Context Free Grammar and Context Free Language

Formal definition of a Context-Free Grammar (CFG):

A CFG is a quadruple (4-tuple), that is, a system which consists of 4 elements. We describe a CFG, G as follows:

G = (V, T, P, S), where

V - finite nonempty set of *variables* or *non-terminal symbols*; Each variable represents a language.

T - finite nonempty set of *terminal symbols*, $V \cap T = \Phi$;

P - finite nonempty set of *productions* or grammar *rules* of the form,

$A \rightarrow \alpha$, where

[head] $A \in V$,

[body] $\alpha \in (V \cup T)^*$ and

[production symbol] '$\rightarrow$' means 'could take the value' / 'can be replaced with';

S - One of the variables represents the language being defined; it is called the *start symbol*, $S \in V$.

Example:

G_1:

$S \rightarrow Abb$, $A \rightarrow \varepsilon$ $A \rightarrow aA$ $A \rightarrow bA$.	OR	$S \rightarrow Abb$, $A \rightarrow \varepsilon \mid aA \mid bA$.

According to the Formal Definition,

$$G_1 = (\{S, A\}, \{a, b\}, \{S \rightarrow Abb, A \rightarrow \varepsilon, A \rightarrow aA, A \rightarrow bA\}, S).$$

- Derivation of Strings of terminals, for example, babb:
  $S \Rightarrow Abb \Rightarrow bAbb \Rightarrow baAbb \Rightarrow babb$.
  In short, $S \Rightarrow^* babb$. [From start symbol to the string]
- We can check in another way that this string can be derived with the given grammar: $babb \Rightarrow baAbb \Rightarrow bAbb \Rightarrow Abb \Rightarrow S$. [From the string to the start symbol; With these two types of derivation, we can relate top-down and bottom-up approaches.]

- We derive similarly, $S \Rightarrow^* bb$, $S \Rightarrow^* abb$, $S \Rightarrow^* bbb$, $S \Rightarrow^* aabb$, $S \Rightarrow^* abbb$, $S \Rightarrow^* babb$, $S \Rightarrow^* bbbb$, $S \Rightarrow^* aaabb$, …
- G_1 'generates' {bb, abb, bbb, aabb, abbb, babb, bbbb, aaabb, …}, we denote it by $L(G_1)$, which is said to be a *Context Free Language (CFL),* as it is generated by a CFG.

In a production of a CFG, substitution of 'A' within '$\alpha_1 A \alpha_2$' does not depend on α_1 and α_2, that is, context α_1-α_2, and so, these types of grammars are called *context-free* grammars.

Describing a CFL by CFG:

Let's consider the language of Palindromes, L_{pal}. A palindrome is a string that reads the same forward and backward, such at 'otto'. To make things simple, we shall consider describing only the palindromes with alphabet {0, 1}. This language includes strings like 0110, 11011, and ε, but not 011 or 0101.

- A few obvious strings are in L_{pal} - ε, 0 and 1.
- If a string is a palindrome, it must begin and end with the same symbol or when the first and last symbols are removed, the resulting string must be a palindrome. So, if w is a palindrome, so are 0w0 and 1w1.

Now, representing the language of palindrome by the variable, *P* – the grammar describing P might be:

$$P \rightarrow \varepsilon$$
$$P \rightarrow 0$$
$$P \rightarrow 1$$
$$P \rightarrow 0P0$$
$$P \rightarrow 1P1$$