Context Free Grammar (CFG)

Context Free Grammar is formal grammar, the syntax or structure of a formal language can be described using context-free grammar (CFG), a type of formal grammar.

The grammar has four tuples: (V,T,P,S).

- V - It is the collection of variables or nonterminal symbols.
- T - It is a set of terminals.
- P - It is the production rules that consist of both terminals and nonterminal.
- S - It is the Starting symbol.

A grammar is said to be the Context-free grammar if every production is in the form of: G -> (V∪T) *, where G ∈ V

- And the left-hand side of the G, here in the example can only be a Variable, it cannot be a terminal.
- But on the right-hand side here it can be a Variable or Terminal or both combination of Variable and Terminal.

Above equation states that every production which contains any combination of the 'V' variable or 'T' terminal is said to be a context-free grammar.

For example the grammar A = { S, a,b, P,S} having production :

Here S is the starting symbol.

{a,b} are the terminals generally represented by small characters.

P is variable along with S.

S-> aS

S-> bSa

but

a->bSa, or

a->ba

is not a CFG as on the left-hand side there is a variable which does not follow the CFGs rule.

In the computer science field, context-free grammars are frequently used, especially in the areas of formal language theory, compiler development, and natural language processing. It is also used for explaining the syntax of programming languages and other formal languages.

In Natural Language Processing (NLP), a Context-Free Grammar (CFG) is used to describe the syntax of a language in a formal way. It consists of a set of production rules that define how sentences in the language can be generated. CFGs are powerful because they can capture many of the hierarchical structures inherent in natural languages, such as nested clauses and phrases.

Components of CFG

- Terminals: The basic symbols from which strings are formed. In natural languages, these are typically words or tokens.
- Non-terminals: Symbols that can be expanded into sequences of terminals and non-terminals. They represent grammatical categories or constituents such as noun phrases (NP) or verb phrases (VP).
- Production Rules: Rules that define how non-terminals can be replaced by sequences of terminals and non-terminals. They are written in the form $A \rightarrow \alpha$, where A is a non-terminal and α is a string of terminals and non-terminals.
- Start Symbol: A special non-terminal from which the generation of strings begins. It usually represents a complete sentence (S).

In the context of a Context-Free Grammar (CFG), non-terminals are symbols that represent sets of strings and can be further expanded into other non-terminals and terminals according to the production rules. Each non-terminal typically represents a grammatical category or a part of the sentence structure. Here's the meaning of the non-terminals used in the example:

1. S (Sentence): Represents a complete sentence. It is the start symbol in the grammar and encompasses the entire structure of a valid sentence.

2. NP (Noun Phrase): Represents a noun phrase, which can be a subject, object, or complement in a sentence. It typically consists of a determiner followed by a noun but can also include adjectives, possessives, and other modifiers.

3. VP (Verb Phrase): Represents a verb phrase, which includes the verb and its complements, objects, or modifiers. A verb phrase can consist of a verb alone or be expanded to include additional elements like noun phrases and prepositional phrases.

4. Det (Determiner): Represents determiners, which are words that introduce nouns and express aspects such as definiteness, quantity, and possession. Examples include "the", "a", "an", "some", "many", etc.

5. N (Noun): Represents nouns, which are words that denote people, places, things, or concepts. Examples include "cat", "mouse", "book", etc.

6. V (Verb): Represents verbs, which are words that denote actions, occurrences, or states of being. Examples include "saw", "eat", "run", etc.

7. PP (Prepositional Phrase): Represents prepositional phrases, which begin with a preposition and are followed by a noun phrase. Prepositional phrases typically provide additional information about time, location, direction, and other relationships.

8. P (Preposition): Represents prepositions, which are words that show the relationship between a noun (or pronoun) and other parts of the sentence. Examples include "with", "in", "on", "at", etc.

Example CFG for a Simple Language

Let's consider a simple CFG for a fragment of English:

Terminals: {"the", "cat", "saw", "mouse", "with", "a"}

Non-terminals: {S, NP, VP, Det, N, V, PP, P}

Start Symbol: S

Production Rules

S → NP VP
NP → Det N
VP → VP PP |V NP
PP → P NP
Det → "the" | "a"
N → "cat" | "mouse"
V → "saw"
P → "with"

Generating a Sentence

Let's generate the sentence "the cat saw the mouse with a mouse" using the CFG:

1. Start with the start symbol S.
2. Apply the rule $S \rightarrow \text{NP VP}$:
   - $S \rightarrow \text{NP VP}$
3. Expand NP using $NP \rightarrow \text{Det N}$:
   - $S \rightarrow \text{Det N VP}$
4. Choose $\text{Det} \rightarrow \text{"the"}$ and $\text{N} \rightarrow \text{"cat"}$:
   - $S \rightarrow \text{"the""cat"VP}$
5. Expand VP using $VP \rightarrow \text{V NP}$:
   - $S \rightarrow \text{"the""cat"V NP}$
6. Choose $\text{V} \rightarrow \text{"saw"}$:
   - $S \rightarrow \text{"the""cat""saw"NP}$
7. Expand NP using $NP \rightarrow \text{Det N}$:
   - $S \rightarrow \text{"the""cat""saw"Det N}$
8. Choose $\text{Det} \rightarrow \text{"the"}$ and $\text{N} \rightarrow \text{"mouse"}$:
   - $S \rightarrow \text{"the""cat""saw""the""mouse"}$
9. Expand VP using $VP \rightarrow \text{VP PP}$:
   - $S \rightarrow \text{"the""cat""saw""the""mouse"PP}$
10. Expand PP using $PP \rightarrow \text{P NP}$:
    - $S \rightarrow \text{"the""cat""saw""the""mouse"P NP}$
11. Choose $\text{P} \rightarrow \text{"with"}$:
    - $S \rightarrow \text{"the""cat""saw""the""mouse""with"NP}$
12. Expand NP using $NP \rightarrow \text{Det N}$:
    - $S \rightarrow \text{"the""cat""saw""the""mouse""with"Det N}$
13. Choose $\text{Det} \rightarrow \text{"a"}$ and $\text{N} \rightarrow \text{"mouse"}$:
    - $S \rightarrow \text{"the""cat""saw""the""mouse""with""a""mouse"}$

- The final generated sentence is "the cat saw the mouse with a mouse".