

①/④ ①/③

7284696
Sem 8
NLP
CHOICE BASED

Q3)
(AR 1)

1) Word sense disambiguation, in NLP, may be defined as the ability to determine which meaning of word is activated by the use of word in a particular context.

2) Lexical ambiguity, Syntactic or semantic, is one of the very first problem that any NLP system faces.

3) part of speech taggers with high-level of accuracy can solve word's syntactic ambiguity.

4) On the other hand, the problem of resolving semantic ambiguity is called WSD (Word Sense Disambiguation)

For e.g.

I can hear bass sound

The occurrence of the word bass clearly denotes the distinct meaning.

Hence disambiguated in WSD as

I can hear bass/frequency sound.

Dictionary based

1) As the name suggests, for disambiguation, these methods primarily rely on dictionaries, thesauruses and lexical knowledge base.

2) They do not use corpora evidences for disambiguation. The first method is the seminal dictionary-based method introduced by Michael Lesk in 1986.

3) The Lesk definition, on which the Lesk algorithm is based is 'measure overlap between sense definitions for all words in context.'

4) However in 2000, Kilgariff and Rosenzweig gave the simplified Lesk definition as 'measure overlap between sense definitions of word and current context', which further means identify the correct sense for one word at a time.

5) Here the current context is the set of words in surrounding sentence or paragraph.

(2) / (3)

7284696
NLP
Sem 8
CHOICE BASED

Q3B1

e.g

Pine cone

1) pine

- a) kinds of evergreen tree with needle shaped leaves
- b) waste away through sorrow or illness.

2) Cone

- a) solid body which narrows to a pt.
- b) something of this shape whether solid or hollow
- c) fruit of certain evergreen trees.

$$\text{Pine}_a \cap \text{Cone}_a = 0$$

$$\text{Pine}_a \cap \text{Cone}_b = 0$$

$$\text{Pine}_a \cap \text{Cone}_c = 2$$

$$\text{Pine}_b \cap \text{Cone}_a = 0$$

$$\text{Pine}_b \cap \text{Cone}_b = 0$$

$$\text{Pine}_b \cap \text{Cone}_c = 0$$

(Q3) A1°

1) Context-free grammar is a list of rules that define the set of all well-formed sentence in a language. Each rule has a left-hand side, which identifies a syntactic category and right hand side, which defines its alternative component parts, reading from left to right.

2) There are the main new ideas

consistency, grammatical relations and subcategorization and dependencies.

3) The fundamental idea of consistency is that groups of words may behave as a single unit or phrase called a constituent. For e.g we will see that group of words may behave

(4) as a single unit.

(1) For e.g the verb want can be followed by an infinitive. as in I want to fly to Detroit or a noun phrase, as in I want to flight to Detroit. But the verb find cannot be formed followed by an infinitive. These are called facts about the subcategory.

③/④③

7284696
NLP
Sem8
CHOICE BASED

Q3A1)

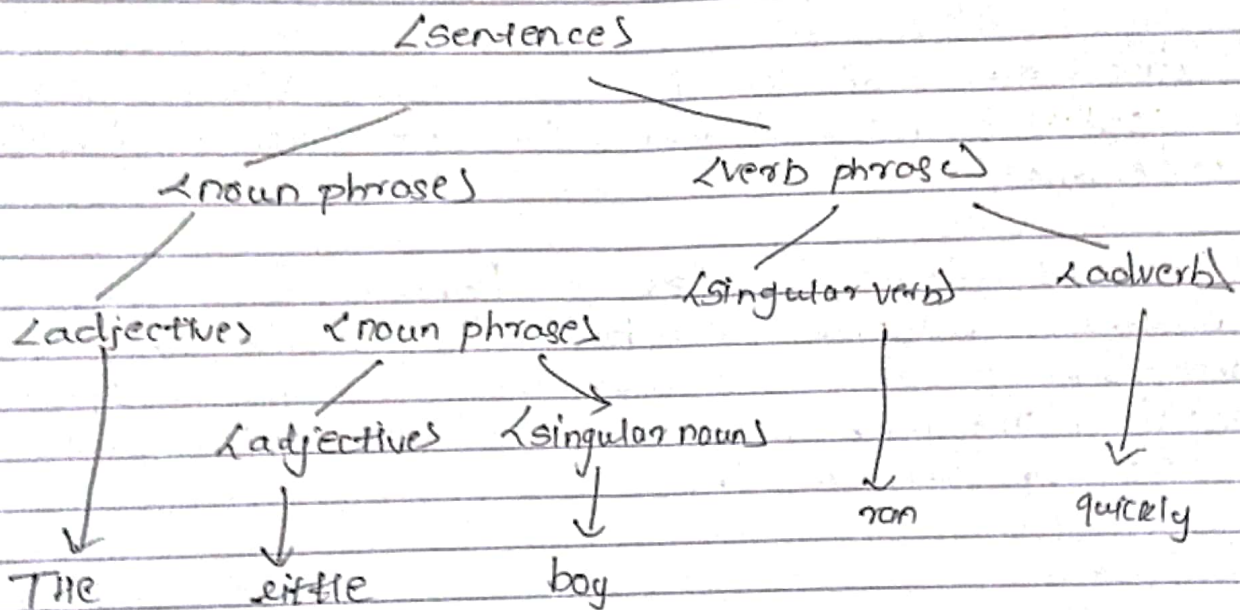
of the grammar.

5) CFG are thus the backbone of many models of the syntax of natural language.

6) As such they are integral to most models of natural language understanding of grammar checking and more recently of speech understanding.

e.g

The little boy ran quickly.



Q3A11)

IR

1) Task of finding text documents which are relevant to a user's informatⁿ need.

2) Document retrieval

3) Actual information buried inside document

4) long listing of documents

5) Describe details of Google which is best IR System for the web.

IE

Goal is to extract pre-specified features from documents or display information.

Feature retrieval

Extract informatⁿ from within the document.

Aggregate over entire collectⁿ

Extracted features are usually entered into a d.B automatically