1. Introduction

1. Introduction

- History of NLP
- Generic NLP system
- Levels of NLP
- Knowledge in language processing
- Ambiguity in Natural language
- Stages in NLP
- Challenges of NLP
- Applications of NLP

NLP

Language

- Mode of Communication

Two types

- Artificial Language
- Natural language

Natural Language

Natural language

- Two types
  - Spoken form
  - **Written form**

Why processing??

NLP

- Natural Language Processing
  - The process of computer analysis of input provided in a human language (any natural language), and conversion of this input into a useful form of representation.

History of NLP

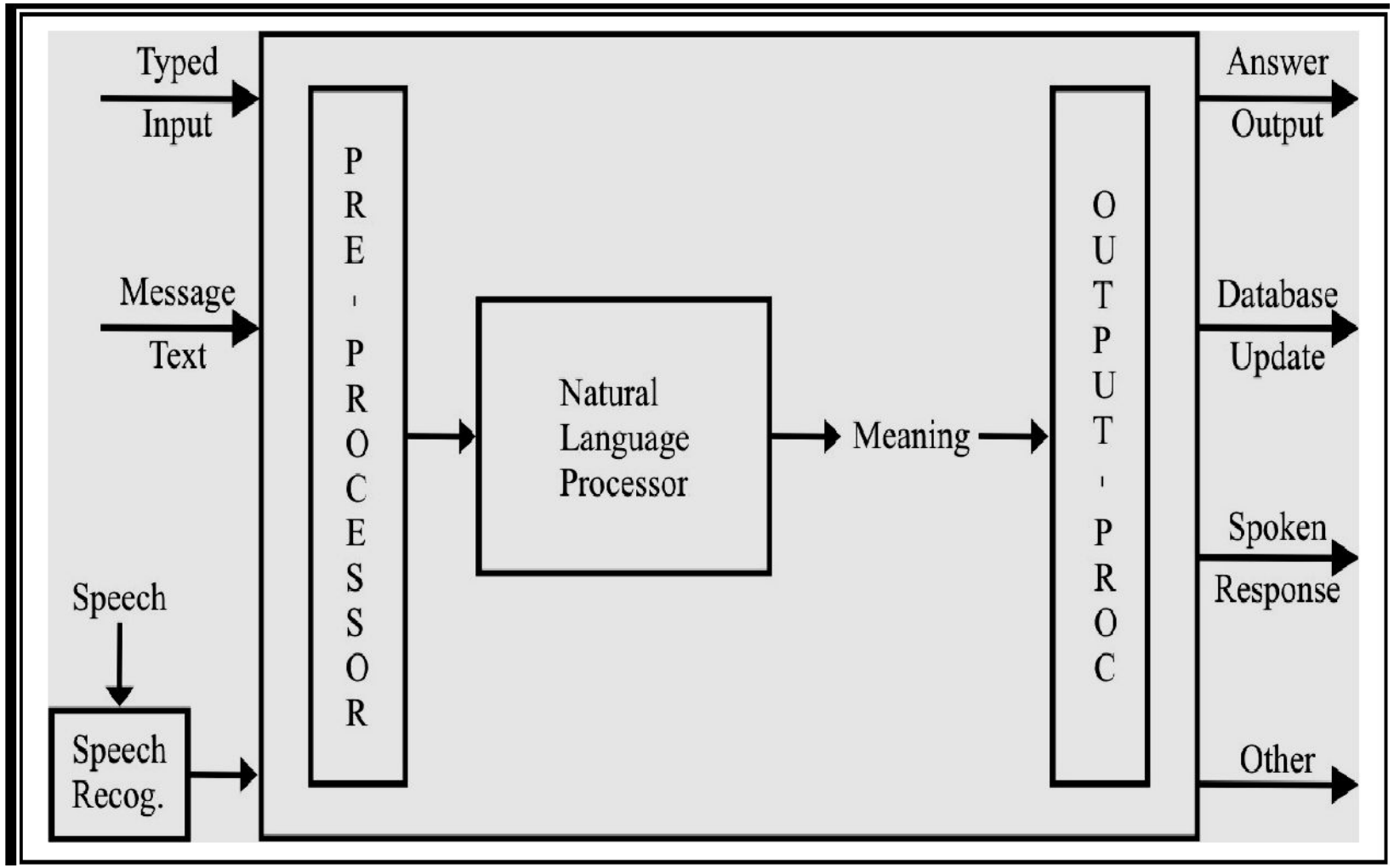
- **1950-** Attempts to automate translation between Russian and English
- **1960-** The work of Chomsky and others on formal language theory and generative syntax
- **1990-** Probabilistic and data-driven models had become quite standard
- **2000-** A Large amount of spoken and textual data become available

History of NLP

- 1950
- The obvious place to look for help was from Linguistics.
- The literature of the 1950s shows a growing awareness of work in mainstream Linguistics, and
- young researchers in Linguistics joined Machine Translation teams.

- 1960
- **John bought a ticket for Mary in the Symphony Hall Booking Office.**
- We know from the *position* of the words *John* and *ticket* that John is the agent instigating the action and that the ticket is the patient (or object) of the action.
- We know that Mary is the beneficiary of the action because of the use of the preposition *for* before her name.

Generic NLP systems



Levels of NLP

Phonology

Morphology

Syntax

Semantics

Reasoning

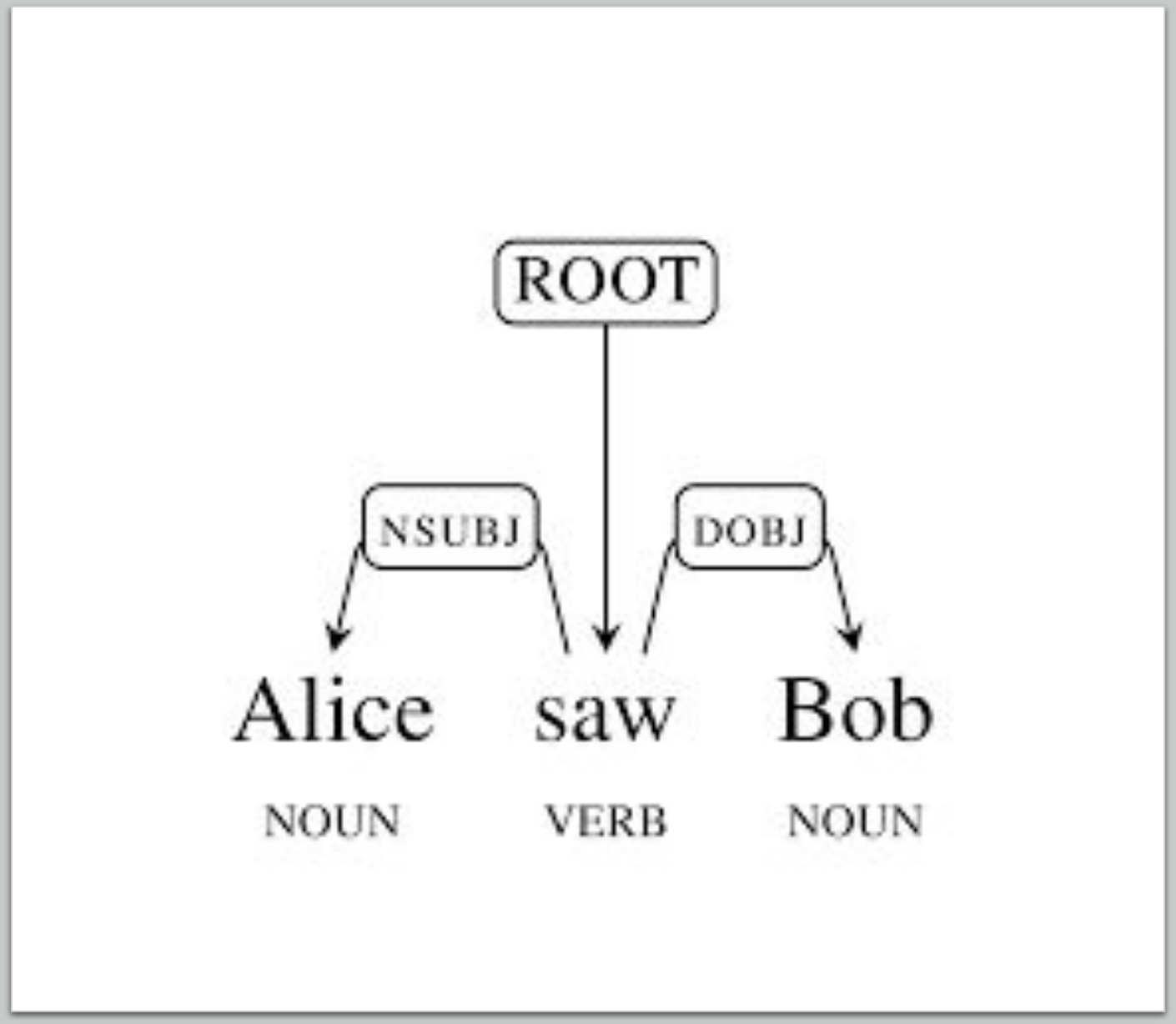
Levels of NLP

- Phonology
  - Speech processing
- Morphology
  - The **cats** are fighting

Levels of NLP

- Syntax/Syntactic processing
  - It is a study of formal relationships between words.

- The cat chased the mouse
  - Parts of speech
  - The--- determinant
  - Cat- noun
  - Chase- verb
  - Mouse - noun



Levels of NLP

- Semantics
  - Determining the meaning of the words
  - **Bat** is flying in the sky

Levels of NLP- reasoning

- To produce an answer to a question which is not explicitly stored in a database;
- Natural Language Interface to Database (NLIDB) carries out reasoning based on data stored in the database.
- For example, consider the database that holds the academic information about student,
- and user posed a query such as:
- **‘Which student is likely to fail in Maths subject?’.**
- To answer the query, NLIDB needs a domain expert to narrow down the reasoning process.

Knowledge in language processing

Bat is flying in the sky

[bat, flying, in, sky] [is, the]

Flying ---- fly. (Morphology)

[batNN, flyVB, inPP, skyNN]

Bat fly, sky. (dependency / syntactic)

Bat – mammal (semantics)

Ambiguity in Natural Language

Lexical Semantic Ambiguity	• *"The crane is loaded."* • *"The beak of the crane is very big."*
Lexical Ambiguity	• *"She received three silver vessels.* • *"Reena gave a silver talk"*
Syntactic Ambiguity	• *"He saw a man with binoculars."*
Anaphoric Ambiguity	• *"Cat went up the hill. It was slipery. It got angry"*

Ambiguity in Natural Language

Pronoun Translation

- *“Monkey ate the banana as it was hungry.”*
- *“Monkey ate the banana as it was ripe”*

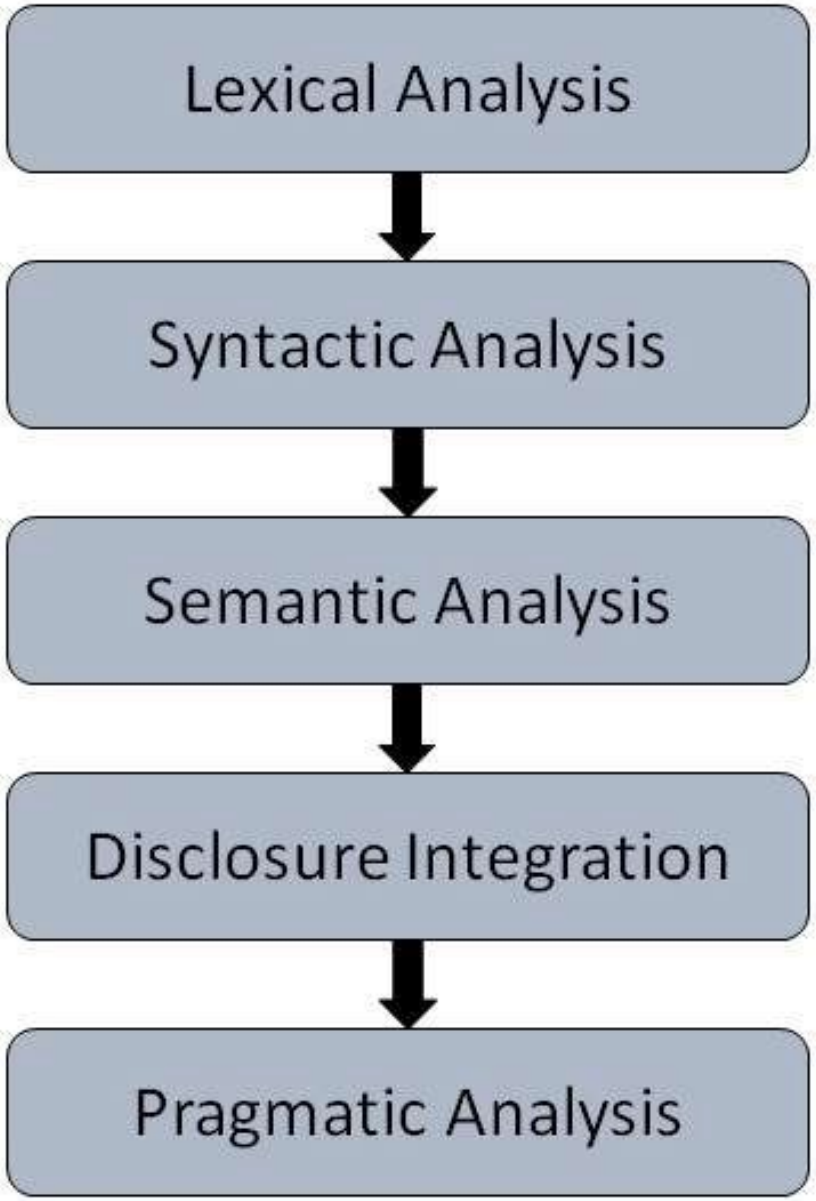
Parallel Corpus

- *“Monkey ate the banana as it was hungry”*
- *“माकडा ने केळ खाल्ल कारण तो भुकेला होता.”*
- *“बंदर ने केला खाया क्योंकि उसे भूख लगी थी।”*

Ambiguity

- *“Bank”*
- *“उत्तर”*

Stages in NLP



Stages in NLP

- Lexical Analysis
  - Refer the dictionary and obtain the properties of the word

  - **Eg: Dog**
    - Noun
    - Take ‘s’ in plural
    - Animate
    - 4 legged
    - carnivore

Stages in NLP

Challenges in Lexical analysis

First step: *part of Speech Disambiguation*

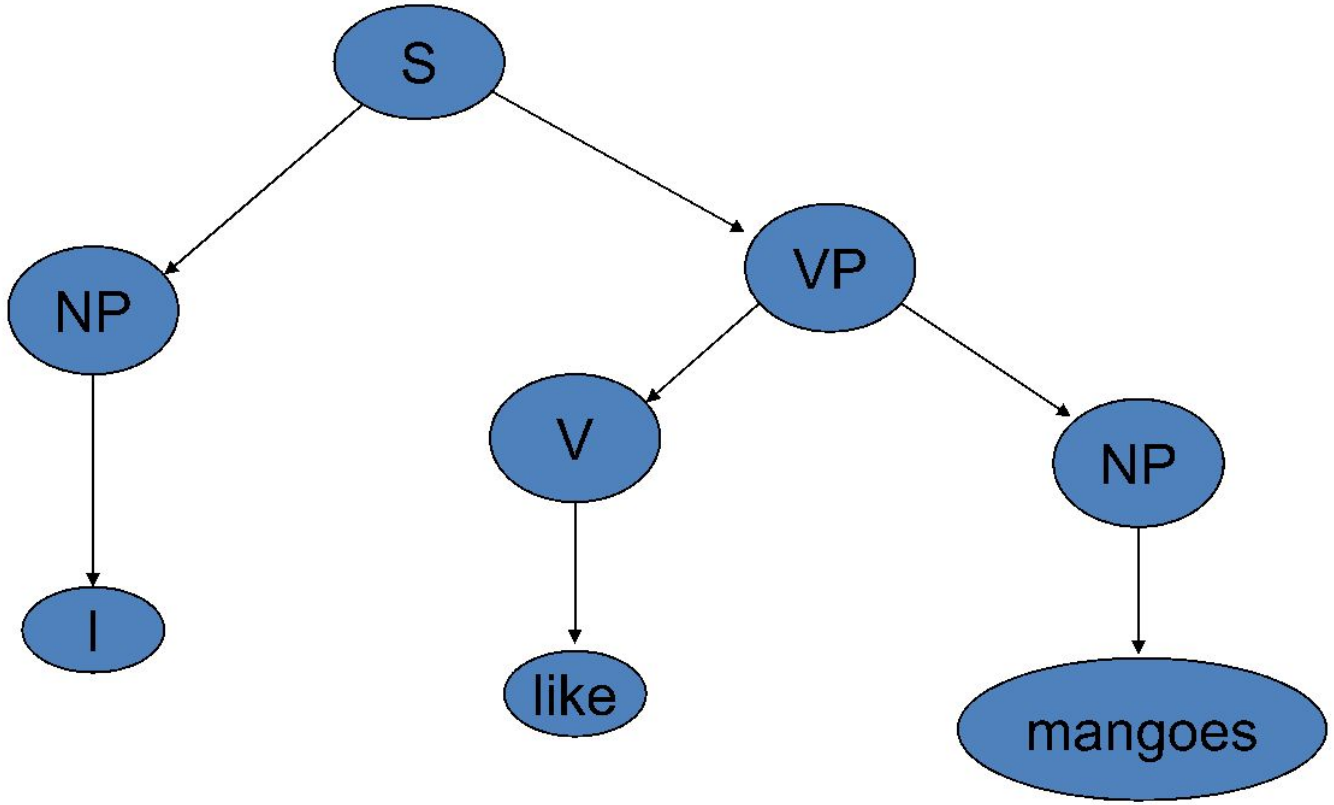
- *Dog* as a *noun* (animal)
- *Dog* as a verb (*to pursue)*

Sense Disambiguation

- *Dog (*as *animal)*
- *Dog (*as *a very detestable person)*

Stages in NLP

Structure Detection



Syntactic Parsing strategy

Driven by grammar

- S-> NP VP
- NP-> N | PRON
- VP-> V NP | V PP
- N-> Mangoes
- PRON-> I
- V-> like

Stages of NLP

- Challenges in Syntactic Processing

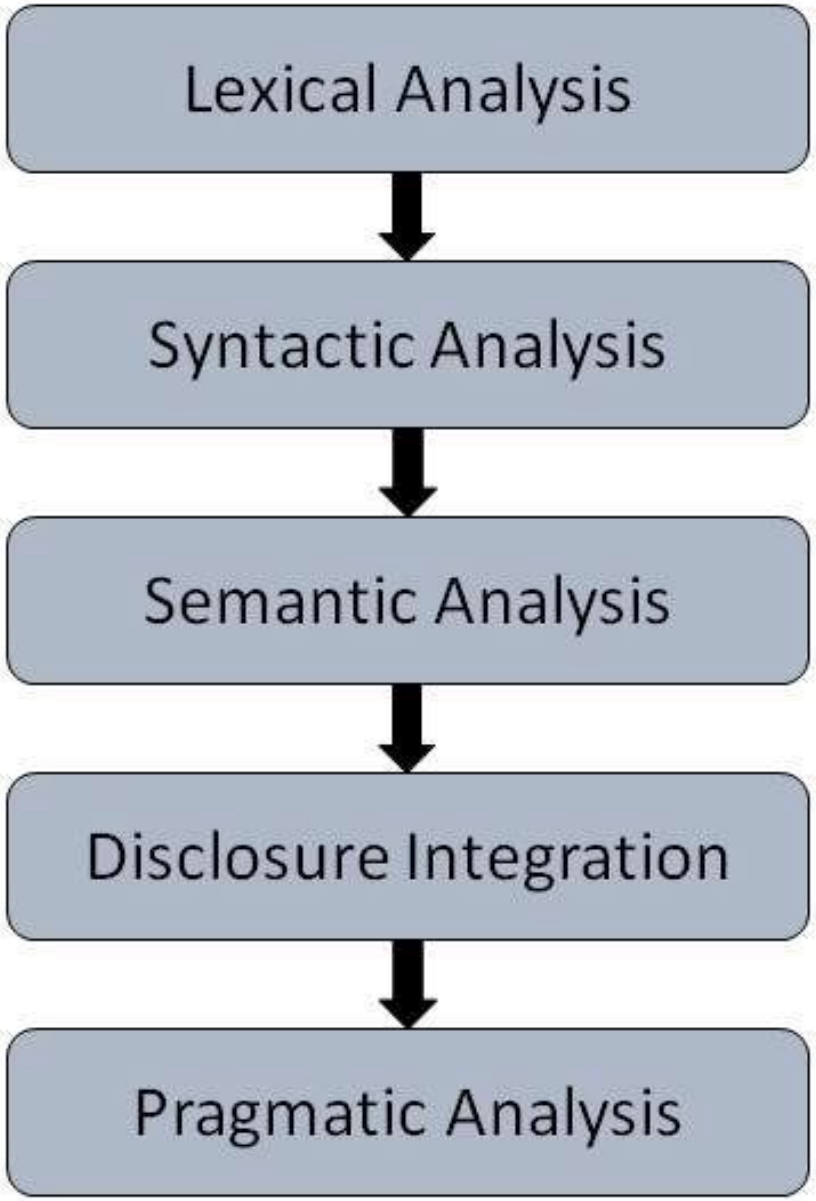
Scope

1.The old men and women were taken to safe locations
(old men and women) vs. *((old men) and women)*
2. No smoking areas will allow Hookas inside

Preposition Phrase Attachment

- *I saw the boy with a telescope*
  (who has the *telescope?*)
- I saw the mountain with a telescope
  (world knowledge: *mountain* cannot be an *instrument of seeing*)
- I saw the boy with the pony-tail
  (world knowledge: *pony-tail* cannot be an *instrument of seeing*)

Stages in NLP



Stages in NLP- semantic analysis

- John gave book to Mary
- Agent is john
- Object is book
- Mary is recipient

- Challenges:
- Bat/crane/ tank
- Uttar/peru in marathi

Stages in NLP- Disclosure Integration

- Disclosure Integration
  - 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.
  - Example
  - Manoj went to the bank. **He** said **it** was crowded

Stages in NLP- Disclosure Integration

Processing of *sequence* of sentences

Mother to *John*:

John go to school. It is open today. Should you bunk? Father will be very angry.

Ambiguity of *open*

bunk what?

Why will the father be angry?

Complex chain of reasoning and application of world knowledge

Ambiguity of *father*

father as *parent*

or

father as *headmaster*

Stages in NLP - Pragmatics

- 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.

Model user intention

- *Tourist (in a hurry, checking out of the hotel, motioning to the service boy): Boy, go upstairs and see if my sandals are under the divan. Do not be late. I just have 15 minutes to catch the train.*
- *Boy (running upstairs and coming back panting): yes sir, they are there.*

World knowledge

Challenges of NLP

- POS tagging

”Khaanaa” : can be noun (food) or verb (to eat)

Mujhe khaanaa khaanaa hai. (first khaanaa is noun and second is verb)

Challenges of NLP

- Pronoun resolution

- a. *The thieves stole the paintings. They were subsequently sold.*
- b. *The thieves stole the paintings. They were subsequently caught.*
- c. *The thieves stole the paintings. They were subsequently found.*

Challenges of NLP

Word sense disambiguation

a. *He went to the bank to withdraw money*

b. *He went near the river bank*

Applications of NLP

- Machine translation
- Information Retrieval
- Question Answering
- Sentiment analysis
- Text summarization