Describe the set of arithmetic expressions with correctly matched parenthesis.

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

$u = (^i, \; v = (^j \quad (i \neq j)$

$w = 0)^i$. $\quad uw$ ~~...~~ has correctly matched parenthesis

vw does not.

Ignore numbers, variables, operations, etc.

<u>Definition</u>: ϵ 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. $\begin{cases} M \xrightarrow{\Rightarrow} \epsilon \\ \quad \rightarrow MM \\ M \rightarrow (M) \end{cases}$

$G = (\{M\}, \{(,)\}, \{M \to \epsilon, M \to MM, M \to (M)\}, M)$

$L_{0^n1^n} = \{0^n 1^n \mid n \geq 0\}$. not regular.

$G = (\{S\}, \{0,1\}, \{S \to \epsilon, S \to 0S1\}, S)$

$S \rightarrow \epsilon$

$\quad \rightarrow 0S1$

$S \rightarrow \epsilon \mid 0S1$

<S> → <NP> <VP>

<NP> → <CN> | <CN><PP>

<VP> → <CV> | <CV><PP>

<PP> → <P> <CN>

<CN> → <A> <N>

<CV> → <V> | <V><NP>

<A> → a | the

<N> → boy | girl | flower | bat

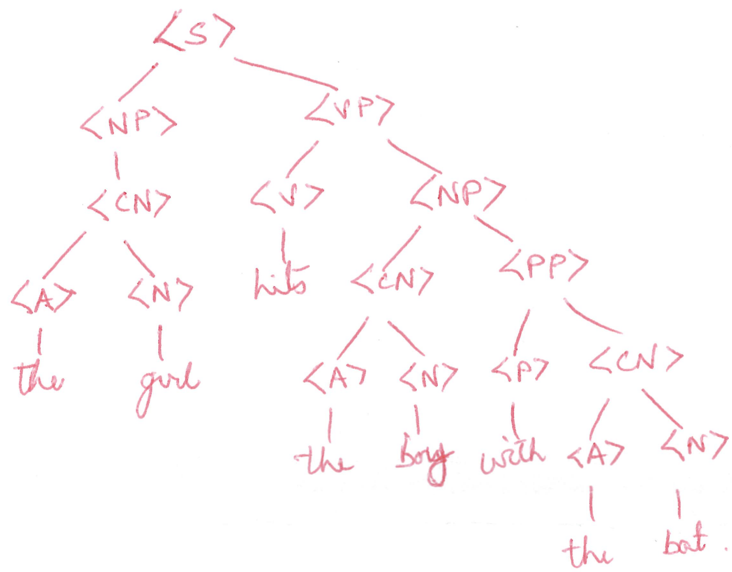
<V> → hits | sees | likes

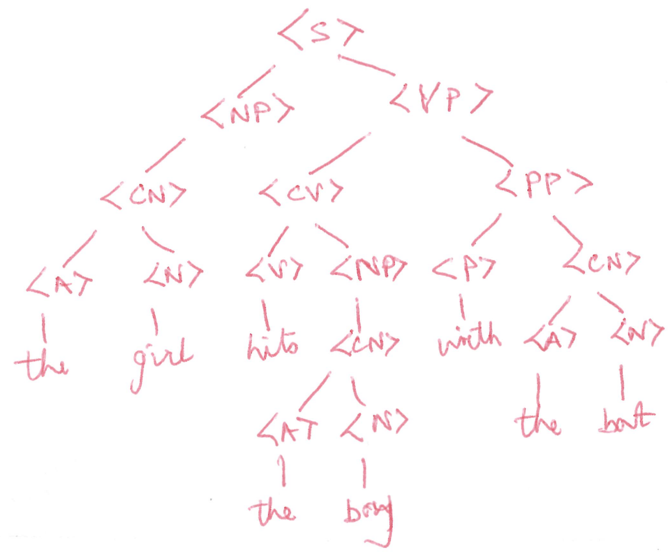
<P> → with

$\overbrace{\underbrace{\text{a}}_{A}\ \underbrace{\text{boy}}_{N}}^{NP}\ \overbrace{\underbrace{\text{sees}}_{V}}^{VP}$

a boy sees a flower

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

- N is a finite set whose elements are called **variables/non-terminals**
- Σ is a finite set whose elements are called **terminals**
- P is a finite set of **production rules** of the form $A \to w$, where $A \in N$ and $w \in (\Sigma \cup N)^*$
- $S \in N$ 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 $xAz \xrightarrow[G]{1} xyz$ ("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 $x \xrightarrow[G]{1} y$ and $y \xrightarrow[G]{*} z$ (inductively).
3. For $w \in (\Sigma \cup N)^*$, $\mathbf{L}(w) = \{x \in \Sigma^* \mid w \xrightarrow[G]{*} x\}$.
4. The **language** defined by grammar G is $\mathbf{L}(G) = \mathbf{L}(S)$.
5. A language $L \subseteq \Sigma^*$ is **context-free** if there is a grammar G such that $L = \mathbf{L}(G)$.

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

- Each leaf is labeled by either ϵ or a symbol in Σ. Concatenating these in order from left to right yields the string w.

~~→ Root is labeled S.~~

- 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 $A \to X_1 X_2 \cdots X_k$ 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

$$S \to aSb \mid bY \mid Ya$$
$$Y \to bY \mid aY \mid \epsilon$$

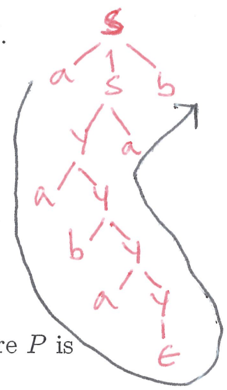
1. Which of the following strings are in $\mathbf{L}(G)$: $aabb$, $aabbb$, $aabaab$?

$S \xrightarrow[G]{1} aSb \xrightarrow[G]{1} aaSbb \xrightarrow[G]{\not *} aabb \notin L(G)$

$S \xrightarrow[G]{1} aSb \xrightarrow[G]{1} aaSbb \xrightarrow[G]{1} aabYbb \xrightarrow[G]{1} aabbb \in L(G)$

$S \xrightarrow[G]{1} aSb \xrightarrow[G]{1} aYab \xrightarrow[G]{1} aaYab \xrightarrow[G]{1} aabYab \xrightarrow[G]{1} aabaYab$
$\xrightarrow[G]{1} aabaab \in L(G)$

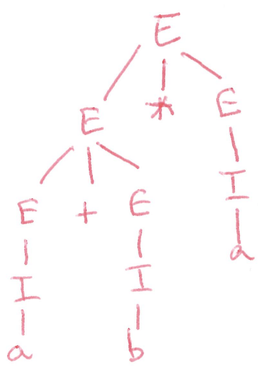
2. Draw a parse tree for the string $aabaab$ with root labeled S.

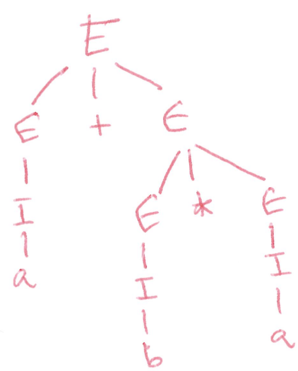


Problem 2. $G_{\text{exp}} = (\{E, I, N\}, \{a, b, 0, 1, (,), +, *, -\}, P, E)$ where P is

$$E \to I \mid N \mid E + E \mid E * E \mid (E)$$
$$I \to a \mid b \mid Ia \mid Ib$$
$$N \to 0 \mid 1 \mid N0 \mid N1 \mid -N \mid +N$$

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





Problem 3. Design a CFG for the language $\{a^i b^j \mid j \geq 2i\}$.

$\ni abb$

$\not\ni aab$

$\ni b$

$G = (\{S\}, \{a, b\}, P, S)$

$S \to \epsilon \mid aSbb \mid Sb$

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

$w \in L(G) \iff w \in A$

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

$w \in A \implies S \xrightarrow[G]{*} w$ (proof by induction on $|w|$)

Proposition 2. *Let $A, B \subseteq \Sigma^*$ be context-free languages and let $h : \Sigma \to \Gamma^*$ be a homomorphism. Then the following languages are also context free: $A \cup B$, AB, A^*, and $h(A)$.*

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

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

$A \cup B$: $G_{A \cup B} = (N_1 \cup N_2 \cup \{S\}, \Sigma, P, S)$ $\Big|$ $N_1 \cap N_2 = \emptyset$

$P = \{S \to S_1 \mid S_2\} \cup P_1 \cup P_2$.

AB : $G_{AB} = (N_1 \cup N_2 \cup \{S\}, \Sigma, P, S)$ $\Big|$ $N_1 \cap N_2 = \emptyset$.

$P = \{S \to S_1 S_2\} \cup P_1 \cup P_2$

A^* : $S \to \epsilon \mid S_1 S$

$h(A)$: $G_{h(A)} = (N_1 \cup \{X_a \mid a \in \Sigma\}, \Gamma, P, S_1)$

P : Rules from P_1 where a is replaced by the non-terminal X_a

and $\forall a \in \Sigma$.

$X_a \longrightarrow h(a)$.