**1.7. Stages/Phases in NLP [Levels]**

The analysis of NL is broken down into various broad levels such as phonological, morphological, lexical, syntactic, semantic, pragmatic and discourse analysis.

For example some applications require the first 3 levels only. Also, the levels could be applied in a different order independent of their granularity.

**Level 1 – Phonology**

This level is applied only if the input is a speech. Input is acoustic waveform and output is string of words.

It deals with speech recognition and generation. That is interpretation of speech sounds within and across words

Speech sound might give a big hint about the meaning of a word or a sentence.

The area of computational linguistic that deals with speech analysis is computational phonology.

**Level 2- Morphological Analysis**

Morphology is study of internal structure of words. Given a particular word in a language, what are the different meaningful units it is made up of and each small unit is called a morpheme.

cat : stem cats: cat + s unhappy happily unhappily

cats : N +PL cat: N+SG sort: V+SG sorts: V+PL sort: N+PL

**Computational tool to perform morphological parsing is finite state transducer.**

**[Just read:** In Information Retrieval, document and query terms can be stemmed to match the morphological variants of terms between the documents and query; such that the singular form of a noun in a query will match even with its plural form in the document, and vice

versa, thereby increasing recall. **Compare with find and search of word]**

**Level 3 – Lexical Analysis**

It deals with understanding of everything about distinct words according to their position in the speech, their meanings and their relation to other words.

It identify and analyze the structure of words with respect to their lexical meaning and part-of-speech. Lexicon is a dictionary. Lexicon of a language means the collection of words and phrases in a language. Validity of word is checked according to lexicon.

“Duck”, for example, can take the form of a noun or a verb but its part-of-speech and lexical meaning can only be derived in context with other words used in the phrase/sentence.

e.g.

I cooked duck for her.

He ducked under the overhanging branches.

**Level 4 – Syntactic Analysis**

**Syntactic Analysis (Parsing)** − It involves analysis of words in the sentence for grammar and arranging words in a manner that shows the relationship among the words. The sentence such as “The school goes to boy” is rejected by English syntactic analyzer.

Syntax refers to the study of formal relationships between words of sentences. Validity of a sentence is checked according to rules of grammar.

To perform syntactic analysis, the knowledge of grammar and parsing techniques is required.

Grammar is formal specification of rules allowable in the language. Parsing is a method of analysing a sentence to determine its structure according to the grammar. CFG is used for syntactic analysis. Two basic parsing techniques are top down parsing and bottom-up parsing.

**I eat banana.**

**I banana eat.**

**Level 5 – Semantic Analysis**

Semantics deals with the meaning of natural language sentences. Meaning of the sentences is understood in this phase. It draws the exact meaning or the dictionary meaning from the text. The text is checked for meaningfulness. It is done by mapping syntactic structures and objects in the task domain. The semantic analyzer disregards sentence such as “hot ice-cream”.

e.g.

She eats banana

Machine eats banana.

**Level 6: Pragmatics and Discourse**

**Pragmatic :**

During this, what was said is re-interpreted on what it actually meant. It involves deriving those aspects of language which require real world knowledge.

Explains how extra meaning is read into texts without actually being encoded in them. This requires much world knowledge, including the understanding of intentions, plans, and goals. Consider the following 2 sentences:

* The city counsellors refused the demonstrators a permit because they feared violence.
* The city counsellors refused the demonstrators a permit because they advocated revolution.

The meaning of “they” in the 2 sentences is different. In order to figure out the difference, world knowledge in knowledge bases and inferencing modules should be utilized.

**Discourse :**

The meaning of any sentence depends upon the meaning of the sentence just before it. In addition, it also brings about the meaning of immediately succeeding sentence.

Focuses on the properties of the text as a whole that convey meaning by making connections between component sentences

1. I only like travelling to Europe. So I submitted a paper to ACL.
2. I only like travelling to Europe. Nevertheless, I submitted a paper to ACL.

 In example (1), you can infer ACL-is-in-Europe-this-year and in (2) you can infer the negation of that.