

(1)

Name - Sudhansu Borade

Roll No - 20BAI10302

Code - CSA 4006.

Slot - B21 + B22 / 0328.

Subject - NLP

Q1 Context Free Grammar (CFG)

- A CFG is a list of rules that define the set of all well formed sentences in a language.
- Each rule has a left hand side, which identifies a syntactic category, and a right hand side, which defines its alternative component parts.
Reading ~~to~~ from left to right.
- A CFG only defines a language.
- It cannot say whether a given string belongs to the language it defines.
- To help this, to map a string of words to its parse tree, a parser can be used.
- A parser can work either top-down or bottomup when proceeding through the search space created by the set of parse trees generated by a CFG.
- A tree is built from the top. ~~or~~ when built by the top down parser, working from the start symbol towards the string.

This is a ~~to~~ top down parsing process starting from the symbol S . productions are applied to rewrite a symbol in the generated string until the input string is derived.

eg. $S \Rightarrow NP, VP \Rightarrow PN, VP \Rightarrow PN V_i \Rightarrow \text{John } V_i$

\Downarrow
sudi V_i

\Downarrow
sudi sleeps.

By con

On the flip side, a bottom-up parser starts from the input string & applies the productions in reverse until the start symbol is reached.

~~to~~ For example, assuming a DFS of search space & left to right processing of the input string, a bottom up derivation of sudi sleeps given in the grammar $G1$ would be.

sudi sleeps $\Rightarrow PN \text{ sleeps} \Rightarrow PN V \Rightarrow NP V \Rightarrow NP$
 \Downarrow
 \Downarrow
 S

The parse tree of course remains the same,
~~the remainder of this chapter, we~~

Q 2*) Time Flies like an arrow.

Some of the possible meanings of this sentence.

- a) Time goes as quickly as an arrow
- b) For every 'time fly' (a kind of an insect) 'n', there is an arrow 'y' such that n likes y.
- c) There exists a particular arrow 'k' such that every 'time fly' (a kind of insect) likes n. Many flies one arrow.

The difference between a), (b) & (c) arises because of syntactic or (lexical ambiguity): "like" can be either a comparative preposition or a verb.

Time here if interpreted incorrectly, can give a literal sense of something flying through the air.

2) He crushed a key to my heart.
Possible meanings:

- a) The key to my heart is a common metaphor: If someone holds the key to your heart, it means roughly that you have to love them.

(3)

b) There was ~~a~~ literally a key which was crushed to your heart.

There is another syntactic ambiguity with some semantic ~~over~~ touch.

On reading a) the verb ~~is~~ "crush" is being used in simple transitive structure.

Q32) looked up the tree.

- Verb Phrase.

Because looking up ~~to a tree~~ is ~~an action~~
hence "Looked up a tree" is a Verb Phrase.

Q31) Important to Bill.

- Adjective Phrase.

- Because ^{here} ~~that~~ something is important to someone.

Here, "important" ^{describes} the way ~~a~~ a thing is related to someone.

94. An augmented grammar is any grammar. is any grammar whose productions are augmented with conditions expressed using features.

- Any ~~non~~ nonterminal symbol in a derivation may be associated with features.

- A feature associated with a non terminal symbol is shown following the non terminal ~~sp~~ symbol is shown following that ~~non~~ non terminal ~~sp~~ separated from it by a " "

An augmented context free grammar for generating the strictly indined language $L_n = \{a^n b^n c^n : n \geq 0\}$. Non terminal symbols that occur more than one in a production are distinguished by indices

The non terminal symbol to the left of the rewrite arrow ' \rightarrow ' gets index 0. (eg A_0) & recurring non terminals to the right of the arrow get indices 1, 2, ... from left to right eg A_1 . COUNT is an integer feature. (a feature that takes integer values) of each of the non terminal symbols.

1/9/20

95. Lexeme

In linguistics, a ~~lex~~ lexeme is the fundamental unit of the ~~lex~~ lexicon (or word stock) of a language.

- Also known as a ~~lex~~ lexical unit, lexical item or lexical word.
- In corpus linguistics, lexemes are commonly referred to as lemmas.
- A lexeme is often but not always an individual word (~~a~~ simple lexeme).
- A single simple lexeme may have a number of inflectional forms or grammatical variants.
- A multiword or composite lexeme is a lexeme made up of more than one ~~&~~ phrasal verb, an open compound, or an idiom (make happy when the sun shines).
- The way in which a lexeme ~~to~~ can be used in a sentence is determined by its ~~or~~ word class or grammatical category.

Lexicon

- It is the vocabulary of a language or a branch of knowledge.
- In linguistics, a ~~lex~~ lexicon is a language's inventory of ~~or~~ lexemes.

- Linguistic theory is of 2 parts:- lexicon and grammar
- Items in the lexicon are called lexemes, or lexical items or word forms
- Lexemes are not atomic elements but contain both phonological & morphological components.
- To describe size of lexicon; lexemes are grouped into lemmas