Equivalent definitions to $L(G)$

→ Needed to talk about / This is to discuss other properties of grammars & CFGs in particular

★ Assume discussion applied to CFGs in particular, but could be generalized to any grammar. ★

Left & right-most derivations

Impose an order on the replacement of variables in in the intermediate forms (sometimes called sentential forms). Desirable: Predictability.

CFG G : $S \to SS \mid aSb \mid bSa \mid a \mid b \mid \varepsilon$

Right-most derivation of abba :

$S \to SS \to SbSa \to Sb\varepsilon a \to aSbba \to a\varepsilon bba \to abba$

whenever there's a choice replace the rightmost variable (in CFGs so this is well-defined)

Left-most derivation of abba :

$S \to SS \to aSbS \to a\varepsilon bS \to abbSa \to abb\varepsilon a$
$= abba$

whenever there's a choice replace leftmost variable

Def (n-step ~~left-most~~ derivation rel.) Given a CFG $G = (V, S, T, P)$, $\alpha, \beta \in (V \cup T)^*$, $n \in \mathbb{N}$
$\alpha \xrightarrow[G\ lm]{n} \beta$ if there are n left-most derivation steps that allow you to derive β from α.

Formally: $\alpha \xrightarrow[lm]{0} \alpha \quad \forall \alpha \in (V \cup T)^*$

Drop G for convenience when know what G is.

$\alpha \xrightarrow[lm]{1} \beta$ if $\exists\, x \in T^*, A \in V, \alpha_1, \gamma \in (V \cup T)^*$
s.t. $\alpha = x A \alpha_1$,
$A \to \gamma \in P$,
$\beta = x \gamma \alpha_1$

$\alpha \xrightarrow[lm]{n+1} \beta$ if $\exists\, \gamma \in (V \cup T)^*$ s.t.
$\alpha \xrightarrow[lm]{n} \gamma$, $\gamma \xrightarrow[lm]{1} \beta$.

Def (*-step left-most derivation rel.) Given CFG $G = (V, S, T, P)$, $\alpha, \beta \in (V \cup T)^*$, $\alpha \xrightarrow[lm]{*} \beta$ if $\exists\, n \in \mathbb{N}$ s.t. $\alpha \xrightarrow[lm]{n} \beta$.

The definition of $\xrightarrow[rm]{n}$, $\xrightarrow[rm]{*}$ is similar.

Thm Given CFG $G = (V, S, T, P)$, $w \in T^*$, the following 3 statements are equivalent:

1) $S \overset{*}{\underset{G}{\Rightarrow}} w$
2) $S \overset{*}{\underset{G\,lm}{\Rightarrow}} w$
3) $S \overset{*}{\underset{G\,rm}{\Rightarrow}} w$

Implications: Restricting the order in which productions of CFGs are applied does not restrict their expressiveness.

Pf 2) ⇒ 1) A left-most derivation is a $*$-step derivation.

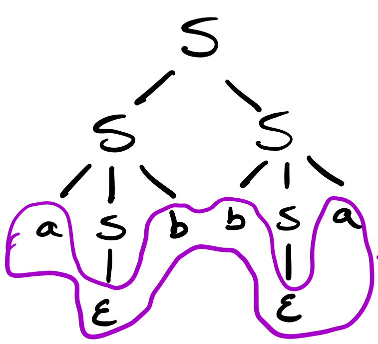
2) ⇐ 1) Trickier. Easier proof version exists where show left-most derivation from a parse tree which we will talk about next.

Parse trees

Rooted tree data structure used to represent grammar derivation.

CFG G: $S \to SS \mid aSb \mid bSa \mid a \mid b \mid \varepsilon$

$w = abba$



→ Yield of the parse tree is $a \varepsilon b b \varepsilon a = abba$.

Def (Parse tree) Given a CFG $G = (V, S, T, P)$, the parse tree of G is a rooted tree T_G where:

1. The root of T is S
2. Every interior node is labeled w/ a variable from V
3. Every leaf is either a terminal or ε.
   (*)[If ε is a leaf, then it must have no siblings => Otherwise could artificially blow up a parse tree w/ $\overset{|}{\varepsilon}$]
   & its yield would be equivalent
4. If an interior node is labeled with A & its children are labeled (from left-to-right) $X_1, X_2, \ldots, X_k$ then $A \to X_1 \ldots X_k \in P$.

Thm Given a CFG $G = (V, S, T, P)$ & $w \in T^*$

$S \xrightarrow[G]{*} w$ <=> G has a parse tree with yield w

Pf
- Map parse tree to some derivation sequence & vice-versa. Really need induction arg. on # nodes & # steps.
- Fyi: Once prove this can show how to produce lm derivation for any w from T proving previous thm.