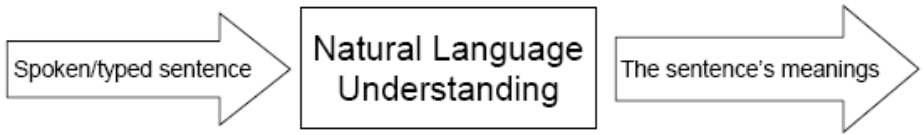
Natural Language Processing (NLP)

- Natural language refers to the way we, humans, communicate with each other i.e., **speech and text**
- Natural language processing (NLP) is a branch of artificial intelligence that helps computers understand, interpret, manipulate and generate human language.
- NLP draws from many disciplines, including computer science and computational linguistics, in its pursuit to fill the gap between human communication and computer understanding
- NLP enables a computer to communicate with humans or other machine the same way we, humans communicate
- NLP is automatic manipulation of natural language that includes
    1. Natural Language Understanding (NLU)
    2. Natural Language Generation (NLG)

Natural Language Processing (NLP)

Natural Language Understanding (NLU)

- NLU is extracting meaning from natural language sentences

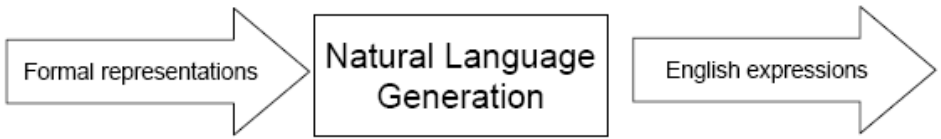


- NLU is one of the most challenging tasks in AI because of several factors we have to consider to understand a speaker's meaning
- We can have infinitely many sentences
- We can put together a sentence in several different ways and words may have different meaning in different context

Natural Language Processing (NLP)

Natural Language Generation (NLG)

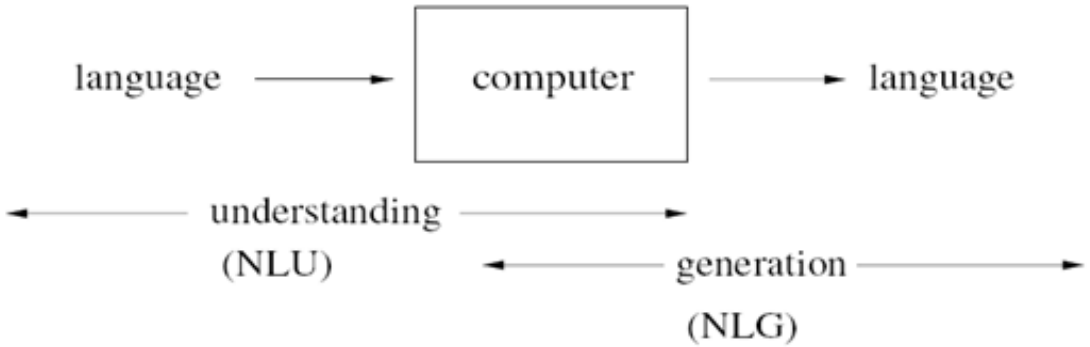
- NLG is converting a computer based representation into a natural language representation



- It includes:
- **Text planning** – retrieving the relevant content from knowledge base.
- **Sentence planning** – choosing required words, forming meaningful phrases
- **Text Realization** – mapping sentence plan into sentence structure.

Natural Language Processing (NLP)

- NLP = NLG + NLU



Knowledge Needed for NLP

- **Phonological** : Relating **sound to the words**
- **Morphological : word construction** from basic units
- **Syntactic** : how words are put together to form **grammatically** correct sentences
- **Semantic** : is concerned with the **meanings** of words and phrases
- **Pragmatic** : how the **context** affects the meaning of the sentence
- **Discourse** : how the immediately **preceding sentence can affect the interpretation of the next sentence**
- **World** : knowledge of the physical world, the world of **human social interaction**

Morphology

- **Morpheme** : smallest grammatical unit in a language
- **Word** : consisting of one or more morphemes
- E.g. **incoming** is a word made up of three morphemes : *in, come, -ing*
- Morpheme may or may not stand alone (has meaning), whereas a word is freestanding (meaningful)
- When a morpheme stands by itself, it is considered as a **root** because it has a meaning of its own (e.g. the morpheme **cat**) and when it depends on another morpheme to express an idea, it is an **affix** because it has a **grammatical function** (e.g. the –s in cats to indicate that it is plural)

- The field of study dedicated to morphemes is called **morphology**.
- Morphology is the study of words, how they are formed, and their relationship to other words in the same language

What can a morpheme do ?

Transform part of speech:

- affix like *-er* can transforms a verb into a noun (e.g. teach → teacher)
- affix like *-ness* can transforms an adjective into a noun (e.g. great → greatness)
- affix like *-er* changes an adjective into another degree of the same adjective (e.g. *small* → *smaller*)

Change features of nouns:

- affix like *-s* can do this. (e.g. cat → cats)

Change features of verbs :

- Aspect (e.g. I walk → I am walking. (present, progressive))
- Tense (e.g. I walked → I will walk. (past, future))
- Number and person (e.g. I walk → They walk. (first person singular, third person plural)

e.t.c

Morphological Analysis

- Morphological analysis is the process of segmenting a sentence into a row of morphemes
- The purpose of the morphological analysis is to determine the minimal units of meaning in a language or morphemes by using comparisons of similar forms
- for example, comparing forms such as “She is walking” and “They are walking” rather than comparing either of these with something completely different like "You are reading"
- Morphological analysis may also be helpful in part of speech tagging

Lexemes and Lexicon

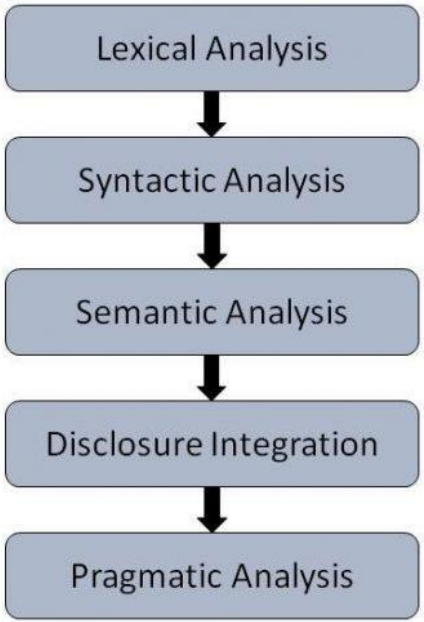
- Simply, a lexicon is a dictionary of all the valid vocabulary
- More precisely, A lexicon is a language's **collection of lexemes**
- **Lexemes**: A lexeme includes all inflected forms of a word

e.g.

WALK – walk, walks, walked, walking
RUN – run, runs, ran, running
SING – sing, sings, sang, sung, singing

- Lexicons can be used to check whether a word in a sentence is **a valid word** in the language or not and infer meaning between sentences that contains words which are part of same lexemes

Steps in Natural Language Processing



Lexical Analysis

- Lexical analysis includes dividing the text into words (**tokenization**) and checking whether the token is a **lexeme** or not
- E.g. I hate sumemr.

  I → is a Lexeme
  hate → is a Lexeme
  sumemr → is not a Lexeme

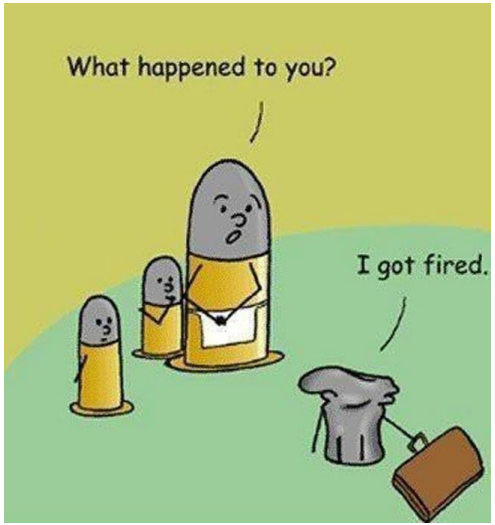
Lexical Ambiguity

- Lexical ambiguity is the presence of two or more possible meanings within a single word
- Example :

  bank : river bank or financial bank

  bat : cricket bat or mammal bat

- However, such words are often used by humans to create all sorts of pun and wordplay



Syntactic Analysis : Parsing

- Parsing is the process of analyzing the grammar of a string of symbols (sentence)
- Parsing is the process of analyzing a sentence by taking it apart word – by – word and determining its structure from its constituent parts and sub parts
- Parsing is formal analysis of a sentence into its constituents, resulting in a **parse tree** showing their syntactic relation to each other

- To find if a sentence is grammatically correct, we require a set of rules : grammar of the language
- The English grammar is almost context-free.

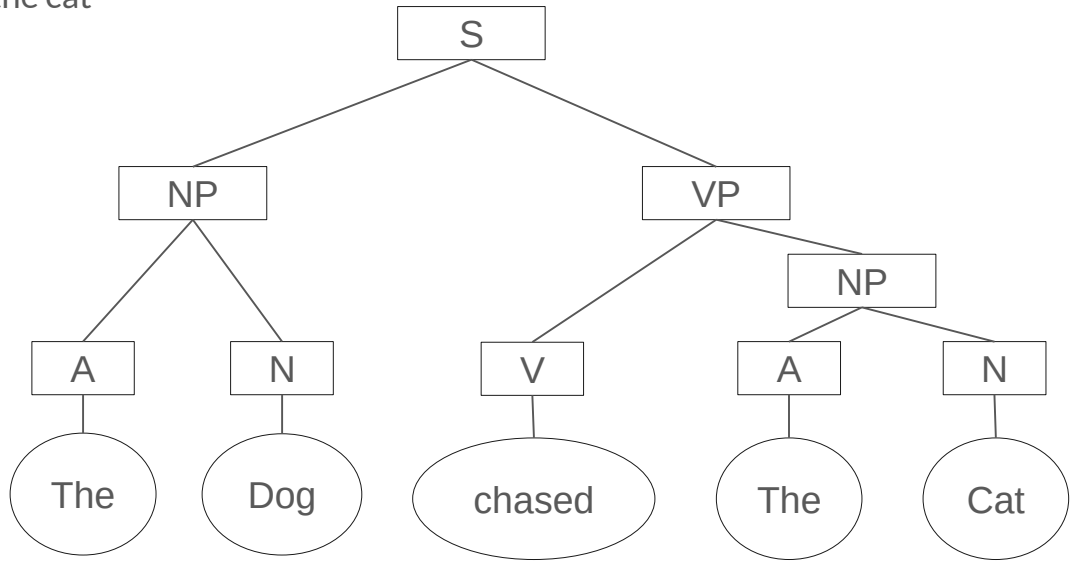
Syntactic Analysis : Phrase Structure Rules

- Sentence (S) → Noun Phrase(NP) + Verb Phrase (VP)
- Noun Phrase (NP) → Noun (N)
- Noun Phrase (NP) → Article (A) + Noun (N)
- Verb Phrase (VP) → Verb (V)
- Verb Phrase (VP) → Verb (V) + Noun Phrase (NP)
- Article(A) → a | an | the
- Noun (N)→ cow | dog | cat | lion | tiger
- Verb → ate | chased | got

etc..

Syntactic Analysis : An example of Parse tree

S= The dog chased the cat



Syntactic Ambiguity

- Syntactic ambiguity is the presence of two or more possible meanings within a single sentence
- A sentence is syntactically ambiguous if there are two or more possible groupings

E.g., The chicken is ready to eat.

- Wait whaaaaaaaaaaat ??
- Are we eating the chicken or the chicken is eating something ???
- The intended meaning of a syntactically ambiguous sentence can often (but not always) be determined by context

Semantic Analysis

- Semantic Analysis extracts the meaning of sentence from the words used
- It checks whether the sentence makes any sense or not
- For example
    - → “bitter sugar” this phrase does not make any sense
    - → “Spaghetti ate the customer” does not make sense

Discourse Integration

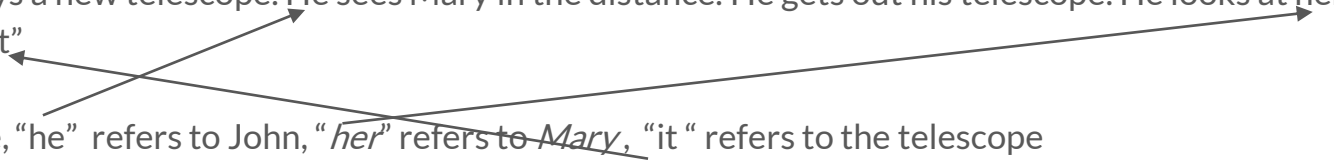
- The meaning of any sentence depends upon the meaning of the sentence just before it.
- In this step, we verify meaning of sentences with the sentences before it

Pragmatics Analysis

- Here, the sentences are re-interpreted to verify the correctness of meaning in **the given context**

Example : **Handling Pronoun**

"John buys a new telescope. He sees Mary in the distance. He gets out his telescope. He looks at her through it"



- Here, "he" refers to John, "*her*" refers to *Mary*, "it " refers to the telescope

"When is the next flight to Sydney?"
"Does it have any seat left?"

- Here, "*it*", refers to a particular flight to Sydney, not Sydney itself.

Pragmatics Ambiguity

- Pragmatic ambiguity arises when the statement is not specific, and the context does not provide the information needed to clarify the statement
- In most cases, we would require real world knowledge to solve pragmatics ambiguity

"I saw an astronomer with a telescope."

- Here, using real world knowledge, since we know that telescope is something an astronomer would have, we can say that the sentence is about an astronomer who has a telescope and is seen by someone (pronoun I)

"I saw a lady with a telescope."

- In this case however, anything may be the case but since telescope and lady has no practical relation as was with astronomer in above sentence, we could be certain that the person used telescope to see the lady

Parameters in NLP

- Auditory Inputs
- Segmentation
- Syntax Structure – see above slides
- Semantics Structure – see above slides
- Pragmatics – see above slides

Parameters in NLP : Auditory Inputs

- Audio devices used to feed sound to the NLP systems are the auditory inputs
- It could be achieved using microphones
- The task is then to map the sound waves to a string of words (speech recognition)
- Stuff to deal with
    - →background noise
    - →inter-speaker variation
    - →intra-speaker variation

Parameters in NLP : Segmentation

- **Text segmentation** is the process of dividing written text into meaningful units, such as words, sentences, or topics
- **Sentence segmentation** is the problem of dividing a string of written language into its component sentences
- **Word segmentation** is the problem of dividing a string of written language into its component words
- Most written languages have explicit boundary markers (like space to separate words or period (.) to separate sentences in English) which aids in segmentation
- **Word splitting** is the process of parsing concatenated text (i.e. text that contains no spaces or other word separators) to infer where word breaks exist