Module 1 : Introduction to Natural Language Processing

Q1 Introduction to Natural Language Processing ?

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

The field of Natural Language Processing is primarily concerned with getting computers to perform useful and interesting tasks with human languages.

The field of Natural Language Processing is secondarily concerned with helping us come to a better understanding of human language.

There are few major components in Natural Language Processing.

1. Natural Language Understanding
   - Mapping the given input in the natural language into a useful representation.
   - There are different level of analysis required here : Morphological Analysis, Syntactic Analysis, Semantic analysis, discourse analysis etc.

2. 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.

Goals of Natural Language Processing
- Design, implement and test systems that process natural languages (Ex: Marathi) for practical applications.

Q2 Steps in Natural Language Processing

1] Tokenization :
- Process of cutting big sentence into small tokens.
- Example :
[Dipesh] [is] [an] [abnormal] [human]

2] Stemming
- Normalizing words into there base or root forms
- Example : Knows, Known, Knowing → Know
↳ root form

Disadvantage : Sometimes root word is not Sensable.

Example : Date, Dated, Dating → Dat

3] Lemmatization :

Groups together different inflected forms of a word called Lemmatization. This is similar to stemming but here the output is always correct.

Example : Die, Died, Dead → Die

4. PoSTags :

This stands for Part of Speech Tags. It indicates how a word function in meaning as well as grammatically within a sentence.

Manan	killed	a	bat	and	ate	it
↓	↓	↓	↓	↓	↓	↓
Noun	verb	det	Noun	conjunction	verb	pronoun

Disadvantages : One word can have multiple part of speech, we use Name entity recognition for this problem.

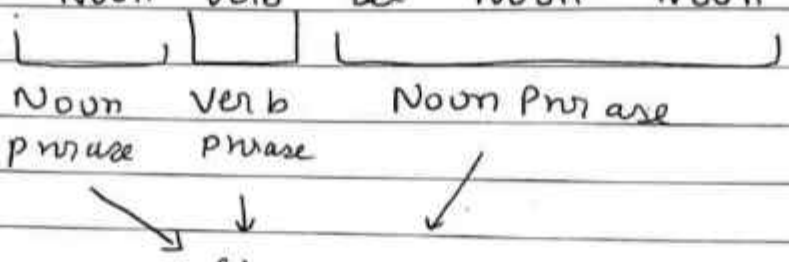
5. Name Entity Recognition :

It is a process of putting Name entity on the words : person, organization, company, location etc.

eg google it → company

6] Chunking:
Picking individual pieces of information and grouping them into bigger pieces.

Anurag ate the black cat
Noun verb Det Noun Noun



A Chunk helps in getting insightful and meaningful info from the text.

Q3 Stages / Levels of Natural Language Processing

=> Morphology : Concerns the way words are built up from smaller meaning bearing units
example : " truthtfultness"

Syntax : Concerns how words are put togetner to form correct sentences and what structural role each word has.
Example : " The dog ate my homework.
Not Syntactically correct.

Semantics : Concerns what words mean and how these meanings combine in sentences to form sentence meanings.
Eg : industrial plant / living organism
Doesn't make a proper meaning.

Pragmatics : concerns how sentences are used in different situations and how it affects the interpretation of the sentence.

Sentence might be correct but it might have semantic ambiguity - which means that sentence has two meanings.
eg : Dipesh loves his girlfriend and Manan does too

5 Discourse: concerns how the immediately preceding sentences affect the interpretation of the next sentence.

Q4. Write a short note on Ambiguity

- Ambiguity in Natural Language Processing can be reffered to as the ability of being understood in more than one way.

- Natural Language is very ambiguous.

Natural Language Processing has following ambiguities.

1. Lexical Ambiguity:
Ambiguity of a single word is called Lexical Ambiguity.

Example: I can play cricket

give me that can

both "can" have different meanings.

2. Syntactic Ambiguity:
This kind of ambiguity occurs when sentence is parsed in different ways.

Ex: Abid saw the man with the binnoculars

- Abid saw the man carrying binoculars

- Abid saw the man through the binoculars.

3. Semantic Ambiguity :

This occurs when actual or exact meaning of the phrase themselves can be mismatched or misinterpreted when even after syntax and meaning of individual word have been resolved

Eg : Manan loves his cat and divesh does too.

4. Anaphoric Ambiguity :

This kind of ambiguity arises due to use of anaphora entities in discourse

anaphora : When a noun is replaced by a pronoun and causes confusion

eg : The house is on a long street. It is very dirty.

5. Pragmatic Ambiguity : It occurs when a sentence gives its multiple interpretation or it is not specific

eg : I love you too

Note : Please Refer my video lectures. if not understood

Q5 Applications of Natural Language Processing.

=> 1. Question Answering by computer.

Question Answering (QA) System is a task of Automatically answering to the questions asked in the natural Language using either a pre-structured database or a collection of natural language documents.

It presents only the requested information instead of searching full documents like search engine.

The basic idea behind this is user just have to ask the question and the system will retrieve the most appropriate and correct answer.

Eg. Q: Who is the father of the Nation [India]
A. Mahatma Gandhi

2. Chatbots:
Intelligent chatbots are offering personalised assistance to the customers already. Analysists predict that the use of chatbots will grow 5 times year on year.

3. Managing the Advertisement Funnel

NLP is a great source for intelligent targeting and placement of advertisements in the right place, at the right time and for the right audience.

4. Market Intelligence

NLP gives exhaustive insights into employment changes and status of the market, tender delays, and closings which help in extracting information from large repositories.

5. Text Summarization:

It refers to the technique of shortning long pieces of text.

Summarization can be mainly classified into extractive and abstractive.

→ Extractive summarization → Extracting few sentences

→ Abstractive Summarization → It builds an internal semantic representation and then uses natural language generation techniques to create a summery.

In both of the cases NLP is used.