Let $G = (V, \Sigma, R, S)$ be a CFG. Let $X, Y$ be strings over $V \cup \Sigma$, i.e. $X, Y \in (V \cup \Sigma)^*$.

**Defn:** $X \xrightarrow{G} Y$ (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:

- $A \rightarrow Z$ is a rule in $R$
- $X = UA W$ (i.e. the variable $A$ got substituted by the "rhs" $Z$)
- $Y = UZW$

Also define $X \xrightarrow{G}^0 X$ ("$X$ can derive itself in 0 steps") $X \xrightarrow{G}^k Z$ if there is a $Y$ such that $X \xrightarrow{G}^k Y$ and $Y \xrightarrow{G} Z$.

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

$L(G) = \{ x \in \Sigma^* : S \xrightarrow{G}^* x \}$. And if $X \in (\Sigma \cup V)^*$ is such that $S \xrightarrow{G} X$, then $X$ is called a sentential form.

eg. $S \xrightarrow{G} (NP) (VP)$. Is this the only form? "Go Figure!"

Alt. defn: $S \xrightarrow{G}^* X$ iff there are $X_0, X_1, X_2, \ldots, X_{k-1}, X_k$ st. $X_k = X$, $X_0 = S$, and for $1 \leq i \leq k$, $X_{i-1} \xrightarrow{G} X_i$. Then $(S, X_1, \ldots, X_k)$ is a derivation.
**Key Defn:** A derivation \((X_0, \ldots, X_k)\) is **leftmost** if for every step \(X_{i-1} \xrightarrow{a} X_i\), writing
\[
X_{i-1} = UAW \quad \text{with the rule} \quad A \rightarrow Z \in R, \\
X_i = UZW
\]
we have \(U \in \Sigma^*\), so that \(A\) was the *leftmost* variable in \(X_{i-1}\).

**Example:** \(S \rightarrow SS | (S) \in V = \{S, \Sigma\}, \Sigma = \{i, (, )\}\).
\[
S \rightarrow SS \rightarrow (S)S \rightarrow (S)(S) \rightarrow (S)(1) \rightarrow (1) (1)
\]
\(S \rightarrow (S)\) commits to just one nested string.

\(S \rightarrow SS \rightarrow SSS\) commits to 3 or more. [unless you]
\[
\text{do } S \rightarrow \varepsilon
\]
\(x \rightarrow \varepsilon (1)(1)(1)\) can re-define a partial parse for a different form \(x\).

**Defn:** A **parse tree** for a string \(x \in L(G)\) has a root labeled \(S\) and leaves labeled by \(u_i \cdots u_n \in \Sigma^*\), and for every internal node labeled \(A \in V\) with children \(U_1, \ldots, U_j \in (\Sigma U V)\), \(A \rightarrow U_1 \cdots U_j\) is a rule in \(R\). (if \(A \rightarrow \varepsilon\) is a rule)

\[
(\varepsilon = 0 \text{ gives } U_i \cdots U_j \rightarrow \varepsilon)
\]

The yield \(u_i \cdots u_n = x\).

\[ X = \left( \varepsilon \right) \]

A different parse tree.
Key Fact: Given a parse tree, we can read off a unique leftmost derivation by expanding the tree in left-to-right order.

1. $S \Rightarrow SS \Rightarrow SSS \Rightarrow^2 (S)SS \Rightarrow^2 (S)(S) \Rightarrow^2 (S)(S)(C) \Rightarrow (S)(S)(C)$
2. $S \Rightarrow SS \Rightarrow (S)S \Rightarrow (S)(S) \Rightarrow (S)(S) \Rightarrow^2 (S)(S)(C) \Rightarrow (S)(S)(C)$

The string $(S)(C)$ has two different leftmost derivations that we got from the two different parse trees for it.

Remark: $(S)(C)$ does have the other LM derivation $S \Rightarrow SS \Rightarrow SSS \Rightarrow (S)SS \Rightarrow (S)(S) \Rightarrow (S)(S) \Rightarrow (S)(S)(C)$

But this doesn’t happen in $S \Rightarrow SS \Rightarrow (S)(C)$.

Defn: A string $x \in \Sigma^*$ is ambiguous 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 ambiguous, else unambiguous.

Example: $(S)(C)$ is ambiguous in $G$ but not in $G'$.

But, $(S)(S)(C)$ is ambiguous in $G'$, so $G'$ is also an ambiguous grammar. Study: Is $G'' = S \Rightarrow \varepsilon | (S)(S)$ unambiguous?
Important Example of Disambiguation (Text simplifications)

\[ G_1 = E \rightarrow \langle \text{const} \rangle | \langle \text{var} \rangle | E + E | E - E | E \cdot E | E / E | (E) \]

Problem: \( x - y + z \) is ambiguous

How to Disambiguate?

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

Use Hierarchy

which gives an unintended value.

E \rightarrow E + T | T

T \rightarrow F | T * F | T / F

F \rightarrow \text{const} | \langle \text{var} \rangle | (E)

E \rightarrow E + T | T

E \rightarrow E - T | T

E \rightarrow E \cdot T | T

E \rightarrow E / T | T

E \rightarrow (E)

FACT (not proved in text): \( G_2 \) is unambiguous

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

E \rightarrow (E)

Yields

\[ x - (y + z) \]

different string!

x / y \cdot z:

Study: How about

x / y \cdot z?