NOȚIUNEA DE GRAMATICĂ

Def: Se numește gramatică un sistem $G=(V_N, V_T, x_0, P)$ unde

V_N – alf. simbolurilor neterminale $\quad V_N \cap V_T = \phi$

V_T – alf. sb. terminale $\quad V_G \overset{def}{=} V_N \cup V_T$

$x_0 \in V_N$ – simbol de start

$P \subset V_G^* V_N V_G^* \times V_G^* \quad |P| < \infty$

P mulțimea de reguli

$(u,v) \in P$ notată $u \to v \in P$ sau $u \to v$ regulă, producție

Definim modul de producere a cuvintelor prin folosirea regulilor.

Def: Se numește "derivare directă" următoarea relație peste V_G^*, notăm $\underset{G}{\Rightarrow}$

$$p \Rightarrow q \text{ dd } \begin{matrix} p = xuy \\ q = xvy \end{matrix} \text{ și } u \to v \in P \quad , x, y \in V_G^*$$

Obs: $u \in V_G^* V_N V_G^*$ – șir arbitrar de terminale și neterminale ce conține cel puțin un neterminal

$v \in V_G^*$ – șir arbitrar (cuvânt peste universul lingvistic peste V_G)

Def: Se consideră închiderea tranzitivă și reflexivă a relației de derivare directă, notată $\overset{*}{\Rightarrow}$, numită derivare (fără mențiunea directă!)

$$p \overset{*}{\Rightarrow} q \overset{def}{\Longleftrightarrow} \exists\, p_1, p_2, \ldots, p_k \text{ a.î. } p = p_1 \Rightarrow p_2 \Rightarrow \ldots \Rightarrow p_k = q \quad , k \geq 1.$$

$k=1$ reflexivitate $p=q$.

Se numește lungime a derivării nr. de derivări directe aplicate în derivare, $l_p = k-1$

Pentru $k \geq 2$ notăm $p \overset{+}{\Rightarrow} q$.

Def: Se numește limbaj generat de gramatica $G=(V_N, V_T, x, P)$ mulțimea

$$\boxed{L(G) = \{ p \in V_T^* \mid x_0 \underset{G}{\overset{*}{\Rightarrow}} p \}}$$

Ex: $G = (\{A,B\}, \{0,1\}, A, \{A \to 0A, A \to 1B, B \to 1\}) \quad L(G) = ?$

$A \to 0A \mid 1B$

$A \Rightarrow 1\underline{B} \Rightarrow 11 \in L(G) \quad A \overset{*}{\Rightarrow} 11$

formă propozițională

$L(G) = \{11, 011, 0011, \ldots, 0^k 11, \ldots\}$

$= \{0^k 11 \mid k \geq 0\}$

$A \Rightarrow 0\underline{A} \Rightarrow 01\underline{B} \Rightarrow 011 \in V_T^* \quad A \overset{*}{\Rightarrow} 011 \in L(G)$

$A \Rightarrow 0\underline{A} \Rightarrow 00A \Rightarrow 001B \Rightarrow 0011 \in L(G)$

$A \Rightarrow 0A \Rightarrow 00A \Rightarrow 000A \Rightarrow \ldots \Rightarrow 0^k A \Rightarrow 0^k 1B \Rightarrow 0^k 11$

Ex: $G_1 = (\{A\}, \{a,b\}, A, \{A \to a \mid aa \mid ba \mid baba\}) \quad L(G_1) = ?$

$A \Rightarrow a \in L(G_1)$

$A \Rightarrow baba \in L(G_1)$

$A \Rightarrow ba \in L(G_1)$

$A \Rightarrow aa \in L(G_1)$

$L(G_1) = \{a, aa, ba, baba\}$

Ex: $G_2 = (\{S\}, \{a,b\}, S, \{S \to ab \mid aSb\})$

$S \Rightarrow ab \in L(G_2)$

$S \Rightarrow a\underline{S}b \Rightarrow aabb \in L(G_2)$

$S \Rightarrow a\underline{S}b \Rightarrow aa\underline{S}bb \Rightarrow aaabbb \in L(G_2)$

$S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow a^3Sb^3 \Rightarrow \ldots \Rightarrow a^kSb^k \Rightarrow a^k ab b^k = a^{k+1}b^{k+1}$

$$L(G_2) = \{a^n b^n \mid n \geq 1\}$$

Ex: $G_3 = (\{S\}, \{0,1\}, S, \{S \to 0S \mid 11\})$

$S \Rightarrow 11 \in L(G_3)$

$S \Rightarrow 0S \Rightarrow 011 \in L(G_3)$

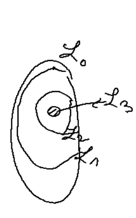
$S \Rightarrow 0S \Rightarrow 00S \Rightarrow \ldots \Rightarrow 0^n S \Rightarrow 0^n 11$

$$L(G_3) = \{0^n 11 \mid n \geq 0\}$$

Obs: Două gramatici ce produc același limbaj se zic "echivalente"

De ex. $L(G) = L(G_3)$ notăm $G \simeq G_3$.

CLASIFICAREA CHOMSKY A GRAMATICILOR



$\mathcal{G}_0$ – tip 0, fără restricții asupra formei regulilor

$\mathcal{G}_1$ – tip 1, gr. dependente de context.

$uAv \to upv$, unde $A \in V_N$, $u, p, v \in V_G^*$, $p \neq \lambda$

se admite și regula $S \to \lambda$ dar $S \notin dr.$

$\mathcal{G}_2$ – tip 2, gr. independente de context

$A \to p$, $A \in V_N$, $p \in V_G^*$

$\mathcal{G}_3$ – tip 3, gr. regulate

$$\begin{cases} A \to pB \\ C \to q \end{cases} , \begin{matrix} A, B, C \in V_N \\ p, q \in V_T^* \end{matrix} \quad \underline{\text{sau echiv}} \quad \begin{cases} A \to Bp \\ C \to q \end{cases} , A, B, C \in V_N, p, q \in V_T^*$$

Tipul gramaticii se consideră cel mai restrictiv tip din clasificarea CHOMSKY

Def: Notăm $\mathcal{L}_i$ = familia limbajelor generate de gramatici de tipul i, $i = \overline{0,3}$

$\mathcal{L}_0$ – lb. de tip 0 $\quad \mathcal{L}_1$ – lb. dependente de context $\quad \mathcal{L}_2$ – lb. independente de context

$\mathcal{L}_3$ – lb. regulate

Def: Se numește IERARHIA CHOMSKY a limbajelor urm. șir de incluziuni

$$\mathcal{L}_3 \subsetneq \mathcal{L}_2 \subsetneq \mathcal{L}_1 \subsetneq \mathcal{L}_0$$

Propoziție: $\forall G \in \mathcal{G}_2$ $\exists G' \in \mathcal{G}_2$ echivalentă cu proprietatea că simpurele regulii care conțin simboluri terminale sunt de forma "$A \to i$", $A \in V_N$, $i \in V_T$.

(Lema "$A \to i$" la lb. de tip 2, mu curs)

#

Fie $G = (V_N, V_T, x_0, P) \in \mathcal{G}_2$ și $L = L(G)$.

Construim o gramatică echivalentă $G' = (V_N \cup \{X_a \mid a \in V_T\}, V_T, x_0, P')$

Construim regulile lui P' astfel

- păstrăm în P' regulile $u \to v \in P$ care conțin numai neterminale în st. dr $u, v \in V_N^*$
- dacă $u \to v \in P$ conține în stânga sau dreapta terminale introduc în P' regula $u' \to v' \in P'$ unde u' și v' se obțin din u și v prin înlocuirea terminalelor a cu X_a.
- adaug regulile $\{X_a \to a \mid a \in V_T\}$

Observăm că G' respectă condiția "$A \to i$".

Rămâne de arătat că $L(G') = L \overset{?}{=} L(G)$.

$L(G) \subset L(G')$

Fie $p \in L(G)$ $p = a_1 \dots a_k \in V_T^*$ și $x_0 \overset{*}{\underset{G}{\Rightarrow}} p$ derivare ce produce p.

Considerăm urm. derivare în gram G':

G': $x_0 \overset{*}{\underset{G'}{\Rightarrow}} X_{a_1} X_{a_2} \dots X_{a_k} \overset{*}{\underset{G'}{\Rightarrow}} a_1 a_2 \dots a_k$ folosind doar regulile nou introduse

Deci $p \in L(G')$, q.e.d.

În Clasificarea Chomsky a limbajelor (gramaticilor) se pot considera definiții echivalente ptr. clasele de gramatici!