CSE 396 Lecture Thu 3/19 Spr 2017

A context-free grammar (CFG) is a 4-tuple $G = (V, \Sigma, R, S)$ where:

V is a finite alphabet of <u>variables</u> aka <u>nonterminals</u>
$(\Sigma \cap V = \emptyset)$

Σ, also called T, is a finite alph. of <u>terminals</u>

S, a member of V, is the <u>start symbol</u>.

R is a finite set of <u>rules</u> (also called <u>productions</u>) of the form

$$A \rightarrow W \quad \text{where } A \in V,\ W \in (\Sigma \cup V)^*.$$

<u>Defⁿ</u>: $uAv \underset{G}{\Rightarrow} uWv$ "uAv can <u>derive</u> uWv in one step of G."

Given two strings $X, Y \in (\Sigma \cup V)^*$, write $X \underset{G}{\Rightarrow}^* Y$ if there are $k \geq 0$ and strings $X_1, X_2, \ldots, X_k \in (\Sigma \cup V)^*$ s.t. $X \underset{G}{\Rightarrow} X_1 \underset{G}{\Rightarrow} \cdots \underset{G}{\Rightarrow} X_k = Y$

__Defn__: A string $X \in (\Sigma \cup V)^*$ is called a __sentential__ __form__ of the CFG G if $S \overset{*}{\underset{G}{\Rightarrow}} X$.

$$\underline{\underline{L(G)}} = \{ x \in \underline{\Sigma^*} : S \overset{*}{\underset{G}{\Rightarrow}} x \}.$$

A language L is a context-free language (CFL) if there is a CFG G s.t. $L = L(G)$.

$$S \overset{::=}{\rightarrow} \langle \text{NounPhrase} \rangle \langle \text{VerbPhrase} \rangle$$

$$S \rightarrow \langle NP \rangle \langle VP \rangle$$

$$\langle NP \rangle \rightarrow \langle \text{Noun} \rangle \mid \langle \text{Adjective} \rangle \langle NP \rangle$$

$$\langle \text{Noun} \rangle \rightarrow \{\text{any } \underline{\text{noun}} \text{ in the dictionary}\}$$

$$\langle \text{Adjective} \rangle \rightarrow \underline{\text{adjective}}.$$

$$\langle \text{Adverb} \rangle \rightarrow \underline{\text{adv}}$$

$$\langle \text{verb} \rangle \rightarrow \underline{\text{verb}}$$

$$\langle VP \rangle \rightarrow \langle \text{verb} \rangle \mid \langle \text{Verb} \rangle \langle \text{Adverb} \rangle.$$

$$
\begin{aligned}
S &\Rightarrow \langle NP \rangle \langle VP \rangle \\
&\Rightarrow \langle \text{Adjective} \rangle \langle NP \rangle \langle VP \rangle \\
&\Rightarrow \langle \text{Adjective} \rangle \langle \text{Adjective} \rangle \langle NP \rangle \langle VP \rangle \\
&\Rightarrow \langle \text{Adj} \rangle \langle \text{Adj} \rangle \langle \text{Noun} \rangle \langle VP \rangle \\
&\Rightarrow \langle \text{Adj} \rangle \langle \text{Adj} \rangle \langle \text{Noun} \rangle \langle \text{Verb} \rangle \langle \text{Adverb} \rangle \\
&\Rightarrow^5 \text{Colorless green ideas sleep furiously.}
\end{aligned}
$$

not leftmost

French, for the most part, has

$\langle NP \rangle \Rightarrow \langle Noun \rangle \langle Adj \rangle$

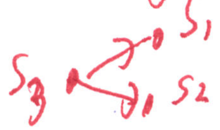
<u>Contrast</u>: $\langle AdjStream \rangle \Rightarrow \varepsilon \mid \langle Adj \rangle \langle AdjStream \rangle$] Abstract pattern for a list* $\langle Adj \rangle$

with $\langle Adj \rangle \to \underline{adj} \mid adj\ \underline{et}\ \langle Adj \rangle$

CFGs can simulate the regular ops.

$G_1 = (V_1, \Sigma, R_1, S_1)$, $G_2 = (V_2, \Sigma, R_2, S_2)$ build G_3.

- $S_3 \to S_1 \mid S_2$ (rest as in R_1, R_2) $\therefore L(G_3) = L(G_1) \cup L(G_2)$



- $S_3 \to S_1 S_2$ (rest as in $R_1 \cup R_2$) $\therefore L(G_3) = L(G_1) \cdot L(G_2)$
- $S_3 \to \varepsilon \mid S_1 S_3$ $\therefore L(G_3) = L(G_1)^*$.

$S_3 \Rightarrow S_1 S_3 \Rightarrow S_1 S_1 S_3 \Rightarrow S_1 S_1 S_1 S_3 \Rightarrow S_1 S_1 S_1$ etc.

Basis: $(\{S\}, \Sigma, \emptyset, S)$ generates $\emptyset$ $(\{S\}, \Sigma, \{S \to \varepsilon\}, S)$ generates $\{\varepsilon\}$

for any $c \in \Sigma$, $(\{S\}, \Sigma, \{S \to c\}, S)$ generates $\{c\}$

$\therefore$ Thm $\boxed{\mathcal{REG} \subseteq \mathcal{CFL}}$ ie. Every regular language is a <u>CFL</u>.