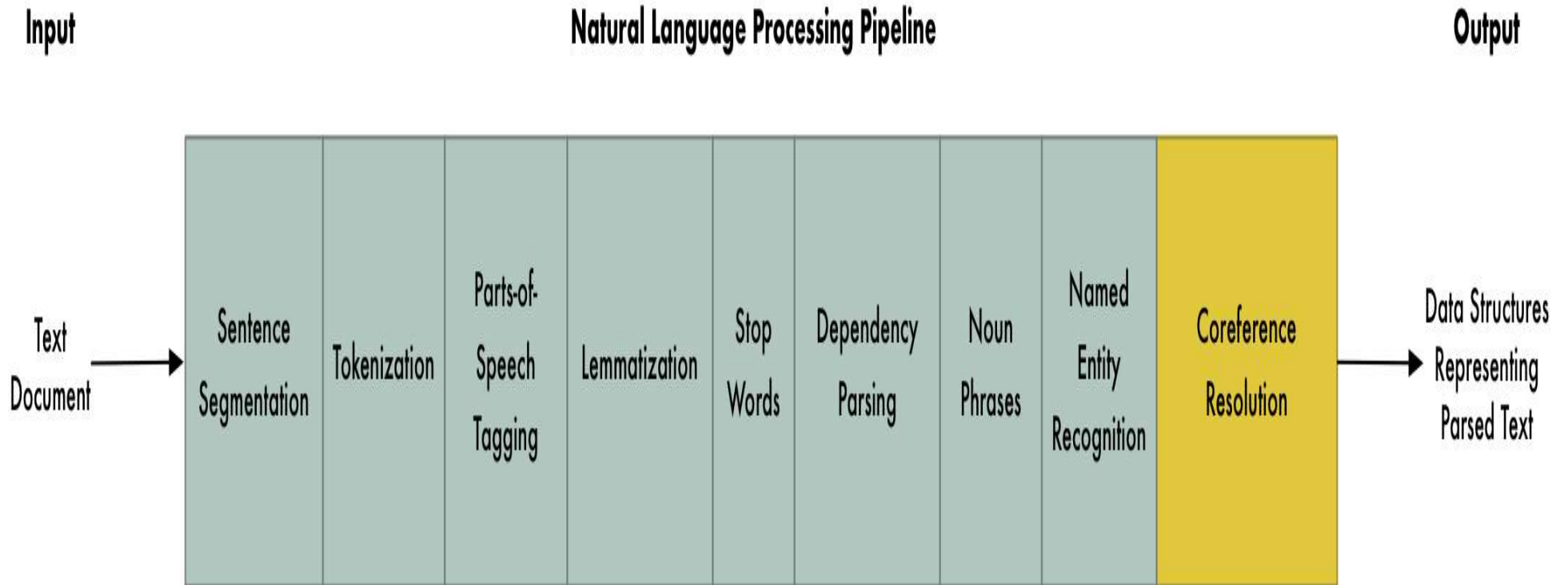
POS Tags- Dividing Sentence into Noun, Verb etc.

Named Entity Recognition – dividing sentence in different entities.

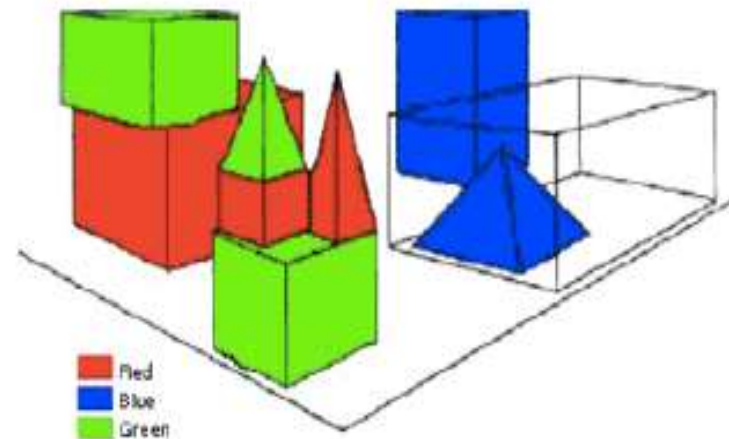
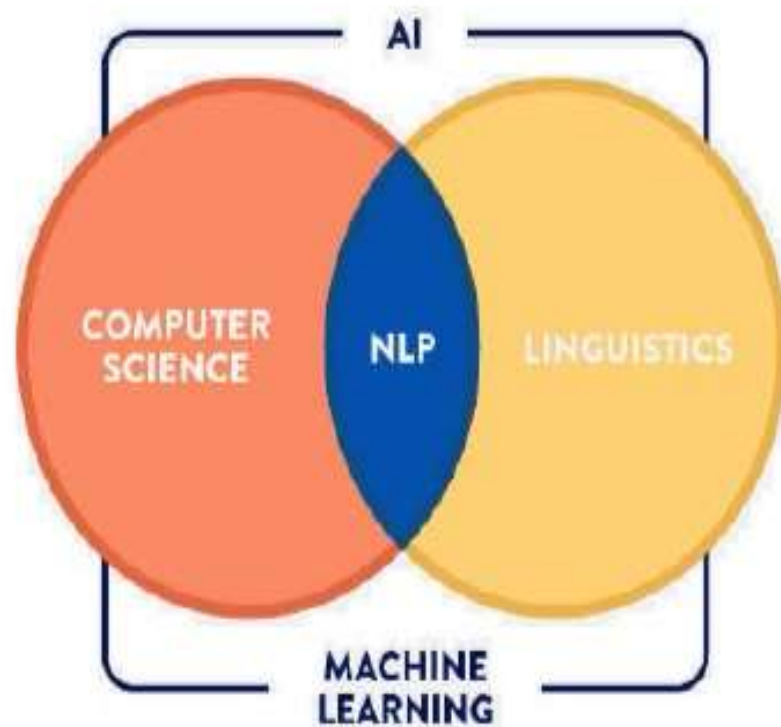
POS Tagger



NLP Pipeline



Approaches to natural language processing



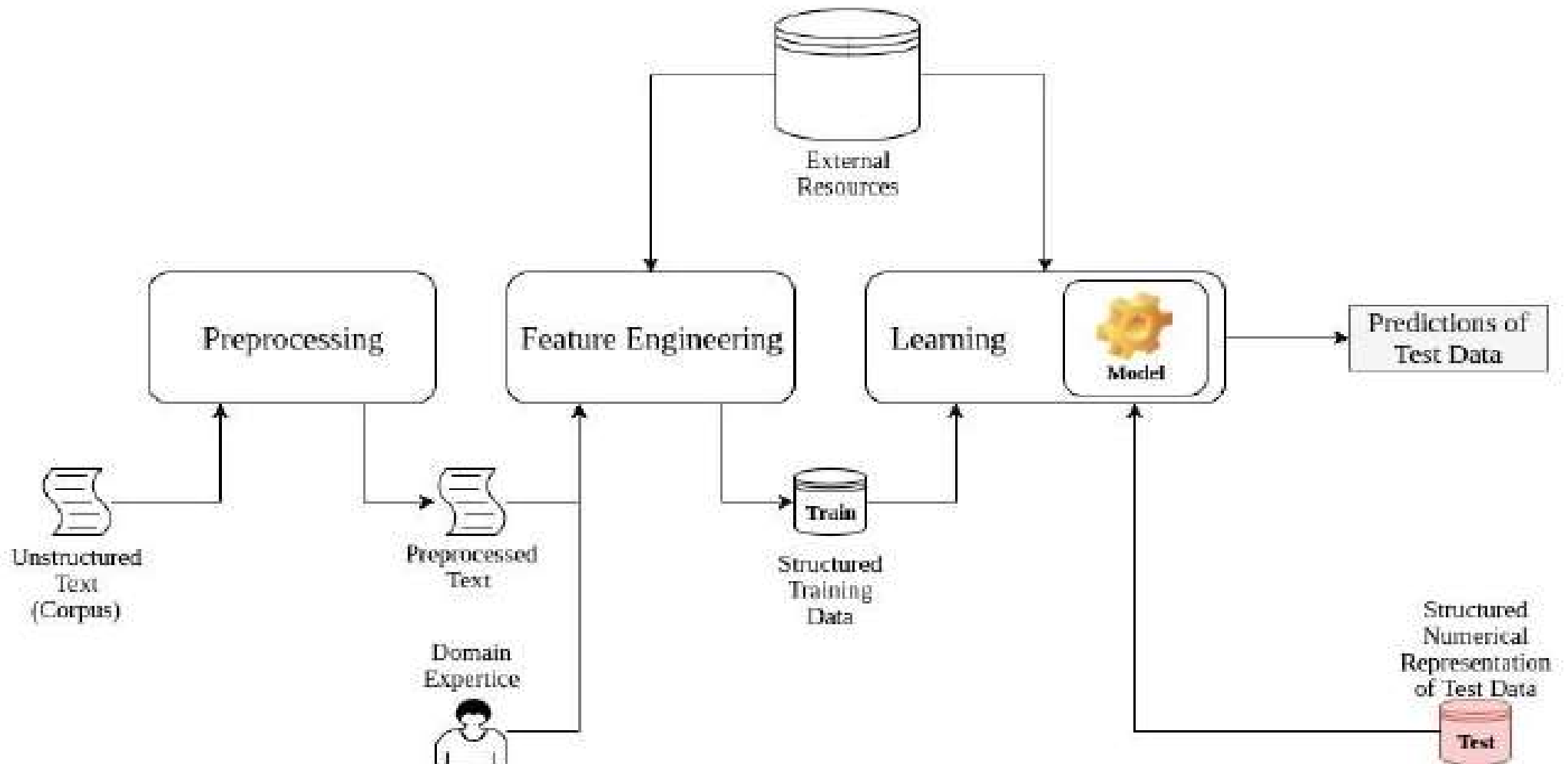
Person: Pick up a big red block.

Computer: OK.

Person: Grasp the pyramid.

Computer: I don't understand which pyramid you mean.

Approaches to natural language processing



Understanding linguistics

Basic ideology starts by making the computer identify the different stages of learning a natural language.

A sentence, traditionally the following are the different stages on how a sentence would be analyzed to gain deeper insights.

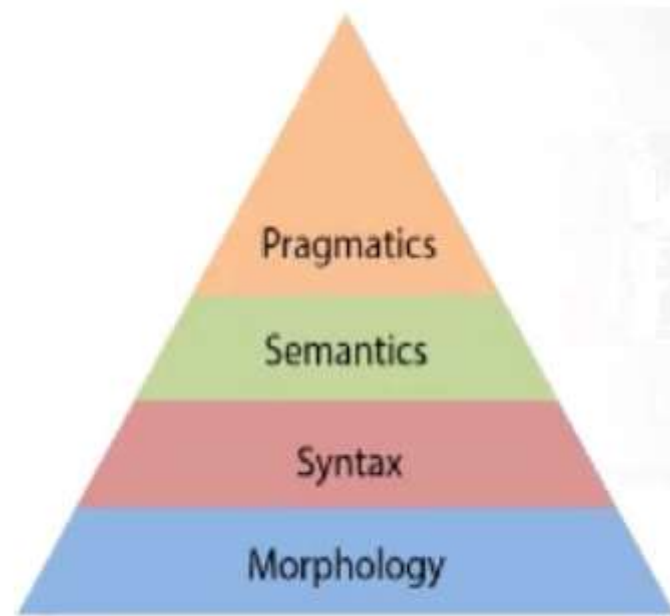


Figure: Major levels in linguistics analysis

Level 1: Morphology

Morphology is a field of linguistic focused on the study of the forms and formation of words in a language. Words in a language consist of one element or elements of the meaning which are called morphemes.

A morpheme is defined as the minimal meaningful unit of a language.

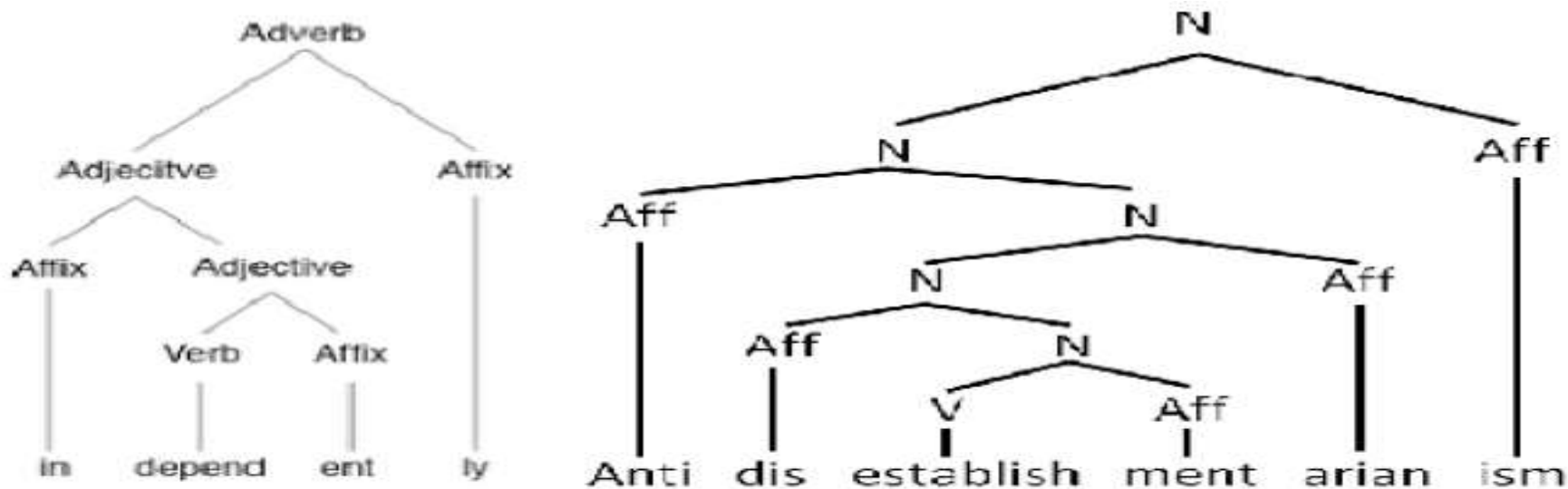


Figure: Morphology makes up the basic constructs

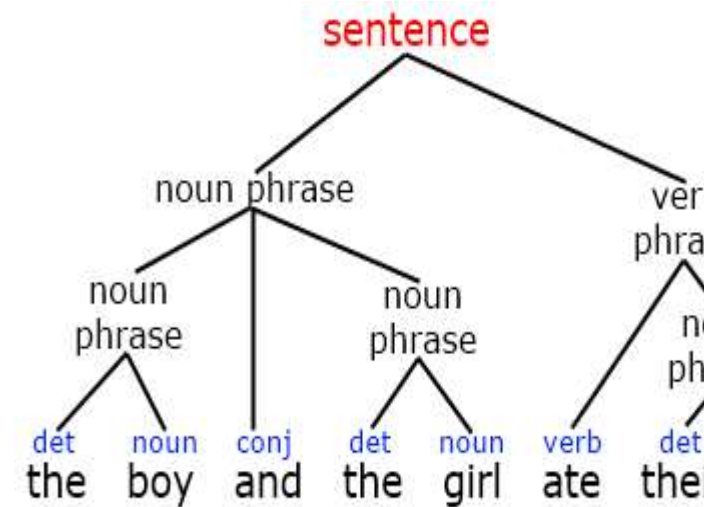
Level 2: Syntax

In linguistics, "syntax" refers to the rules that govern the ways in which words combine to form phrases, clauses, and sentences.

Syntax is the proper order of words in a phrase or sentence.

Syntax is a tool used in writing proper grammatical sentences.

<u>Lexical Categories</u>	<u>Syntactic Description</u>
Nouns	Nouns: Occur after determiners, determiners + adjectives, or adjectives
Adjectives	Adjectives: Determiners + ADJ +Noun, Linking verb +ADJ, Adv. +ADJ
Verbs	Verbs: Aux+Verb, Verb! (as an imperative)
Adverbs	Adverbs: ADV+adj, ADJ+Verb, Verb+ADV, ADV+ADV



Syntax describes the systematic ways in which words are combined to create well-formed phrases, clauses, and sentences as well as the systematic ways in which clauses and sentences combine to become more complex.”

Level 3: Semantics

Deals with the meaning conveyed by creating sentences in that language.

Semantics include tasks like named entity recognition and relationship extraction.

Semantic is the study of meaning that is used to understand human expression through language.



Level 4: Pragmatics

Understanding the sentences created and to understand the conveyed meanings.

we try to understand the text as a whole.

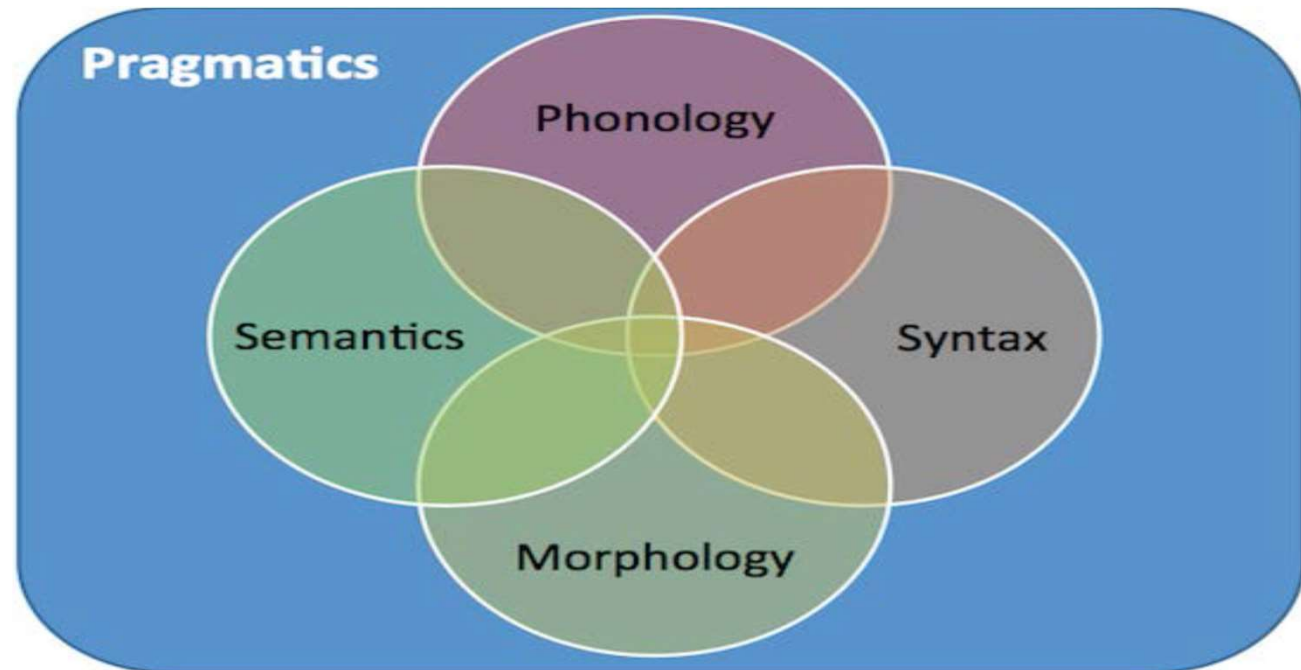
Common problems that are associated with pragmatics:

Co-Referencing

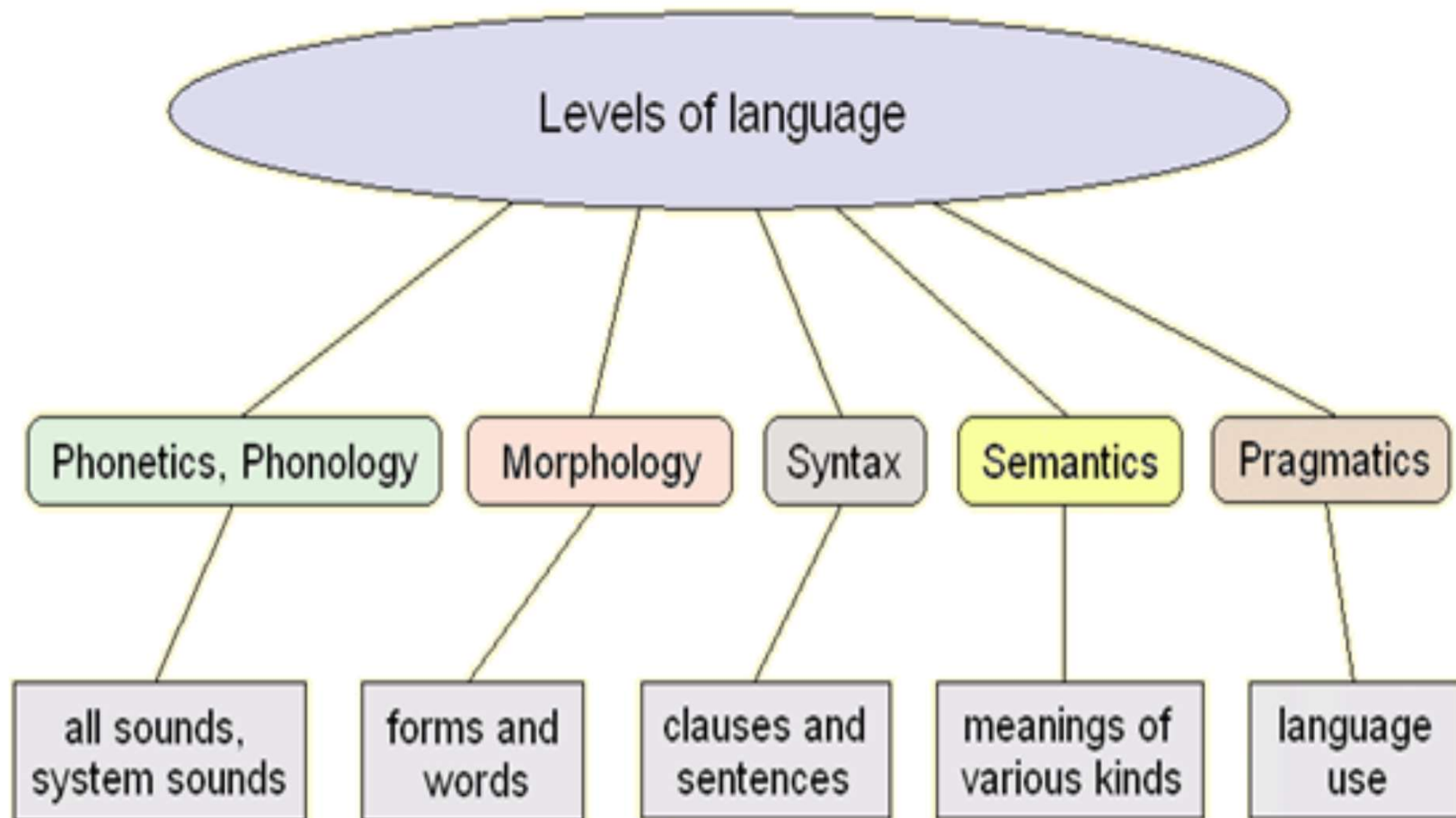
Summarization

Modelling.

Question and answering.



Understanding linguistics



Language Processing

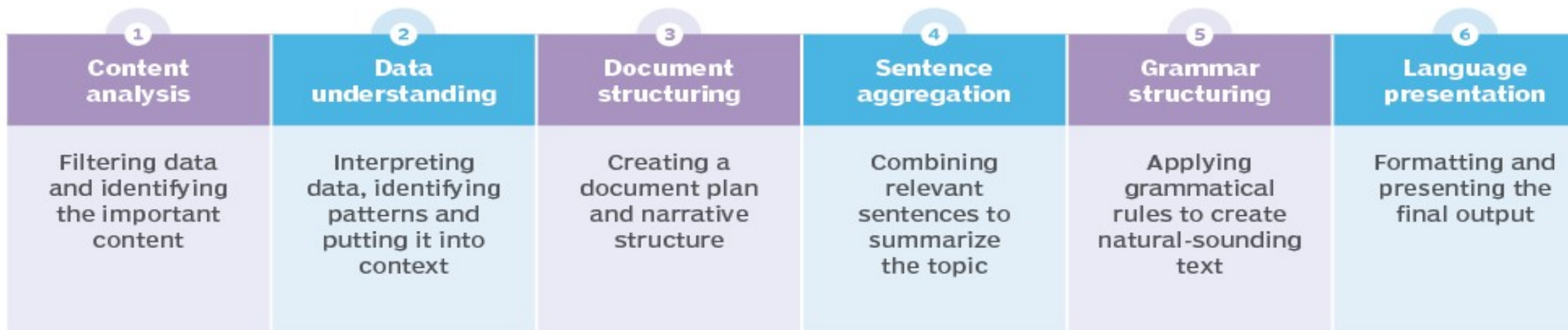
- *Level 1* – Speech sound (*Phonetics & Phonology*)
- *Level 2* – Words & their forms (*Morphology, Lexicon*)
- *Level 3* – Structure of sentences (*Syntax, Parsing*)
- *Level 4* – Meaning of sentences (*Semantics*)
- *Level 5* – Meaning in context & for a purpose (*Pragmatics*)
- *Level 6* – Connected sentence processing in a larger body of text (*Discourse*)

Natural language generation (NLG)

Natural language generation (NLG) is the use of artificial intelligence (AI) programming to produce written or spoken narratives from a data set.

NLG is related to human-to-machine and machine-to-human interaction, including computational linguistics, natural language processing (NLP) and natural language understanding (NLU).

6 steps to natural language generation



Libraries

NLTK

Scikit – learn

Spacy