A context-free grammar (CFG) is an object $G = (V, \Sigma, R, S)$ where

- $V$ is a finite alphabet of variables, aka nonterminals.
- $\Sigma$, also called $I$, is a finite alphabet of terminals.
- We suppose $V \cap \Sigma = \emptyset$.
- $S$, a member of $V$, is the starting variable.
- $R$ is a finite set of rules of the form $A \rightarrow X$ where $A \in V$ and $X \in (\Sigma \cup V)^*$. Really $R \subseteq V \times (\Sigma \cup V)^*$. Also called $P$ for production.

**Defn:** Given $X, Y \in (\Sigma \cup V)^*$, we write $X \Rightarrow_Y Y$ "$X$ derives $Y$ in one step of $G$" if $X$ can be broken as $X = U \cdot A \cdot W$ and there is a rule $A \rightarrow Z$ in $R$ s.t. $Y = U \cdot Z \cdot W$. 


Also define $X \Rightarrow X$ for any $X \in (\Sigma \cup V)^*$ and for $k \geq 1$, $X \Rightarrow^k Z$ if there is a $Y$ in $(\Sigma \cup V)^*$ st. $X \Rightarrow^{k-1} Y$ and $Y \Rightarrow Z$.

i.e., $X$ derives $Z$ in $K$ steps of the grammar $G$.

Finally:

$X \Rightarrow^* Z$ if $X \Rightarrow^k Z$ for some $k \geq 0$.

$L(G) = \{ w \in \Sigma^* : S \Rightarrow^* w \}$

However:

A string $X \in (\Sigma \cup V)^*$ is called a Sentential Form of $G$ if $S \Rightarrow^* X$.

Origin: Noam Chomsky, 1956-57

Syntactic Structures:

$S \Rightarrow N \rightarrow V$

A sentence can be a Noun Verb Phrase Phrase

Variables are often written as "tokens" <noun_phrase> as in <verb_phrase> BNF
Examples: \( L = \text{PAL} = \{ x \in \{0,1\}^* : x = x^R \} \)

Compare to final grammar in Tue 3/7 lecture.

"or" for rule options. \( 01^R = 10 \quad 0110^R = 0110 \)

Understood: \( V = \Sigma \) \( \Sigma = \{0,1\} \), rules and

\( G = S \rightarrow \varepsilon | 0S0 | 1S1 \)

Is \( L(G) \subseteq L ? \) "Is \( G \) sound for \( L ? \)"

Is \( L(G) \supseteq L ? \) "Is \( G \) comprehensive?"

Is \( G \) correct? \( L(G) = L \).

Is this \( G \) sound? Let's try some derivations:

\( S \rightarrow \varepsilon . \ \varepsilon \in \Sigma \). Is \( \varepsilon \in L ? \) \( \checkmark \) yes

\( S \rightarrow 0S0 \rightarrow 0\varepsilon 0 = 00 \in \Sigma \). Is \( 00 \in \text{PAL} ? \) \( \checkmark \) yes

Similarly, \( S \rightarrow 1S1 \rightarrow 11 \in \text{PAL} \).

\( S \rightarrow 0S0 \rightarrow 0 \cdot 1S1 \cdot 0 \rightarrow 0 \cdot 11 \cdot 0 = 0110 \).

Is \( G \) comprehensive for \( \text{PAL} ? \) No: \( G \) cannot derive the palindromes 010, 101, 000, 111, nor 0 or 1.

The cruelest way to show this is to see that \( G \) obeys a stronger soundness condition: every \( x \in \Sigma \) it derives into an even-length palindrome.
In the intuition that the lead clause of a noun phrase must be a (and the 'i' in the rule $S \rightarrow S \iota S_2$)
and hence derivable in $G_i$, it is easier to prove derivability in $G_2$.
Hence $G_2$ is also sound.

Soundness of $G_2$ does imply soundness of $G_1$ because $G_2$ can simulate the rule $S \rightarrow (S \iota S_2) S_2$ via its derivations.

\[ \Sigma = \{ \langle i', i' \rangle \} \]

\[ S \iota S \iota S_2 \]
This is ingrained in the example of (Fully-Parenthesized) Formulas:

\[ E \rightarrow \text{constant} \mid \text{variable} \mid (E) \]

The start symbol, "Expression" can be a numeric constant with digits and \( \alpha \) as terminals, and identifier letters as variables not grammar variables.

\[ E \rightarrow (E+E) \mid (E-E) \mid (E*E) \mid (E/E) \]

The interpretation of \( E \) as "I am a legal expression" shows this grammar is sound.

It does not, however, cover all expressions we tend to write using precedence in place of parentheses: \((x + y*2)\)

However, CFGs can convey precedence as well as balance.