Let $G = (V, \overset{T}{\Sigma}, \overset{P}{R}, S)$ be a CFG. Σ or T: terminals; V: Variables or nonterminals

Let X, Y be strings over $\underline{V \cup \Sigma}$, ie. $X, Y \in (V \cup \Sigma)^*$.

Defn: $X \underset{G}{\Rightarrow} Y$ (read: "derives in one step of G")

if there are strings $U, W \in (\Sigma \cup V)^*$ and a variable $A \in V$ and a string $Z \in (\Sigma \cup V)^*$ such that:

- $A \to Z$ is a rule in R
- $X = UAW$
- $Y = UZW$

} ie. the variable A got substituted by the "rhs" Z.

Also define $X \underset{G}{\Rightarrow}^0 X$ ("X can derive itself in 0 steps")

$X \underset{G}{\Rightarrow}^k Z$ if there is a Y such that $X \underset{G}{\Rightarrow}^{k-1} Y$, $Y \underset{G}{\Rightarrow} Z$

Finally, $X \underset{G}{\Rightarrow}^* Z$ if for some $k \geq 0$, $X \underset{G}{\Rightarrow}^k Z$.

$$\boxed{L(G) =_{def} \{x \in \Sigma^* : S \underset{G}{\Rightarrow}^* x\}.}$$

And if $X \in (\Sigma \cup V)^*$ is such that $S \underset{G}{\Rightarrow}^* X$, then X is called a sentential form.

eg. $S \Rightarrow \langle NP \rangle \langle VP \rangle$. Is this the only form? "Go Figure!"

Alt. defn: $S \underset{G}{\Rightarrow}^* X$ iff there are $X_0, X_1, X_2, \dots, X_{k-1}, X_k$ st. $X_k = X$, $X_0 = S$, and for $1 \leq i \leq k$, $X_{i-1} \underset{G}{\Rightarrow} X_i$. Then $(S, X_1, \dots, X_k)$ is a Derivation

②

Key Defn: A derivation $(X_0, \ldots, X_k)$ is leftmost if for every step $X_{i-1} \underset{G}{\Longrightarrow} X_i$, writing

$$X_{i-1} = UAW \qquad \text{with the rule } A \to Z \in R,$$
$$X_i = UZW$$

we have $U \in \Sigma^*$, so that A was the leftmost variable in X_{i-1}.

Example: G: $S \to SS \mid (S) \mid \varepsilon \quad V = \{S\},\ \Sigma = \{`(', `)'\}$.

not leftmost — ditto

$$\underline{S} \Rightarrow \underline{S}S \Rightarrow (S)\underline{S} \Rightarrow (S)(\underline{S}) \Rightarrow (\underline{S})() \Rightarrow ()()$$

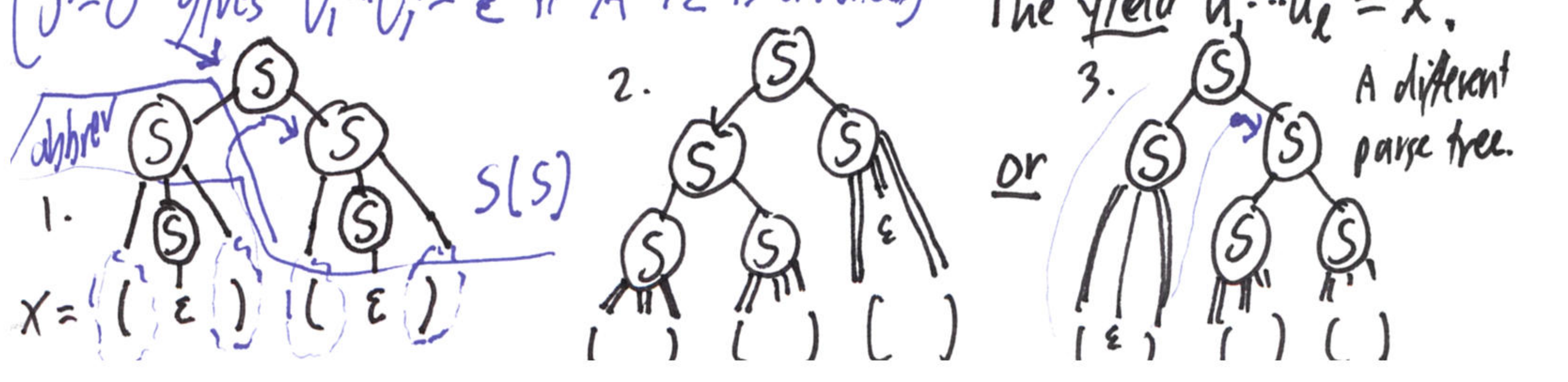
$S \to (S)$ commits to "just one" nested string.

$S \to \underbrace{SS}_{\text{which one did I expand??}} \to SSS$ commits to 3 or more. [unless you do $S \to \varepsilon$]

eg $\overset{⑥}{\Rightarrow} ()()()$

Defn: A parse tree for a string $x \in L(G)$ [can re-define: or a partial parse for a sentential form X] has a root labeled S and leaves labeled by $u_1 \cdots u_\ell \in \Sigma^*$, and for every interior node labeled $A \in V$ with children $U_1, \ldots, U_j \in (\Sigma \cup V)$, $A \to U_1 \cdots U_j$ is a rule in R.

($j = 0$ gives $U_1 \cdots U_j = \varepsilon$ if $A \to \varepsilon$ is a rule.) The yield $u_1 \cdots u_\ell = X$.



<u>Key Fact</u>: Given a parse tree, we can read off a unique leftmost derivation by expanding the tree in left to right pre-order.

1. $S \Rightarrow \underline{S}S \Rightarrow (\underline{S})S \Rightarrow ()\underline{S} \Rightarrow ()(S) \Rightarrow ()()$

2. $S \Rightarrow \underline{S}S \Rightarrow \underline{S}SS \Rightarrow^2 ()\underline{S}S \Rightarrow^2 ()()\underline{S} \Rightarrow^2 ()()()$.
   $\Rightarrow ()(S)S$

3. $S \Rightarrow \underline{S}S \Rightarrow (\underline{S})S \Rightarrow ()S \Rightarrow ()\underline{S}S \Rightarrow^2 ()()\underline{S} \Rightarrow^2 ()()()$

∴ The string $()()()$ has two different <u>leftmost</u> (LM) derivations that we got from the two different parse trees for it.

Remark: $()()$ does have the other LM derivation $S \Rightarrow SS$
$SS \Rightarrow SSS \Rightarrow (S)SS \Rightarrow ()SS \Rightarrow ()(S)S \Rightarrow ()()S \xRightarrow{\text{via } S \to \varepsilon} ()()$

But, this doesn't happen in G': $S \to SS \mid (S) \mid ()$. $\varepsilon \notin L(G')$.

<u>Defn</u>: A string $x \in \Sigma^*$ is <u>ambiguous</u> in a CFG G if x has two different parse trees, equivalently, LM derivations. If any $x \in L(G)$ is ambiguous, then G is <u>ambiguous</u>, else <u>unambiguous</u>.

<u>Example</u>: $()()$ is ambiguous in G but not in G'.

<u>But</u>, "$()()()$" is ambiguous in G', so G' is also an <u>ambiguous grammar</u>. Study: Is G'' unambiguous?

$G'' = S \Rightarrow () \mid \underline{(S)S}$ How about G''': $S \to \varepsilon \mid (S)S$?

Important Example of Disambiguation (Text simplifies it)

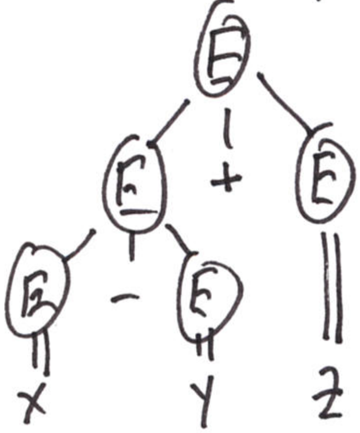
G_1: $E \to \langle\underline{const}\rangle \mid \langle\underline{var}\rangle \mid E+E \mid E-E \mid E\cdot E \mid E/E \mid (E)$.

Problem X − Y + Z is ambiguous

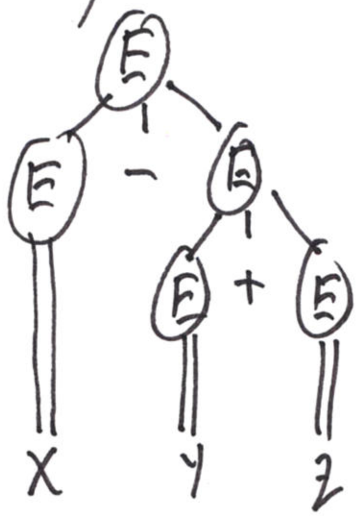
How to Disambiguate?

We could force fully parenthesized expressions or change to Postfix notation, but...

$\boxed{\text{Use Hierarchy}}$



groups as (X − Y) + Z



groups as X − (Y + Z) which gives an unintended value.

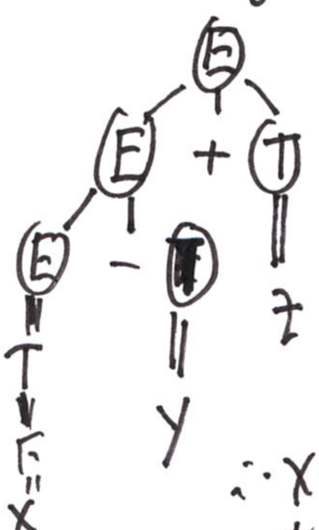
$E \to \overset{E-T}{E+T} \mid T$ "An *expression* can be a single *term* or can be an expression *plus* a *term*"

$T \to F \mid T*F \mid T/F$

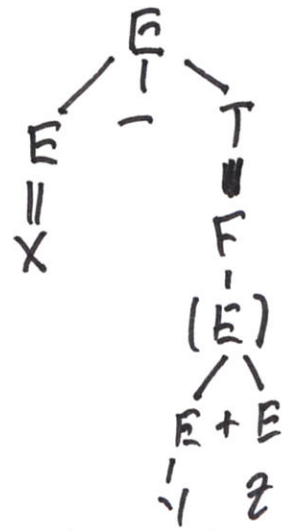
$\boxed{\text{FACT (not proved in text): } G_2 \text{ is unambiguous.}}$

$F \to \underline{const} \mid \underline{var} \mid (E)$

A factor can be a const or var or any expr in parens.



∴ X − Y + Z grouped as (X − Y) + Z



yields X − (Y + Z)

different string!

Student: How about X/Y*Z?