TC - CURS 8

SIMPLIFICAREA GRAMATICILOR INDEPENDENTE DE CONTEXT

1. Eliminarea simbolurilor și a producțiilor nefolositoare („useless")

<u>Def.</u> Fie $G=(N, \Sigma, S, P)$ o g.i.c

Un neterminal $A \in N$ se numește folositor dacă și numai dacă există $w \in L(G)$ astfel încât:

$$S \stackrel{*}{\Rightarrow} xAy \stackrel{*}{\Rightarrow} w, \quad x, y \in (N \cup \Sigma)^*$$

Altfel, spunem că A este un neterminal nefolositor. Toate producțiile care îl $\overbrace{\text{pe}}^{\text{conțin}}$ A în membrul stâng sau drept sunt, de asemenea nefolositoare, deoarece nu intervin în derivarea niciunui șir din L(G).

<u>Exemplu</u> 1)

$$G_1 \quad \begin{aligned} S &\to aS \mid A \\ A &\to aA \mid \lambda \\ B &\to bA \end{aligned}$$

Neterminalul B este nefolositor și deci și producția $B \to bA$ este nefolositoare (inutilizabilă) deoarece nu există nicio derivare în această gramatică pt care $S \stackrel{*}{\Rightarrow} xBy$

2) G_2
$$S \to aSb \mid aA$$
$$A \to aA \mid bB \mid \lambda$$
$$B \to bBb$$

În acest exemplu, în G_2 avem o derivare de forma $S \stackrel{*}{\Rightarrow} xBy$, însă nu există nicio derivare $B \stackrel{*}{\Rightarrow} z$, $z \in \Sigma^*$, astfel că B nu intervine în derivarea niciunui șir din $L(G_2)$. Atunci, din G_2 se va elimina atât B, cât și toate producțiile în care B apare în membrul stâng sau drept, respectiv $A \to bB$, $B \to bBb$.

<u>Teorema 1</u> Fie $G = (N, \Sigma, S, P)$ o g.i.c. Există o gramatică G' care nu conține neterminali sau producții nefolositoare echivalentă cu G (adică $L(G) = L(G')$)

<u>Dem.</u> Vom construi $G' = (N', \Sigma, S, P')$ utilizând următorul algoritm.

<u>Pas 1</u> $N_1 \leftarrow \phi$

<u>Pas 2</u> Repetă următorul pas până când nu se mai adaugă la N_1 noi neterminali:

□ Pentru fiecare $A \in N$ pentru care există $A \to z_1 \dots z_m \in P$, $z_1, \dots, z_m \in \Sigma \cup N_1$, $A \notin N_1$, adaugă A la N_1

Neterminalii din N_1 sunt generativi

Astfel, N_1 conține doar neterminalii A din N pentru care în G există derivarea $A \overset{*}{\underset{G}{\Rightarrow}} w$, $w \in \Sigma^*$.

Evident, $N_1 \subseteq N$. În particular, dacă $S \notin N_1$, atunci rezultă că pentru niciun $w \in \Sigma^*$ nu avem $S \overset{*}{\Rightarrow} w$, deci $L(G) = \emptyset$

Pasul 3. Dacă $S \notin N_1$, STOP

Pasul 4 $N_2 \leftarrow \{S\}$.

Pasul 5 Repetă până nu se mai adaugă noi neterminali la N_2:

□ Pentru fiecare $A \rightarrow x_0 A_1 x_1 \dots A_n x_n \in P$, $n \geq 1$, $x_0, x_1, \dots, x_n \in \Sigma^*$, $A \in N_2$, $A_1, \dots, A_n \in N$:

□ Pentru fiecare $i \in \{1, \dots, n\}$, dacă $A_i \notin N_2$ adaugă A_i la N_2.

Observăm că un neterminal A din N este adăugat la N_2 ddacă în G există derivarea $S \overset{*}{\Rightarrow} xAy$, $x, y \in (N \cup \Sigma)^*$.

Fie $G' = (N', \Sigma, S, P')$, unde $N' = N_1 \cap N_2 \subseteq N$, $P' = \{X_0 \rightarrow X_1 \dots X_n \in P \mid X_0 \in N', X_1, \dots, X_n \in \Sigma \cup N'\}$

Observăm că G' nu conține simboluri inutilizabile, și, deoarece $N' \subseteq N$, $P' \subseteq P$, rezultă că $L(G') \subseteq L(G)$.

Reciproc, dacă $w \in L(G)$ atunci în G există o derivare de forma

$$S \underset{G}{\Rightarrow} w_2 \underset{G}{\Rightarrow} w_2 \underset{G}{\Rightarrow} \dots \underset{G}{\Rightarrow} w_n = w$$

În conformitate cu algoritmul de mai sus, toți neterminalii care intervin în această derivare sunt incluşi atât în N_1, cât şi în N_2, deci sunt incluşi în N', inclusiv S.

Mai mult decât atât, producţiile care s-au aplicat la fiecare pas al acestei derivări sunt incluse în P'. Rezultă că avem $S \underset{G'}{\Rightarrow} w_1 \underset{G'}{\Rightarrow} \dots \underset{G'}{\Rightarrow} w_n = w$,

adică $w \in L(G')$.

În concluzie $L(G') = L(G)$, G' fiind o gramatică fără simboluri inutilizabile.

<u>Exemplu</u> 3 Eliminarea simbolurilor nefolositoare din gramatica

$$G_3 \quad \begin{aligned} S &\to AB \mid CA \\ B &\to BC \mid AB \\ A &\to a \\ C &\to AB \mid b \end{aligned}$$

Aplicând 1-2 obţinem $N_1 = \{A, C, S\}$ şi aplicând 4-5 obţinem $N_2 = N_1 = \{A, C, S\}$ deci $N' = \{S, A, C\}$, $P' = \{S \to CA,\ A \to a,\ C \to b\}$

<u>Exemplu 4</u> $G = (\{S, A, B, C\}, \{a, b\}, S, P)$

unde P:
$$\begin{aligned} S &\to aS \mid A \mid C \\ A &\to a \\ B &\to aa \\ C &\to aCb \end{aligned}$$

Obținem $N_1 = \{S, A, B\}$, $N_2 = \{S, A, C\}$

Atunci $N' = \{S, A\}$, $P' = \{S \to aS, S \to A, A \to a\}$

2. Eliminarea λ-producțiilor

Dacă cuvântul vid, λ, nu aparține limbajului generat de o gramatică independentă de context (g.i.c), atunci toate λ-producțiile (de forma $A \to \lambda$) pot fi eliminate.

Dacă cuvântul vid aparține limbajului generat de o g.i.c, atunci vom putea elimina toate λ-producțiile cu excepția $S \to \lambda$, necesară pt a introduce pe λ în limbajul gramaticii.

Def. Dacă într-o gramatică avem

$$A \overset{*}{\Rightarrow} \lambda, \ A \text{ neterminal},$$

spunem că A este anulabil

Teorema 2 Fie $G = (N, \Sigma, S, P)$ g.i.c. Atunci există $G_1 = (N_1, \Sigma, S_1, P_1)$ g.i.c. fără λ-producții ai $L(G_1) = L(G) - \{\lambda\}$.

Dem Vom construi mai întâi mulțimea $N_a \subseteq N$ a tuturor neterminalilor din N care sunt anulabili

1. $N_a \leftarrow \{A \mid A \to \lambda \in P\}$
2. Repetă până când nu se mai adaugă neterminali noi:

□ Dacă $A \to A_1 \cdots A_m \in P$, $m \geq 1$, $A_1, \ldots, A_m \in N_a$ și $A \notin N_a$, adaugă A la N_a.

Observăm că un neterminal N_a este introdus în N_a ddacă este anulabil.

Dacă $S \notin N_a$, atunci $S_1 = S$, $N_1 = N$

Dacă $S \in N_a$, introducem un neterminal nou, S_1, astfel că $N_1 = N \cup \{S_1\}$

Pentru a obține P_1, procedăm în felul următor:

1. $P_1 = \{A \to \beta \in P \mid |\beta|_{N_a} = 0\}$
2. Pentru fiecare producție $B \to \beta_0 B_1 \beta_1 \cdots B_m \beta_m \in P$, cu $\beta_0, \beta_1, \ldots, \beta_n \in (\Sigma \cup (N - N_a))^*$, $B_1, \ldots, B_n \in N_a$, $n \geq 1$, adaugă la P_1 toate producțiile de forma: $B \to \beta_0 X_1 \beta_1 \cdots X_m \beta_n$, unde $X_i \in \{\lambda, B_i\}$, $i = 1, \ldots, n$.
Dacă $\beta_0 = \beta_1 = \ldots = \beta_n = \lambda$ atunci nu toți X_i pot fi λ (pt a nu introduce $B \to \lambda \in P_1$)

Observăm că în P_1 nu sunt introduse producții de forma $A \to \lambda$.

3. Dacă $S \in N_a$, atunci la P_1 adăugăm $S_1 \to S$.

Observăm că:

i) În G_1 nu există producții anulabile.

ii) Producțiile lui G_1 fie sunt producții din G care nu conțin în membrul drept (MD) neterminali anulabili; fie sunt producții care provin din producțiile lui G care au în MD cel puțin un neterminal anulabil, prin substituirea în toate modurile posibile a neterminalilor anulabili cu λ, mai puțin cazul în care s-ar adăuga o prod. de forma $A \to \lambda$

iii) Dacă $A \to \beta \in P_1$, atunci $A \overset{*}{\underset{G}{\Rightarrow}} \beta$

iv) din ii) și iii) rezultă că $L(G_1) \subseteq L(G) - \{\lambda\}$

v) Pentru orice $A \in N$, pentru orice derivare $A \overset{n}{\underset{G}{\Rightarrow}} z$, $z \in (N \cup \Sigma)^+$, vom avea $A \overset{*}{\underset{G_1}{\Rightarrow}} z$. Arătăm aceasta prin inducție după n.

<u>Baza</u> <u>$n=1$</u> Rezultă $A \to z \in P$, $z \neq \lambda$.

Atunci $A \to z \in P_1$, deci $A \underset{G_1}{\Rightarrow} z$.

<u>Ipoteza</u> Pp că $(\forall) A \in N$, $(\forall)$ $A \overset{m}{\underset{G}{\Rightarrow}} z$, $z \in (N \cup \Sigma)^+$, $m \leq n$, avem $A \overset{*}{\underset{G_1}{\Rightarrow}} z$

<u>Saltul inductiv</u> Fie $A \overset{n+1}{\underset{G}{\Longrightarrow}} z$, $z \in (N \cup \Sigma)^+$.

Punem în evidență primul pas:

$A \Rightarrow x_0 A_1 x_1 \dots A_k x_k \overset{n}{\Rightarrow} z$, unde

$A \to x_0 A_1 x_1 \dots A_k x_k \in P$, $x_0, x_1, \dots, x_k \in \Sigma^*$, $k \geq 1$.

Atunci $A \to x_0 A_1 \dots A_k x_k \in P_1$ și, totodată,

$z = x_0 \alpha_1 x_1 \dots \alpha_k x_k$, unde $A_i \overset{\leq n}{\underset{G}{\Longrightarrow}} \alpha_i$, $i = 1, \dots, n$.

Dacă $\alpha_i \neq \lambda$, atunci conform ipotezei de inducție $A_i \overset{*}{\underset{G_1}{\Longrightarrow}} \alpha_i$, $i = 1, \dots, n$.

Dacă $\alpha_i = \lambda$, rezultă că A_i anulabil.

Fie $i_1, \dots, i_t$ cu $1 \leq i_1, \dots, i_t \leq k$, toți indicii cu $\alpha_{i_1} \neq \lambda, \dots, \alpha_{i_t} \neq \lambda$. Rezultă că în P_1 avem producția:

$A \to x_0 x_1 \dots x_{i_1 - 1} A_{i_1} x_{i_1} \dots x_{i_t - 1} A_{i_t} x_{i_t} x_{i_t + 1} \dots x_k \in P_1$

$A_{i_1} \overset{*}{\underset{G_1}{\Longrightarrow}} \alpha_1, \dots, A_{i_k} \overset{*}{\underset{G_1}{\Longrightarrow}} \alpha_{i_k}$,

$z = x_0 x_1 \dots x_{i_1 - 1} x_{i_1} x_{i_1 + 1} \dots x_{i_t - 1} x_{i_t} \dots x_k$. Deci

$A \underset{G_1}{\Rightarrow} x_0 \dots x_{i_1 - 1} A_{i_1} x_{i_1} \dots x_{i_t - 1} A_{i_t} x_{i_t} \dots x_k$

$\overset{*}{\underset{G_1}{\Longrightarrow}} x_0 \dots x_{i_1 - 1} \alpha_{i_1} x_{i_1} \dots x_{i_t - 1} \alpha_{i_t} x_{i_t} \dots x_k = z$

adică $A \overset{*}{\underset{G_1}{\Longrightarrow}} z$

vi) Din (v), pt $S \in N$, $z \in \Sigma^+$ cu $S \overset{*}{\underset{G}{\Rightarrow}} z$, $z \neq \lambda$ rezultă $S \overset{*}{\underset{G_1}{\Rightarrow}} z$, adică $z \in L(G_1)$,

vii) deci $L(G) - \{\lambda\} \subseteq L(G_1)$

viii) În cazul în care $\lambda \in L(G)$, adică S este anulabil, la P_1 adăugăm producția $S_1 \to \lambda$; obținem astfel $G_1' = (N \cup \{S_1\}, \Sigma, S_1, P_1 \cup \{S_1 \to \lambda\})$, cu $L(G_1') = L(G)$, iar în G_1' unica λ-producție este $S_1 \to \lambda$, iar S_1 nu mai apare în membrul drept al niciunei producții.

End Teorema 2

<u>Exemplu 3</u>

G_3 $\quad S \to aSbS \mid bSaS \mid \lambda$

$N_a = \{S\}$

G_1: $S_1 \to S$ (S_1 simbol de start)

$S \to aSbS \mid abS \mid aSb \mid ab$

$S \to bSaS \mid baS \mid bSa \mid ba$

G_1' La producțiile lui G_1 se adaugă $S_1 \to \lambda$

3) Eliminarea redenumirilor

Def. Fie $G = (N, \Sigma, S, P)$ g.i.c. Numim redenumire o producție de forma
$$A \to B, \quad A, B \in N.$$

Teorema 3 Fie G o g.i.c. Atunci există G' g.i.c. fără redenumiri cu $L(G') = L(G)$.

Dem Fie $G = (N, \Sigma, S, P)$ g.i.c.

Construim $G' = (N, \Sigma, S, P')$ g.i.c fără redenumiri astfel.

1. $P' \leftarrow P - \{A \to A \mid A \in N\}$
2. Cât timp există $A \to B \in P'$
   2.1. Șterge $A \to B$ din P'
   2.2 Pentru orice $B \to \alpha \in P$, $\alpha \neq A$, adaugă $A \to \alpha$ la P' (dacă nu era deja în P')

Observații

(i) algoritmul de mai sus se termină întotdeauna, deoarece există un număr finit de producții în P, un nr. finit de producții ce pot fi introduse în P' la fiecare pas

(ii) Se poate demonstra prin inducție structurală că $L(G') = L(G)$

Exercițiu

Exemplu
$$S \to TU \mid V$$
$$T \to aTb \mid \lambda$$
$$U \to cU \mid \lambda$$
$$V \to aVc \mid W$$
$$W \to bW \mid \lambda$$

Eliminare simboluri nefolositoare: nu sunt

Eliminare λ-producții:

Neterminali anulabili

$\{T, U, W\}$ $\{T, U, W, S, W\}$

Noile producții

$S_1 \to S \mid \lambda$

$S \to TU \mid T \mid U \mid V$

$T \to aTb \mid ab$

$U \to cU \mid c$

$V \to aVc \mid ac \mid W$

$W \to bW \mid b$

Eliminare redenumiri

$S_1 \to S \mid \lambda$

$S \to TU \mid aTb \mid ab \mid cU \mid c \mid aVc \mid ac \mid W$

(1) $T \to aTb \mid ab$

$U \to cU \mid c$

$V \to aVc \mid ac \mid W$

$W \to bW \mid b$

$S_1 \to S \mid \lambda$; $S \to TU \mid aTb \mid ab \mid cU \mid c \mid aVc \mid ac \mid bW \mid b$

(2) $T \to aTb \mid ab$; $U \to cU \mid c$

$V \to aVc \mid ac \mid bW \mid b$; $W \to bW \mid b$

<u>Corolar</u> Orice g.i.c. G este echivalentă cu o g.i.c. G' fără redenumiri, fără λ-producții și fără simboluri inutilizabile, ai

$$L(G') = L(G) - \{\lambda\}$$

<u>Dem</u> Rezultă din Teoremele 1 - 3.