leftmost derivation

Canonical name	LeftmostDerivation
Date of creation	2013-03-22 18:54:50
Last modified on	2013-03-22 18:54:50
Owner	CWoo (3771)
Last modified by	CWoo (3771)
Numerical id	7
Author	CWoo (3771)
Entry type	Definition
Classification	msc 68Q45
Classification	msc 68Q42
Defines	rightmost derivation

Let $G = (\Sigma, N, P, \sigma)$ be a context-free grammar. Recall that a word v is generated by G if it is

- a word over the set $T := \Sigma - N$ of terminals, and
- derivable from the starting symbol σ.

The second condition simply says that there is a finite sequence of derivation steps, starting from σ, and ending at v:

$$\sigma = v_0 \to v_1 \to \cdots \to v_n = v$$

For each derivation step $v_i \to v_{i+1}$, there is a production $X \to w$ in P such that

$$\begin{aligned} v_i &= aXb, \\ v_{i+1} &= awb. \end{aligned}$$

Note that X is a non-terminal (an element of N), and a, b, w are words over Σ.

Definition. Using the notations above, if X is the leftmost non-terminal occurring in v_i, then we say the derivation step $v_i \to v_{i+1}$ is *leftmost*. Dually, $v_i \to v_{i+1}$ is *rightmost* if X is the rightmost non-terminal occurring in v_i.

Equivalently, $v_i \to v_{i+1}$ is leftmost (or rightmost) if a (or b) is a word over T.

A derivation is said to be *leftmost* (or *rightmost*) if each of its derivation steps is leftmost (or rightmost).

Example. Let G be the grammar consisting of a, b as terminals, σ, X, Y, Z as non-terminals (with σ as the starting symbol), and $\sigma \to XY$, $X \to a$, $Y \to b$, $\sigma \to ZY$, and $Z \to X\sigma$ as productions. G is clearly context-free.

The word $a^2b^2 = aabb$ can be generated by G by the following three derivations:

1. $\sigma \to ZY \to X\sigma Y \to XXYY \to XaYY \to XaYb \to Xabb \to aabb$,
2. $\sigma \to ZY \to X\sigma Y \to a\sigma Y \to aXYY \to aaYY \to aabY \to aabb$,
3. $\sigma \to ZY \to Zb \to X\sigma b \to XXYb \to XXbb \to Xabb \to aabb$.

The second is leftmost, the third is rightmost, and the first is neither.

Note that in every derivation, the first and last derivation steps are always leftmost and rightmost.

Remarks.

- One of the main properties of a context-free grammar G is that every word generated by G is derivable by a leftmost (correspondingly rightmost) derivation, which can be used to show that for every context-free language L, there is a pushdown automaton accepting every word in L, and conversely that the set of words accepted by a pushdown automaton is context-free.

- Leftmost derivations may be defined for any arbitrary formal grammar G satisfying the condition

  > (*) no terminal symbols occur on the left side of any production.

  Given two words u, v, we define $u \Rightarrow_L v$ if there is a production $A \rightarrow B$ such that $u = xAy$ and $v = xBy$ such that x is a terminal word. By taking the reflexive transitive closure of $\Rightarrow_L$, we have the leftmost derivation $\Rightarrow_L^*$. It can be shown that for any grammar G satisfying condition (*), the language $L_L(G)$ consisting of terminal words generated by G via leftmost derivations is always context-free.

References

[1] S. Ginsburg *The Mathematical Theory of Context-Free Languages*, McGraw-Hill, New York (1966).

[2] D. C. Kozen *Automata and Computability*, Springer, New York (1997).