

ASSIGNMENT 1

- 1)
a) Explain ambiguity in each stage of NLP with an example
 → i) lexical ambiguity

The ambiguity of a single word is called lexical ambiguity.
 For eg, treating the word silver as noun, an adjective,
 or a verb.

eg: she made a silver speech

- ii) Syntactic ambiguity

This kind of ambiguity occurs when a sentence is parsed
 in different ways. For eg, the sentence "The man saw a girl
 with telescope". It is ambiguous whether the man saw
 the girl carrying a telescope or he saw her through
 his telescope.

- iii) Semantic Ambiguity

This kind of ambiguity occurs when the meaning of
 words themselves can be misinterpreted.

For eg "The car hit the pole while it was moving" is
 having semantic ambiguity because interpretations can be
 "The car, while moving hit the pole" and "The car hit
 the pole while the pole was moving".

- iv) Anaphoric Ambiguity

This kind of ambiguity arises due to use of anaphora
 entities in discourse. For eg, the horse ran up the hill.
 It was very steep. It soon got tired. Here, the anaphoric
 reference of 'it' in two situations cause ambiguity.

v) Pragmatic Ambiguity

such kind of ambiguity refers to the situation where the context of a phrase gives it multiple interpretations. For eg, the sentence "I like you too" can have multiple interpretations like I like you (just like you like me), I like you (just like someone else does).

- b) explain derivational and inflectional morphology in detail with suitable examples.

→ inflectional morphology

Inflection is a morphology process that adapts existing words so that they function effectively in sentences without changing category of base morphemes. Inflections morphemes of a word do not have to be listed in dictionary since we guess their meaning from root word.

To illustrate this consider following two sentences

- 1) To help my grandmother in her garden.
- 2) He is my grandmother's help.

use our general knowledge of words and their meaning shows that 1. Help is used as a verb and requires us working with our grandmother in order to support her. 2. Help is noun and stands for the person that regularly supports my grandmother.

* Derivational morphology

Derivational morphology is a type of word formation that creates new lexemes, either by changing syntactic category or by adding substantial new meaning to a free or bound base. Derived words may fit into number of semantic categories. For nouns, agent & result, personal and participant. For verbs, causative and applicative categories are well attested. Eg of derivational morphology.

- a) Lexical item (free morpheme) = like (verb)
+ prefix (bound morpheme) dis- = dislike (verb)
- b) Lexical item : like
+ suffix -able = likable
+ prefix -un = unlikeable
+ suffix -ness = unlikableness

c) Identify and describe the ambiguities in the following sentences

i) The man kept the dog in the house

→ Syntactic Ambiguity : There are two ways to parse the sentence

- (a) (The man) (Kept) (the dog in the house), meaning the man took ownership of dog that was in the house
- (b) (The man) (Kept) (the dog) (in the house) means the man forced the dog to remain in the house

ii) Book the flight

Lexical ambiguity: The word book can be treated as noun for eg book → number of printed pages on the other hand book can be treated as verb

The word flight has two meanings one i.e running away or avoiding something, another is journey by air

→ Derive any three applications of NLP. Explain the various stages involved in NLP process with suitable example.

→ Applications of NLP

i) Automatic summarization : Text summarization may be defined as the technique to create short, accurate summary of longer text documents . Automatic text summarization will help us with relevant information in less time, NLP plays an important role in developing an automatic text summarization .

ii) Question - Answering : Question answering is a computer science discipline within the fields of AI & NLP. It focuses on building systems that automatically answer questions posted by human beings in their natural language. The exact answer can be generated by doing syntax & semantic analysis of question.

iii) Sentiment Analysis : Another important application of NLP is sentiment analysis . As name suggests , sentiment analysis is used to identify sentiments among several posts. It is also used to identify the sentiment where

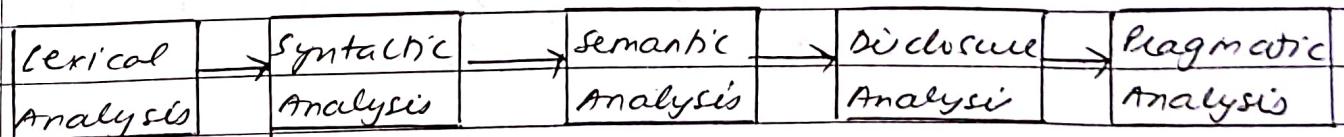
emotions are not expressed explicitly

* Stages of NLP

i) Lexical Analysis : It involves identifying and analyzing the structure of words. Lexicon of language means the collection of words and phrases in language. Lexical analysis is dividing the whole chunk of text into paragraphs, sentences and words.

"Duck" for eg, 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 phrase / sentence. This in fact is an early step towards a more sophisticated information retrieval system.

ii) Syntactic Analysis (Parsing) : It involves analysis of words in 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.



iii) Semantic Analysis : It concerns what words mean and how these meanings combine in sentences to form sentence meanings. It draws the exact meaning or dictionary meaning from the text. The semantic analyzer

disregards sentence such as "hot-ice-cream". Another eg can be (plant: industrial plant / living organism)

iv) Disclosure integration: The concern how to immediately preceding sentences after interpretation of next sentence. The meaning of any sentence depends upon the meaning of sentence just before it. In addition, it also brings about meaning of the immediately succeeding sentence

v) Pragmatic Analysis: This concern how sentences are used in different situation & how it affects the interpretation of the sentence. During this, what was said is interpreted on what it actually meant. It involves deriving those aspects of language which require real world knowledge.

3) What is the role of FSA in morphological analysis? Explain FST in detail. Design a finite transducer with E-insertion orthographic rule that passes from surface level "Foxes" to lexical levels "Fox+NPL" using FST

→ An FSA defines a formal language of defining a set accepted strings, while FST defines relations between sets of strings. In morphological parsing, an example would be inputting a string of letter into the FST, the FST would then output a string of morpheme

Finite State Transducer

- FSA automata

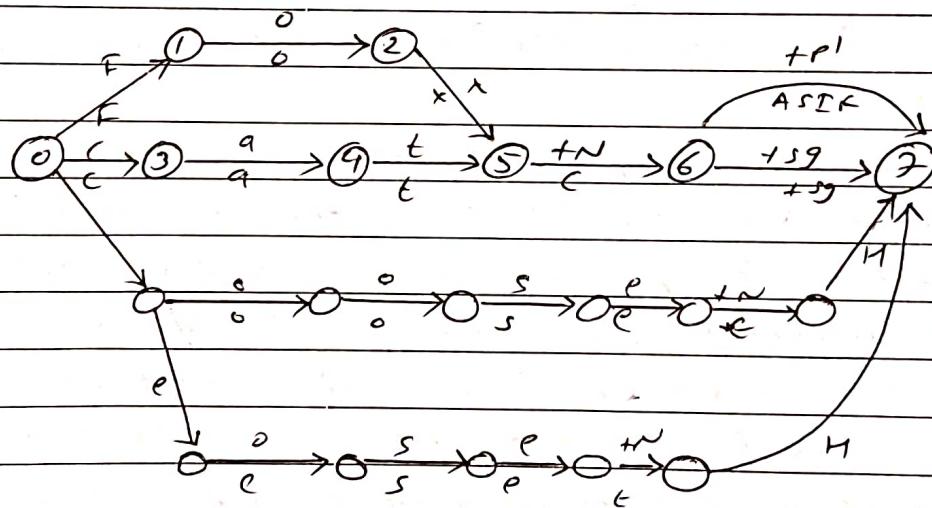
- An FSA represents a set of strings e.g. walk, walks, walked
- A recognise function
 - recognize (ct1) → true or false

(4)

- FST transducers

. An FST represents a set of pairs of strings (think of as iC, oP pairs) can return multiple answers if ambiguity.

e.g.: If you don't have pos tagged input "walk" could be the verb "They walk to the store" versus the noun "I took a walk"

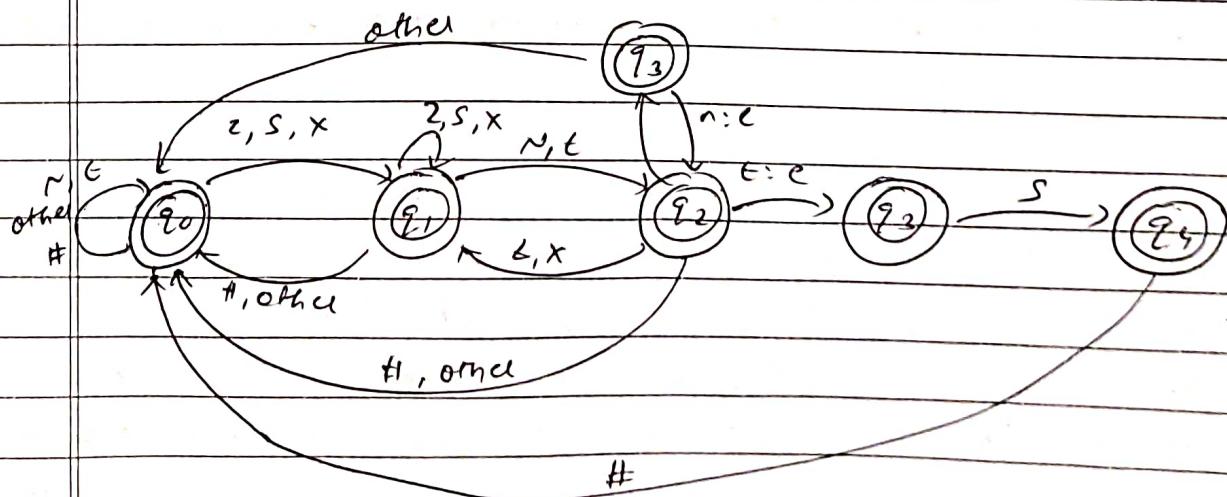


lexical

F	o	x	+N	+PL
---	---	---	----	-----

intermediate

f	o	x	Λ	s	H
---	---	---	---	---	---



lexical

f	o	x	+N	+PL
---	---	---	----	-----

Intermediate

f	o	x	\n	s#
---	---	---	----	----

surface

f	o	x	E	S
---	---	---	---	---

4) what is language model? write a note on n-gram language model

→ language modelling is one of the most important parts of modern NLP. There are many sort of applications for language modeling like machine translation, spell correction, speech recognition, summarization, question answering, sentiment analysis etc. Each of those tasks require use of language model. Language model is required to represent the text to a form understand from the machine point of view.

n-gram language model

n-gram models are widely used in statistical NLP. In search recognition, phonemes & sequences of phonemes are modeled using n-gram distribution

e.g:

This is big data AI book

unigram: This, is, big, data, AI, book = 6

Bigram: This is, is big, big data, data AI, AI Book = 5

Trigram: This is big, is big data, big data AI, data AI Book = 4

The chain rule probability is

$$P(w_1 \dots w_n) = P(w_1) \cdot P(w_2 | w_1) \dots P(w_3 | w_1, w_2) \dots P(w_n | w_1, \dots, w_{n-1})$$

The maximum likelihood estimate

$$P(w_i | w_{i-1}) = \frac{\text{Count}(w_{i-1}, w_i)}{\text{Count}(w_{i-1})}$$

e.g:

$\langle s \rangle$ I am Sam $\langle /s \rangle$

$\langle s \rangle$ Sam I am $\langle /s \rangle$

$\langle s \rangle$ I do not like green eggs and ham $\langle /s \rangle$

$$P(I | \langle s \rangle) = 2/3$$

$$P(\text{Sam} | \langle s \rangle) = 1/3$$

$$P(\text{am} | I) = 2/3$$

$$P(\langle ss \rangle | \text{Sam}) = 1/2$$

$$P(\text{Sam} | \text{am}) = \frac{1}{2}$$

$$P(\text{do} | I) = 1/3$$

5) Explain various approaches to perform Part of speech (POS) tagging.

- Part of speech tagging may be defined as the process of assigning one of the POS to given word. In simple words, POS tagging is a task of labelling each word in a sentence with its appropriate POS.
- Most of POS tagging falls under Rule base, stochastic and transformation based tagging.

i) Rule Based Pos tagging

One of the oldest techniques of tagging is rule based pos tagging. Rule Based taggers use dictionary or lexicon for getting possible tags for tagging each word. If word has more than one feasible tag, then rule-based taggers use hand-written rules to identify the correct tag. Disambiguation can also be performed in this by analyzing the linguistic function of a word along with its preceding as well as following word.

Rules may be either

- context - pattern rules
- regex into FSA

ii) Stochastic pos tagging

The model that includes frequency or probability can be called stochastic.

The simplest stochastic trigger applies two following approaches for pos tagging

Word frequency approach

In this the stochastic tagger disambiguates the words based on probability that word occurs with a particular tag. The main issue is that it may yield inadmissible sequence of tags.

Tag sequence probabilities

It is another approach where tagger calculates the probability of given sequence of tags occurring. It is also called n-gram approach.

* Transformation Based Tagging

It is also called as TBT tagging . It is an instance of transformation based learning which is a rule based algorithm for automatic tagging of PoS. TBT allows us to have linguistic knowledge . In readable form, transforms one state to another using transformation rules if we see similarity between rule based and transformation tagger . Other like rule - based it is also based on rules which specify what tags to be needed to assigned specific words .