CLASSIFICATION OF GRAMMARS

07/01/2008

E7.1

Classification of grammars into 4 groups, depending on the form of the productions:

- grammars of type 0 : no limitations
- —— " —— 1 : context-sensitive
- —— " —— 2 : context-free
- —— " —— 3 : regular (or right linear)

<u>Definition</u> : grammar of type 0

Productions have the most general form $\alpha \to \beta$, with $\alpha \in V^* \cdot V_N \cdot V^*$ and $\beta \in V^*$

A language generated by a grammar of type 0 is called of type 0.

<u>Example</u>: grammar for $L = \{0^n 1^m \mid m, n \in \mathbb{N} \;\&\; n < m\}$

$G = (\{S\}, \{0,1\}, P, S)$

with P : 1) $S \longrightarrow 0S1$ generates $\underbrace{0 \cdots 0}_{k\text{ times}} S \underbrace{1 \cdots 1}_{k\text{ times}}$

2) $0S \longrightarrow S$ removes a 0 on the left of S

3) $S \longrightarrow 1$ generates the final 1

Derivation of 001111 :

$\underline{S} \overset{1}{\Longrightarrow} 0\underline{S}1 \overset{1}{\Longrightarrow} 00\underline{S}11 \overset{1}{\Longrightarrow} 00\underline{0S}111 \overset{2}{\Longrightarrow} 00\underline{S}111 \overset{3}{\Longrightarrow} 001111$

Derivation of 111 :

$\underline{S} \overset{1}{\Longrightarrow} 0\underline{S}1 \overset{1}{\Longrightarrow} 0\underline{0S}11 \overset{2}{\Longrightarrow} \underline{0S}11 \overset{2}{\Longrightarrow} \underline{S}11 \overset{3}{\Longrightarrow} 111$

Note: 1) the application of productions could go on forever (e.g. rule 1 in the previous example)

2) grammars of type 0 allow for derivations that shorten the sentential form (see for instance the previous derivations)

<u>Definition</u>: grammar of type 1 (context-sensitive)

Productions have the form $\alpha \to \beta$, with $\alpha \in V^* \cdot V_N \cdot V^*$, $\beta \in V^+$, and $|\alpha| \leq |\beta|$

These productions cannot shorten the length of the sentential form to which they are applied

A language generated by a grammar of type 1 is called of type 1 or context-sensitive

<u>Example</u>: grammar for $L = \{a^n b^n c^n \mid n \geq 1\}$

$G = (\{S, A, B, C\}, \{a, b, c\}, P, S)$

with P:

$$\left.\begin{array}{ll} 1) & S \to aSBC \\ 2) & S \to aBC \end{array}\right\} \text{generate } \underbrace{aa\cdots a}_{k\text{ times}}\underbrace{BCBC\cdots BC}_{k\text{ times}}$$

3) $CB \to BC$ } moves the C's to the end

4) $aB \to ab$
5) $bB \to bb$
6) $bC \to bc$
7) $cC \to cc$

} generate the terminals from left to right

Note: we cannot simply have $B \to b$ and $C \to c$ because this would generate many words not in L

Example (con't)

Derivation of aaabbbccc:

$S \overset{1}{\Rightarrow} a\underline{S}BC \overset{1}{\Rightarrow} aa\underline{S}BCBC \overset{2}{\Rightarrow} aaaBCB\underline{CB}C \overset{3}{\Rightarrow} aaaB\underline{CB}BCC$

$\overset{3}{\Rightarrow} aaaBB\underline{CB}CC \overset{3}{\Rightarrow} aa\underline{aB}BBCCC \overset{4}{\Rightarrow} aaa\underline{bB}BCCC$

$\overset{5}{\Rightarrow} aaab\underline{bB}CCC \overset{5}{\Rightarrow} aaabb\underline{bC}CC \overset{6}{\Rightarrow} aaabbb\underline{cC}C$

$\overset{7}{\Rightarrow} aaabbbc\underline{cC} \overset{7}{\Rightarrow} aaabbbccc$

Note: not each sequence of direct derivations leads to a sentence in $\mathcal{L}(G)$

$S \Rightarrow a\underline{S}BC \Rightarrow aa\underline{S}BCBC \Rightarrow aaaBCB\underline{CB}C \Rightarrow aa\underline{aB}CBBCC$

$\Rightarrow aaa\underline{bC}BBCC \Rightarrow \underbrace{aaabcBBCC}_{\text{we cannot apply any other production}}$

Definition: grammar of type 2 (context-free)

Productions have the form $A \to \beta$, with $A \in V_N$ and $\beta \in V^+$

These productions are productions of type 1, with the additional requirement that on the left there is a single nonterminal

A language generated by a grammar of type 2 is called of type 2 or context-free

Example: grammar for $L = \{0^n 1^m \mid m, n \in \mathbb{N} \ \& \ n < m\}$

$G = (\{S, A, B\}, \{0, 1\}, P, S)$

with P:
1) $S \longrightarrow A \mid B$
2) $A \longrightarrow 0A1 \mid B$ generates $\underbrace{0\cdots0}_{k\text{ times}}\underbrace{1\cdots1}_{l\text{ times}}$ $k < l$
3) $B \longrightarrow 1 \mid 1B$ generates $\underbrace{1\cdots1}_{k\text{ times}}$ $k > 1$

<u>Example</u>: grammar for $L = \{0^m 1^n 1^n 0^m \mid m,n \in \mathbb{N} \;\&\; m+n > 0\}$

$G = (\{S,A\}, \{0,1\}, P, S)$

with P: 1) $S \to 00 \mid 0S0 \mid A$ generates $\underbrace{0\cdots0}_{k \text{ times}}\underbrace{1\cdots1}_{l \text{ times}}\underbrace{0\cdots0}_{k \text{ times}}$ l even, possibly $l=0$ or $k=0$ but not both

2) $A \to 11 \mid 1A1$ generates $\underbrace{1\cdots1}_{k \text{ times}}$ $k \geq 2$ & k even

<u>Definition</u>: grammar of type 3 (regular, right linear)

Productions have the form $A \to \delta$, with $A \in V_N$

$\delta \in V_T \cup (V_T \cdot V_N)$

(i.e. $A \to aB$ or $A \to a$ with $A, B \in V_N$ and $a \in V_T$)

A language generated by a grammar of type 3 is called of type 3 or regular

<u>Example</u>: $\{a^n b \mid n \geq 0\}$ is of type 3, since it is generated by the grammar $S \to aS \mid b$

Note: a grammar of type 3 is called <u>linear</u> because on the righthand side of a production there is at most one nonterminal; it is called <u>right</u>-linear because the nonterminal is on the right of the terminal

<u>Exercise</u>: Show that grammars of type 3 generate the class of regular languages that do not contain ε

Solution to the exercise: (E7.5)

Given a grammar $G = (V_N, V_T, P, S)$ we construct an NFA

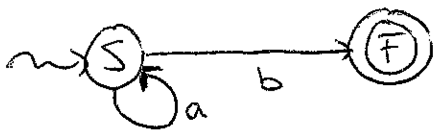
$A_G = (V_N \cup \{F\}, V_T, \delta, S, \{F\})$

with $B \in \delta(A, a)$ iff $A \rightarrow aB$

$F \in \delta(A, a)$ iff $A \longrightarrow a$

Note that A_G is constructed in such a way that $w \in \mathcal{L}(A_G)$ if and only if $w \in \mathcal{L}(G)$

Example: If $G = (\{S\}, \{a, b\}, \{S \rightarrow aS \mid b\}, S)$ then A_G look as follows



Conversely, given an NFA $A = (Q, \Sigma, \delta, q_0, F)$ we construct a grammar

$G_A = (Q, \Sigma, P, q_0)$

with $p \longrightarrow aq \in P$ iff $q \in \delta(p, a)$

$p \longrightarrow a \in P$ iff $q \in \delta(p, a)$ and $q \in F$

Note that G_A is constructed in such a way that $w \in \mathcal{L}(G_A)$ if and only if $w \in \mathcal{L}(A)$

Example: If $A = (\{q_0, q_1\}, \{a, b\}, \delta, q_0, \{q_1\})$ with $\delta(q_0, a) = \{q_0\}$, $\delta(q_0, b) = \{q_1\}$, and $\delta(q_1, a) = \delta(q_1, b) = \emptyset$ then A_G has the productions $q_0 \longrightarrow aq_0 \mid bq_1 \mid b$

Note: We are never able to produce a sentence out of this sentential form