Describe the set of arithmetic expressions with correctly matched parenthesis.

Not regular. \( F = \{ i^i \mid i \geq 0 \} \)

\[ u = (^{i \cdot i}) \quad v = (^{i + i}) \]

\( w = 0 \). \( uvw \) has correctly matched parenthesis
\( vw \) does not.

Ignore numbers, variables, operations, etc.

**Definition**: \( \epsilon \) has correctly matched parentheses.

- If \( u \) and \( v \) have correctly matched parentheses, then \( uv \) also has correctly matched parentheses.
- If \( u \) has correctly matched parentheses, then \( (u) \) also has correctly matched parentheses.

Let \( M \) denote the set of correctly matched strings.

**CFG**: \[
M \rightarrow \epsilon \quad G = (\{ M \}, \{ 0, 1 \}, \{ M \rightarrow \epsilon, M \rightarrow MM, M \rightarrow (M) \}, \epsilon, M)
\]

**L_{min} = \{ 0^n 1^n \mid n \geq 0 \}**. Not regular.

\[
\begin{align*}
S & \rightarrow \epsilon \\
& \rightarrow 0S1 \\
S & \rightarrow e | 0S1
\end{align*}
\]

\[
G = (\{ S \}, \{ 0, 1 \}, \{ S \rightarrow \epsilon, S \rightarrow 0S1 \}, S)
\]
\[ S \rightarrow NP \ VP \]
\[ NP \rightarrow CN | CN PP \]
\[ VP \rightarrow CV | CV NP \]
\[ PP \rightarrow P7 CN \]
\[ CN \rightarrow A7 \ N \]
\[ CV \rightarrow V | V NP \]
\[ A7 \rightarrow a | the \]
\[ N \rightarrow boy | girl | flower | bat \]
\[ V \rightarrow hit | sees | likes \]
\[ P7 \rightarrow with \]

The girl hits the boy with the bat.
Lecture 5: Context-Free Languages

Date: September 5, 2023.

Context-free Grammar (CFG) is $G = (N, \Sigma, P, S)$ where

1. $N$ is a finite set whose elements are called variables/non-terminals
2. $\Sigma$ is a finite set whose elements are called terminals
3. $P$ is a finite set of production rules of the form $A \to w$, where $A \in N$ and $w \in (\Sigma \cup N)^*$
4. $S \in N$ is the starting non-terminal

For such a grammar we have the following definitions.

1. For $x, y, z \in (\Sigma \cup N)^*$ and $A \in N$, we say $xAz \xrightarrow{1} xyz$ ("$xyz$ can be derived from $xAz$ in one step") if $A \to y$ is a rule of $G$.
2. For $x, z \in (\Sigma \cup N)^*$ we say $x \xrightarrow{G} z$ ("$z$ can be derived from $x$ in zero or more steps") if either $x = z$,
or there exists $y \in (\Sigma \cup N)^*$ such that $x \xrightarrow{G} y$ and $y \xrightarrow{G} z$ (inductively).
3. For $w \in (\Sigma \cup N)^*$, $L(w) = \{ x \in \Sigma^* \mid w \xrightarrow{G} x \}$.
4. The language defined by grammar $G$ is $L(G) = L(S)$.
5. A language $L \subseteq \Sigma^*$ is context-free if there is a grammar $G$ such that $L = L(G)$.

Parse Tree for a string $w \in \Sigma^*$ in grammar $G$ is a labeled, rooted, ordered tree where

- Each leaf is labeled by either $\epsilon$ or a symbol in $\Sigma$. Concatenating these in order from left to right yields
  the string $w$.
- Each internal node is labeled with a non-terminal.
- If an interior node is labeled by $A$ with children labeled by $X_1, X_2, \ldots, X_k$ (in order from left to right),
  then $A \to X_1X_2\cdots X_k$ must be a rule in $G$.

Proposition 1. Let $G = (N, \Sigma, P, S)$ be a CFG. A string $w$ has a parse tree with root labeled $A \in N$ iff
$A \xrightarrow{G} w$.

Ambiguity: A CFG $G = (N, \Sigma, P, S)$ is said to be ambiguous if there is a string $w \in \Sigma^*$ such that there
are two different parse trees for $w$ with root labeled $S$.

$G$ is unambiguous if it is not ambiguous.
Problem 1. Consider the CFG $G = \{S, Y, \{a, b\}, P, S\}$ where $P$ is the set of rules

$$
S \rightarrow aSb \, | \, bY \, | \, Ya \\
Y \rightarrow bY \, | \, aY \, | \, \epsilon
$$

1. Which of the following strings are in $L(G)$: $aabb$, $aabbb$, $aabaab$?

2. Draw a parse tree for the string $aabaab$ with root labeled $S$.

Problem 2. $G_{\text{exp}} = \{(E, I, N), \{a, b, 0, 1, (,), +, *, -, \}, P, E\}$ where $P$ is

$$
E \rightarrow I \, | \, N \, | \, E + E \, | \, E \cdot E \, | \, (E) \\
I \rightarrow a \, | \, b \, | \, Ia \, | \, Ib \\
N \rightarrow 0 \, | \, 1 \, | \, N0 \, | \, N1 \, | \, -N \, | \, +N
$$

Draw a parse tree for $a + b * a$. 

\[ \text{Parse tree for } a + b * a \]
Problem 3. Design a CFG for the language \( \{ a^i b^j \mid j \geq 2i \} \).

\[ G = ( \Sigma, \Gamma, S, P, \epsilon) \]

\[ S \rightarrow \epsilon \mid aSbb \mid Sb \]

\[ A = \{ a^i b^j \mid j \geq 2i \} \]

\[ w \in L(G) \iff w \in A \]

\[ S \xrightarrow{G}^* w \Rightarrow w \in A \] (proof by induction on \# derivation steps)

\[ w \in A \Rightarrow S \xrightarrow{G}^* w \] (proof by induction on \#w\)
Proposition 2. Let $A, B \subseteq \Sigma^*$ be context-free languages and let $h: \Sigma \rightarrow \Gamma^*$ be a homomorphism. Then the following languages are also context free: $A \cup B$, $AB$, $A^*$, and $h(A)$.

$$G_1 = (N_1, \Sigma, P_1, S_1) \quad G_2 = (N_2, \Sigma, P_2, S_2)$$

$$A = L(G_1) \quad B = L(G_2)$$

$A \cup B$:

$$G_{A \cup B} = (N_1 \cup N_2 \cup \Sigma^*, \Sigma, P, S) \mid N_1 \cap N_2 = \emptyset$$

$$P = \begin{cases} S \rightarrow S_1 & S \rightarrow S_2 \end{cases} \cup P_1 \cup P_2.$$

$AB$:

$$G_{AB} = (N_1 \cup N_2 \cup \Sigma^*, \Sigma, P, S) \mid \Sigma, \emptyset \cap N_2 = \emptyset.$$

$$P = \begin{cases} S \rightarrow S_1 S_2 \end{cases} \cup P_1 \cup P_2.$$

$A^*$:

$$S \rightarrow \varepsilon \mid S_1 S$$

$h(A)$:

$$G_{h(A)} = (N, \cup \Sigma^* X_a \mid a \in \Sigma, \Pi, P, S_1)$$

$$P : \text{Rules from } P_1 \text{ where } a \text{ is replaced by the non-terminal } X_a$$

and $\not\exists a \in \Sigma$.

$$X_a \rightarrow h(a).$$