5.2. Simplificarea gramaticilor independente de context (II)

Definiția 3. Fie $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$ o gramatică independentă de context și $X \in \boldsymbol{N}$. Se spune că X este **neterminal/simbol productiv** dacă există cel puțin o derivare de forma $X \overset{*}{\Rightarrow} w$ unde $w \in \boldsymbol{T}^*$. În caz contrar se spune că X este **neterminal/simbol neproductiv**.

Observația 3. Pentru generarea cuvintelor din $\boldsymbol{L}(\boldsymbol{G})$ sunt utile numai neterminalele productive şi deci producţiile care conțin în partea stângă sau în partea dreaptă neterminale neproductive se pot elimina din $\boldsymbol{P}$, fără ca $\boldsymbol{L}(\boldsymbol{G})$ să se modifice.

Propoziția 3. Neterminalele/simbolurile productive și cele neproductive ale unei gramatici independente de context $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$ se pot obţine prin metoda şirului crescător de mulţimi.

Demonstrație. Fie șirul de mulțimi:

$$\begin{cases} \boldsymbol{M_0} = \{X \in \boldsymbol{N} \mid \exists X \to \alpha \in \boldsymbol{P} \text{ cu } \alpha \in \boldsymbol{T}^*\} \\ \boldsymbol{M_{n+1}} = \boldsymbol{M_n} \cup \{X \in \boldsymbol{N} \mid \exists X \to \alpha \in \boldsymbol{P} \text{ cu } \alpha \in (\boldsymbol{T} \cup \boldsymbol{M_n})^*\} \end{cases}.$$

Se constată că $\boldsymbol{M_0} \subset \boldsymbol{M_1} \subset \cdots \subset \boldsymbol{M_n} \subset \boldsymbol{M_{n+1}} \subset \boldsymbol{N}$. Deoarece mulţimea $\boldsymbol{N}$ este finită, rezultă că există un indice $k \in \mathbb{N}^*$ pentru care șirul de mulțimi se stabilizează, adică

$\boldsymbol{M_k} = \boldsymbol{M_{k+1}} = \boldsymbol{M_{k+2}} = \ldots$. Rezultă că $\boldsymbol{M_k}$ este mulţimea neterminalelor productive, iar $\boldsymbol{N} \setminus \boldsymbol{M_k}$ este mulțimea neterminalelor neproductive.

Exemplul 6. Fie gramatica independentă de context $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$, unde

$\boldsymbol{N} = \{S, A, B, C, D\}$, $\boldsymbol{T} = \{a, b\}$ și mulțimea producțiilor $\boldsymbol{P}$ este următoarea:

$S \to AB \mid aBC$	(1)
$A \to BA \mid a$	(2)
$B \to b \mid AC$	(3)
$C \to AC \mid CB$	(4)
$D \to AD \mid a$	(5)

Pentru a determina neterminalele/simbolurile neproductive ale gramaticii $\boldsymbol{G}$ se aplică propoziția 3.

Se obține:

$\boldsymbol{M_0} = \{X \in \boldsymbol{N} \mid \exists X \to \alpha \in \boldsymbol{P} \text{ cu } \alpha \in \boldsymbol{T}^*\} = \{A, B, D\}$, în baza producțiilor (2), (3) şi (5);

$\boldsymbol{M_1} = \boldsymbol{M_0} \cup \{X \in \boldsymbol{N} \mid \exists X \to \alpha \in \boldsymbol{P} \text{ cu } \alpha \in (\boldsymbol{T} \cup \boldsymbol{M_0})^* = (a, b, A, B, D)^*\} = \{S, A, B, D\}$, în baza producţiei (1);

$\boldsymbol{M_2} = \boldsymbol{M_1} \cup \{X \in \boldsymbol{N} \mid \exists X \to \alpha \in \boldsymbol{P} \text{ cu } \alpha \in (\boldsymbol{T} \cup \boldsymbol{M_1})^* = \{a, b, S, A, B, D\}^*\} = \{S, A, B, D\}$.

Lect. univ. dr. TATIANA-CORINA DOSESCU
Facultatea de Informatică
Universitatea Titu Maiorescu

Deoarece $\boldsymbol{M_2} = \boldsymbol{M_1}$ şirul de mulţimi s-a stabilizat, $\boldsymbol{M_2}$ este mulţimea neterminalelor productive, iar $\boldsymbol{N} \setminus \boldsymbol{M_2} = \{C\}$ este mulţimea neterminalelor neproductive.

Deci neterminalul *C* poate fi eliminat, împreună cu producțiile în care apare *C*, simplificând astfel gramatica independentă de context $\boldsymbol{G}$ şi obţinând o gramatică independentă de context $\boldsymbol{G^s} = (\boldsymbol{N^s}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P^s})$ echivalentă cu ***G***, unde $\boldsymbol{N^s} = \{S, A, B, D\}$, iar mulţimea producţiilor $\boldsymbol{P^s}$ este alcătuită din producțiile:

$S \to AB$ (1)
$A \to BA|\ a$ (2)
$B \to b$ (3)
$D \to AD|\ a$ (4)

Definiția 4. Fie $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$ o gramatică independentă de context și $X \in \boldsymbol{N} \cup \boldsymbol{T}$. Se spune că *X* este **simbol accesibil** dacă există cel puțin o derivare de forma $S \overset{*}{\Rightarrow} \alpha X \beta$ unde α, $\beta \in (\boldsymbol{N} \cup \boldsymbol{T})^*$. În caz contrar se spune că *X* este **simbol inaccesibil**.

Observația 4. Pentru generarea cuvintelor din $\boldsymbol{L}(\boldsymbol{G})$ sunt utile numai simbolurile accesibile şi deci producțiile care conțin în partea stângă sau în partea dreaptă simboluri inaccesibile se pot elimina din ***P***, fără ca $\boldsymbol{L}(\boldsymbol{G})$ să se modifice.

Propoziția 4. Simbolurile accesibile și cele inaccesibile ale unei gramatici independente de context $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$ se pot obţine prin metoda şirului crescător de mulţimi.

Demonstrație. Fie șirul de mulțimi:

$$\begin{cases} \boldsymbol{M_0} = \{S\} \\ \boldsymbol{M_{n+1}} = \boldsymbol{M_n} \cup \{X \in \boldsymbol{N} \cup \boldsymbol{T} |\ \exists Y \in \boldsymbol{M_n} \cap \boldsymbol{N}, \text{astfel încât } Y \to \alpha X \beta \in \boldsymbol{P} \text{ cu } \alpha, \beta \in (\boldsymbol{N} \cup \boldsymbol{T})^*\} \end{cases}$$

Se constată că $\boldsymbol{M_0} \subset \boldsymbol{M_1} \subset \cdots \subset \boldsymbol{M_n} \subset \boldsymbol{M_{n+1}} \subset \boldsymbol{N} \cup \boldsymbol{T}$. Deoarece mulțimea $\boldsymbol{N} \cup \boldsymbol{T}$ este finită, rezultă că există un indice $k \in \mathbb{N}^*$ pentru care șirul de mulțimi se stabilizează, adică $\boldsymbol{M_k} = \boldsymbol{M_{k+1}} = \boldsymbol{M_{k+2}}$= ... . Rezultă că $\boldsymbol{M_k}$ este mulțimea simbolurile accesibile, iar

$(\boldsymbol{N} \cup \boldsymbol{T}) \setminus \boldsymbol{M_k}$ este mulțimea simbolurile inaccesibile.

Exemplul 7. Fie gramatica independentă de context $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$, unde

N = $\{S, A, B, D\}$, ***T*** = $\{a, b\}$ și mulțimea producțiilor ***P*** este următoarea:

$S \to AB$ (1)
$A \to BA|\ a$ (2)
$B \to b$ (3)
$D \to AD|\ a$ (4)

Pentru a determina simbolurile accesibile și simbolurile inaccesibile ale gramaticii ***G*** se aplică propoziția 4.

Se obține:

$\boldsymbol{M_0} = \{S\}$;

$\boldsymbol{M_1} = \boldsymbol{M_0} \cup \{X \in \boldsymbol{N} \cup \boldsymbol{T} |\ \exists Y \in \boldsymbol{M_0} \cap \boldsymbol{N}, \text{a î.}\, Y \rightarrow \alpha X \beta \in \boldsymbol{P} \text{ cu } \alpha, \beta \in (\boldsymbol{N} \cup \boldsymbol{T})^*\} =$

$=\{S\} \cup \{X \in \boldsymbol{N} \cup \boldsymbol{T} |\ \exists Y \in \{S\} \text{ a î.}\, Y \rightarrow \alpha X \beta \in \boldsymbol{P} \text{ cu } \alpha, \beta \in (\boldsymbol{N} \cup \boldsymbol{T})^*\} =$

$=\{S\} \cup \{A, B\} = \{S, A, B\}$, deoarece $S \rightarrow \lambda AB$ şi $S \rightarrow AB\lambda$;

$\boldsymbol{M_2} = \boldsymbol{M_1} \cup \{X \in \boldsymbol{N} \cup \boldsymbol{T} |\ \exists Y \in \boldsymbol{M_1} \cap \boldsymbol{N}, \text{a î.}\, Y \rightarrow \alpha X \beta \in \boldsymbol{P} \text{ cu } \alpha, \beta \in (\boldsymbol{N} \cup \boldsymbol{T})^*\} =$

$=\{S, A, B\} \cup \{X \in \boldsymbol{N} \cup \boldsymbol{T} |\ \exists Y \in \{S, A, B\} \text{ a î. } Y \rightarrow \alpha X \beta \in \boldsymbol{P} \text{ cu } \alpha, \beta \in (\boldsymbol{N} \cup \boldsymbol{T})^*\} =$

$=\{S, A, B\} \cup \{A, B, a, b\} = \{S, A, B, a, b\}$, deoarece $S \rightarrow \lambda AB$, $S \rightarrow AB\lambda$, $A \rightarrow \lambda a\lambda$ și

$B \rightarrow \lambda b\lambda$;

$\boldsymbol{M_3} = \boldsymbol{M_2} \cup \{X \in \boldsymbol{N} \cup \boldsymbol{T} |\ \exists Y \in \boldsymbol{M_2} \cap \boldsymbol{N}, \text{a î.}\, Y \rightarrow \alpha X \beta \in \boldsymbol{P} \text{ cu } \alpha, \beta \in (\boldsymbol{N} \cup \boldsymbol{T})^*\} =$

$=\{S, A, B, a, b\} \cup \{X \in \boldsymbol{N} \cup \boldsymbol{T} |\ \exists Y \in \{S, A, B\}, \text{a î.}\, Y \rightarrow \alpha X \beta \in \boldsymbol{P} \text{ cu } \alpha, \beta \in (\boldsymbol{N} \cup \boldsymbol{T})^*\}=$

$= \{S, A, B, a, b\} \cup \{A, B, a, b\} = \{S, A, B, a, b\}.$

Deoarece $\boldsymbol{M_3} = \boldsymbol{M_2}$ şirul de mulţimi s-a stabilizat, $\boldsymbol{M_3}$ este mulţimea simbolurilor accesibile, iar $(\boldsymbol{N} \cup \boldsymbol{T}) \setminus \boldsymbol{M_2} = \{D\}$ este mulţimea simbolurilor inaccesibile.

Deci simbolul D poate fi eliminat, împreună cu producțiile în care apare D, simplificând astfel gramatica independentă de context $\boldsymbol{G}$ şi obţinând o gramatică independentă de context $\boldsymbol{G^a} = (\boldsymbol{N^a}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P^a})$ echivalentă cu $\boldsymbol{G}$, unde $\boldsymbol{N^a} = \{S, A, B\}$, iar mulţimea producţiilor $\boldsymbol{P^a}$ este alcătuită din producțiile:

$S \rightarrow AB$ (1)

$A \rightarrow BA |\ a$ (2)

$B \rightarrow b$ (3)

Aplicație. Fie gramatica independentă de context $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$, unde

$\boldsymbol{N} = \{S, X, Y, Z\}$, $\boldsymbol{T} = \{a, b\}$ și mulțimea producțiilor $\boldsymbol{P}$ este următoarea:

$S \rightarrow bb |\ aXY |\ Xa |\ a |\ XY |\ Zb$ (1)

$X \rightarrow aXY |\ XY |\ Zb |\ bb |\ Xa |\ a$ (2)

$Y \rightarrow Xa |\ a$ (3)

$Z \rightarrow XY |\ Zb |\ bb |\ aXY |\ Xa |\ a$ (4)

i) Să se determine gramatică independentă de context $\boldsymbol{G^s} = (\boldsymbol{N^s}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P^s})$ echivalentă cu $\boldsymbol{G}$, obținută prin eliminarea neterminalelor neproductive ale gramaticii $\boldsymbol{G}$.

ii) Să se determine gramatică independentă de context $\boldsymbol{G^a} = (\boldsymbol{N^a}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P^a})$ echivalentă cu $\boldsymbol{G^s}$, obţinută prin eliminarea simbolurilor inaccesibile ale gramaticii $\boldsymbol{G^a}$.

Rezolvare.

i) Pentru a determina neterminalele neproductive ale gramaticii ***G*** se aplică propoziția 3.

Se obține:

$\boldsymbol{M_0} = \{X \in \boldsymbol{N} |\ \exists X \to \alpha \in \boldsymbol{P} \text{ cu } \alpha \in \boldsymbol{T}^*\}$ = $\{S, X, Y, Z\}$, în baza producțiilor (1), (2), (3) și (4);

Cum $\boldsymbol{M_0} = \boldsymbol{N}$, șirul de mulțimi este stabilizat și ***N*** este mulțimea tuturor neterminalelor productive, iar mulțimea tuturor neterminalelor neproductive este vidă. Prin urmare, gramatica $\boldsymbol{G^s}$ coincide cu gramatica ***G***.

ii) Pentru a determina simbolurile accesibile și simbolurile inaccesibile ale gramaticii ***G*** se aplică propoziția 4.

Se obține:

$\boldsymbol{M_0} = \{S\}$;

$\boldsymbol{M_1} = \boldsymbol{M_0} \cup \{W \in \boldsymbol{N} \cup \boldsymbol{T} |\ \exists U \in \boldsymbol{M_0} \cap \boldsymbol{N}, \text{a î.}\, U \to \alpha W \beta \in \boldsymbol{P} \text{ cu } \alpha, \beta \in (\boldsymbol{N} \cup \boldsymbol{T})^*\} =$

$= \{S\} \cup \{W \in \boldsymbol{N} \cup \boldsymbol{T} |\ \exists U \in \{S\} \text{ a î.}\, S \to \alpha X \beta \in \boldsymbol{P} \text{ cu } \alpha, \beta \in (\boldsymbol{N} \cup \boldsymbol{T})^*\} =$

$= \{S\} \cup \{X, Y, Z, a, b\} = \boldsymbol{N} \cup \boldsymbol{T}$, deoarece în baza lui (1) se poate scrie:

$$\begin{array}{lll} S \to \lambda bb & \Longrightarrow & b \in M_1; \\ S \to aXY & \Longrightarrow & X \in M_1; \\ S \to aXY\lambda & \Longrightarrow & Y \in M_1; \\ S \to Xa\lambda & \Longrightarrow & a \in M_1; \\ S \to \lambda Zb & \Longrightarrow & Z \in M_1. \end{array}$$

Cum $\boldsymbol{M_1} = \boldsymbol{N} \cup \boldsymbol{T}$, șirul de mulțimi este stabilizat și ***N*** ∪ ***T*** este mulțimea tuturor simbolurile accesibile, iar mulțimea tuturor simbolurile inaccesibile este vidă. Prin urmare, gramatica $\boldsymbol{G^a}$ coincide cu gramatica ***G***.

5.3. Forma normală Chomsky a unei gramatici independente de context

Definiția 1. O gramatică independentă de context $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$ este în **forma normală Chomsky** dacă orice producție a sa este fie de forma $A \to a$, fie de forma $A \to BC$, unde $A, B, C \in \boldsymbol{N}$ și $a \in \boldsymbol{T}$.

Propoziția 1. Fie o gramatică independentă de context $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$. Atunci există o gramatică independentă de context $\widetilde{\boldsymbol{G}}$ echivalentă cu ***G*** ale cărei producții sunt fie de forma *A* → *a*, fie de forma *A* → $B_1B_2 \dots B_m$, unde *a* ∈ ***T*** și $A, B_1B_2 \dots B_m$ ∈ ***N***.

Demonstrație.

În baza teoremei din cursul 6, se poate presupune că în gramatica ***G*** nu există redenumiri, adică producții de forma *A* → *B*.

Se consideră $\widetilde{\boldsymbol{N}}$ o dublură a lui ***N*** \ $\{S\}$. Atunci există o bijecție ***f*** : ***N*** \ $\{S\}$ → $\widetilde{\boldsymbol{N}}$, unde

$$\boldsymbol{f}(X) \overset{notat}{=} \tilde{X}.$$

Fie gramatica $\widetilde{\boldsymbol{G}} = \left(\boldsymbol{N} \cup \widetilde{\boldsymbol{N}}, \boldsymbol{T}, \boldsymbol{S}, \widetilde{\boldsymbol{P}}\right)$, unde pentru construcţia lui $\widetilde{\boldsymbol{P}}$ se consideră pe rând producțiile *A* → α ∈ ***P***, astfel:

- dacă $|\propto| = 1$, atunci α ∈ ***T*** și producția este trecută în $\widetilde{\boldsymbol{P}}$;
- dacă $|\propto| > 1$, atunci în $\widetilde{\boldsymbol{P}}$ se înscrie producția *A* → $\tilde{\propto}$, unde $\tilde{\propto}$ se obține din α înlocuind neterminalele cu corespondentul lor din $\widetilde{\boldsymbol{N}}$.

În final, se adaugă la $\widetilde{\boldsymbol{P}}$ producţiile din mulţimea $\{\tilde{X} \rightarrow a \big| \; a \in \boldsymbol{T}\}$.

Teorema 1. Fie ***G*** o gramatică independentă de context. Atunci există o gramatică independentă de context $\boldsymbol{G}^{(\boldsymbol{n})} = \left(\boldsymbol{N}^{(\boldsymbol{n})}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P}^{(\boldsymbol{n})}\right)$ în forma normală Chomsky echivalentă cu ***G***.

Demonstrație.

În baza propoziției de mai sus, se poate presupune că toate producțiile gramaticii ***G*** sunt fie de forma *A* → *a*, fie de forma *A* → $B_1B_2 \dots B_m$, unde *a* ∈ ***T*** și $A, B_1B_2 \dots B_m$ ∈ ***N***.

Pentru a se obține forma normală Chomsky, o producție de forma *A* → $B_1B_2 \dots B_m$, unde *a* ∈ ***T*** și $A, B_1B_2 \dots B_m$ ∈ ***N*** trebuie înlocuită cu producții de forma *X* → *YZ*, astfel încât limbajul generat de gramatica ***G*** să nu se modifice, în felul următor:

$$\begin{cases} A \rightarrow B_1D_1 \\ D_1 \rightarrow B_2D_2 \\ \dots \\ D_{m-2} \rightarrow B_{m-1}B_m \end{cases},$$

unde $D_1, D_2, \dots, D_{m-2}$ sunt simboluri neterminale noi ce se adaugă la neterminalele din ***N*** și contribuie la obținerea mulțimii $\boldsymbol{N}^{(\boldsymbol{n})}$. De asemeni, și cele *m*-1 producții de mai sus contribuie la obținerea mulțimii producțiilor $\boldsymbol{P}^{(\boldsymbol{n})}$.

Exemplul 1. Fie o gramatica independentă de context $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$, unde

$\boldsymbol{N} = \{S, A, B\}$, $\boldsymbol{T} = \{a, b\}$, iar mulțimea producțiilor ***P*** este:

Lect. univ. dr. TATIANA-CORINA DOSESCU
Facultatea de Informatică
Universitatea Titu Maiorescu

$$S \to ab|\ bA|\ A \qquad (1)$$
$$A \to bAa|\ aS|\ B|\ a \qquad (2)$$
$$B \to aBb|\ bS|\ b \qquad (3)$$

Se observă că gramatica ***G*** conține redenumiri, cum ar fi *S → A*, *A → B,* și pentru a o aduce la forma normală Chomsky trebuie eliminate, aplicând algoritmul corespunzător:

Redenumiri	**Producții noi**
$S \to A$	$S \to bAa\|\ aS\|\ a$
$A \to B$	$A \to aBb\|\ bS\|\ b$
$S \overset{*}{\Rightarrow} B$	$S \to aBb\|\ bS\|\ b$

Conform tabelului se obține gramatica $\boldsymbol{G_1} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P_1})$ echivalentă cu gramatica ***G***, fără redenumiri, ale cărei producții ce alcătuiesc mulțimea $\boldsymbol{P_1}$ sunt:

$$S \to ab|\ bA|\ bAa|\ aS|\ a|\ aBb|\ bS|\ b \qquad (1)$$
$$A \to bAa|\ aS|\ a|\ aBb|\ bS|\ b \qquad (2)$$
$$B \to aBb|\ bS|\ b \qquad (3)$$

Asupra gramaticii $\boldsymbol{G_1}$ se aplică propoziția 1, obținându-se o gramatică echivalentă notată prin $\boldsymbol{G_2} = (\boldsymbol{N_2}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P_2})$, unde $\boldsymbol{N_2} = \{\boldsymbol{S}, \boldsymbol{A}, \widetilde{\boldsymbol{A}}, \boldsymbol{B}, \widetilde{\boldsymbol{B}}\}$ şi producţiile care alcătuiesc mulţimea $\boldsymbol{P_2}$ sunt:

$$\tilde{A} \to a \qquad (1)$$
$$\tilde{B} \to b \qquad (2)$$
$$S \to \tilde{A}\tilde{B}\big|\ \tilde{B}A\big|\ \tilde{B}A\tilde{A}\big|\ \tilde{A}S|\ a\big|\ \tilde{A}B\tilde{B}\big|\ \tilde{B}S|\ b \qquad (3)$$
$$A \to \tilde{B}A\tilde{A}\big|\ \tilde{A}S|\ a\big|\ \tilde{A}B\tilde{B}\big|\ \tilde{B}S|\ b \qquad (4)$$
$$B \to \tilde{A}B\tilde{B}\big|\ \tilde{B}S|\ b \qquad (5)$$

Asupra gramaticii $\boldsymbol{G_2}$ se aplică teorema de mai sus, obţinându-se o gramatică independentă de context $\boldsymbol{G^{(n)}} = \left(\boldsymbol{N^{(n)}}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P^{(n)}}\right)$ în forma normală Chomsky echivalentă cu ***G***, unde:

- $\boldsymbol{N^{(n)}} = \{S, A, B, \tilde{A}, \tilde{B}, D_1, D_2, D_3, D_4, D_5\}$;
- mulțimea producțiilor $\boldsymbol{P^{(n)}}$ este alcătuită din:

$$
\begin{array}{ll}
\tilde{A} \rightarrow a & (1) \\
\tilde{B} \rightarrow b & (2) \\
S \rightarrow \tilde{A}\tilde{B} \big| \tilde{B}A \big| \tilde{B}D_1 \big| \tilde{A}S \big| a \big| \tilde{A}D_2 \big| \tilde{B}S \big| b & (3) \\
D_1 \rightarrow A\tilde{A} & (4) \\
D_2 \rightarrow B\tilde{B} & (5) \\
A \rightarrow \tilde{B}D_3 \big| \tilde{A}S \big| a \big| \tilde{A}D_4 \big| \tilde{B}S \big| b & (6) \\
D_3 \rightarrow A\tilde{A} & (7) \\
D_4 \rightarrow B\tilde{B} & (8) \\
B \rightarrow \tilde{A}D_5 \big| \tilde{B}S \big| b & (9) \\
D_5 \rightarrow B\tilde{B} & (10)
\end{array}
$$

Exemplul 2. Fie o gramatica independentă de context $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$, unde

$\boldsymbol{N} = \{S, A, B\}$, $\boldsymbol{T} = \{a, b\}$, iar mulțimea producțiilor $\boldsymbol{P}$ este:

$$
\begin{array}{ll}
S \rightarrow aABA & (1) \\
A \rightarrow AaA \big| AB \big| a & (2) \\
B \rightarrow BbB \big| BAb \big| b & (3)
\end{array}
$$

Se observă că gramatica $\boldsymbol{G}$ nu conține redenumiri și pentru a o aduce la forma normală Chomsky se aplică propoziția 1, obținându-se o gramatică echivalentă notată prin

$\boldsymbol{G_2} = (\boldsymbol{N_2}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P_2})$, unde $\boldsymbol{N_2} = \{S, A, \tilde{A}, B, \tilde{B}\}$ şi producţiile care alcătuiesc mulţimea $\boldsymbol{P_2}$ sunt:

$$
\begin{array}{ll}
\tilde{A} \rightarrow a & (1) \\
\tilde{B} \rightarrow b & (2) \\
S \rightarrow \tilde{A}ABA & (3) \\
A \rightarrow A\tilde{A}A \big| AB \big| a & (4) \\
B \rightarrow B\tilde{B}B \big| BA\tilde{B} \big| b & (5)
\end{array}
$$

Asupra gramaticii $\boldsymbol{G_2}$ se aplică teorema de mai sus, obţinându-se o gramatică independentă de context $\boldsymbol{G}^{(\boldsymbol{n})} = \left(\boldsymbol{N}^{(\boldsymbol{n})}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P}^{(\boldsymbol{n})}\right)$ în forma normală Chomsky echivalentă cu $\boldsymbol{G}$, unde:

- $\boldsymbol{N}^{(\boldsymbol{n})} = \{S, A, B, \tilde{A}, \tilde{B}, D_1, D_2, D_3, D_4, D_5\}$;
- mulțimea producțiilor $\boldsymbol{P}^{(\boldsymbol{n})}$ este alcătuită din:

$$
\begin{array}{ll}
\tilde{A} \rightarrow a & (1) \\
\tilde{B} \rightarrow b & (2) \\
S \rightarrow \tilde{A}D_1 & (3) \\
D_1 \rightarrow AD_2 & (4)
\end{array}
$$

$$D_2 \rightarrow BA \quad (5)$$
$$A \rightarrow AD_3 \quad (6)$$
$$D_3 \rightarrow \tilde{A}A \quad (7)$$
$$A \rightarrow AB \quad (8)$$
$$A \rightarrow a \quad (9)$$
$$B \rightarrow BD_4 \quad (10)$$
$$D_4 \rightarrow \tilde{B}B \quad (11)$$
$$B \rightarrow BD_5 \quad (12)$$
$$D_5 \rightarrow A\tilde{B} \quad (13)$$
$$B \rightarrow b \quad (14).$$

Exemplul 3. Fie o gramatica independentă de context $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$, unde

$\boldsymbol{N} = \{S, A, B, C, D, E\}$, $\boldsymbol{T} = \{a, b\}$, iar mulțimea producțiilor ***P*** este:

$$S \rightarrow aB \mid AC \quad (1)$$
$$A \rightarrow ASC \mid BC \mid aD \mid a \quad (2)$$
$$B \rightarrow bS \mid b \quad (3)$$
$$C \rightarrow BA \mid \lambda \quad (4)$$
$$D \rightarrow abC \quad (5)$$
$$E \rightarrow aB \quad (6)$$

Se observă că gramatica ***G*** are o λ - producție, simboluri neproductive și simboluri inaccesibile, care trebuie eliminate pentru a o aduce la forma normală Chomsky, eliminarea se face în 5 etape succesive.

Etapa 1. Eliminarea λ - producției.

În această etapă se aplică algoritmul pentru eliminarea λ - producțiilor dintr-o gramatică independentă de context, astfel:

Pasul 1. $\boldsymbol{M} = \left\{A \in \boldsymbol{N} \middle| A \overset{*}{\Rightarrow} \lambda\right\} = \{C\}$;

Pasul 2. $\boldsymbol{P^1} = \boldsymbol{P} \setminus \{A \rightarrow \lambda \mid A \in \boldsymbol{M}\}$, numită $\boldsymbol{P^1}$ iniţială, este alcătuită din:

$$S \rightarrow aB \mid AC \quad (1)$$
$$A \rightarrow ASC \mid BC \mid aD \mid a \quad (2)$$
$$B \rightarrow bS \mid b \quad (3)$$
$$C \rightarrow BA \quad (4)$$
$$D \rightarrow abC \quad (5)$$
$$E \rightarrow aB \quad (6)$$

Pasul 3. $\boldsymbol{P^1}$ finală este alcătuită din:

$S \rightarrow aB \mid AC$	(1)
$A \rightarrow ASC \mid BC \mid aD \mid a$	(2)
$B \rightarrow bS \mid b$	(3)
$C \rightarrow BA$	(4)
$D \rightarrow abC$	(5)
$E \rightarrow aB$	(6)
$S \rightarrow A$	(7)
$A \rightarrow AS$	(8)
$A \rightarrow B$	(9)
$D \rightarrow ab$	(10)

În acest fel s-a obținut gramatica independentă de context $\boldsymbol{G_1} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P_1})$, care nu conține λ - producții și este echivalentă cu gramatica ***G***.

Etapa 2. Eliminarea simbolurilor neproductive din gramatica $\boldsymbol{G_1}$.

În această etapă se determină, mai întâi, simbolurile productive cu metoda șirului crescător de mulțimi astfel:

$\boldsymbol{M_0} = \{X \in \boldsymbol{N} \mid \exists X \rightarrow \alpha \in \boldsymbol{P_1} \text{ cu } \alpha \in \boldsymbol{T}^*\} = \{A, B, D\}$, deoarece $A \rightarrow a$, $B \rightarrow b$ și respectiv $D \rightarrow ab$;

$\boldsymbol{M_1} = \boldsymbol{M_0} \cup \{X \in \boldsymbol{N} \mid \exists X \rightarrow \alpha \in \boldsymbol{P_1} \text{ cu } \alpha \in (\boldsymbol{T} \cup \boldsymbol{M_0})^* = (a, b, A, B, D)^*\} =$

$= \{A, B, D\} \cup \{S, A, B, C, D, E\} = \boldsymbol{N}$, deoarece $S \rightarrow A$, $C \rightarrow BA$ și respectiv $E \rightarrow aB$.

Deoarece $\boldsymbol{M_1} = \boldsymbol{N}$ şirul s-a stabilizat și ***N*** este mulțimea simbolurilor productive, adică nu există simboluri neproductive.

Prin urmare gramatica $\boldsymbol{G_1}$nu conţine simboluri neproductive.

Etapa 3. Eliminarea simbolurilor inaccesibile din gramatica $\boldsymbol{G_1}$.

În această etapă se determină, mai întâi, simbolurile accesibile cu metoda șirului crescător de mulțimi astfel:

$\boldsymbol{M_0} = \{S\}$;

$\boldsymbol{M_1} = \boldsymbol{M_0} \cup \{X \in \boldsymbol{N} \cup \boldsymbol{T} \mid \exists Y \in \boldsymbol{M_0} \cap \boldsymbol{N}, \text{a î.} Y \rightarrow \alpha X \beta \in \boldsymbol{P} \text{ cu } \alpha, \beta \in (\boldsymbol{N} \cup \boldsymbol{T})^*\} =$

$=\{S\} \cup \{X \in \boldsymbol{N} \cup \boldsymbol{T} \mid \exists Y \in \{S\} \text{ a î.} Y \rightarrow \alpha X \beta \in \boldsymbol{P} \text{ cu } \alpha, \beta \in (\boldsymbol{N} \cup \boldsymbol{T})^*\} =$

Lect. univ. dr. TATIANA-CORINA DOSESCU
Facultatea de Informatică
Universitatea Titu Maiorescu

$=\{S\} \cup \{a, A, B, C\} = \{a, S, A, B, C\}$ deoarece $S \rightarrow \lambda A\lambda$, $S \rightarrow \lambda aB$, $S \rightarrow aB\lambda$ și respectiv $S \rightarrow AC\lambda$;

$\boldsymbol{M_2} = \boldsymbol{M_1} \cup \{X \in \boldsymbol{N} \cup \boldsymbol{T} | \exists Y \in \boldsymbol{M_1} \cap \boldsymbol{N}, \text{a î.} Y \rightarrow \alpha X\beta \in \boldsymbol{P} \text{ cu } \alpha, \beta \in (\boldsymbol{N} \cup \boldsymbol{T})^*\} =$

$= \{a, S, A, B, C\} \cup$

$\cup \{X \in \boldsymbol{N} \cup \boldsymbol{T} | \exists Y \in \{a, S, A, B, C\} \text{ a î.} Y \rightarrow \alpha X\beta \in \boldsymbol{P} \text{ cu } \alpha, \beta \in (\boldsymbol{N} \cup \boldsymbol{T})^*\} =$

$= \{a, S, A, B, C\} \cup \{b, S, A, B, C, D\} = \{a, b, S, A, B, C, D\}$, deoarece $A \rightarrow aD\lambda$;

$\boldsymbol{M_3} = \boldsymbol{M_2} \cup \{X \in \boldsymbol{N} \cup \boldsymbol{T} | \exists Y \in \boldsymbol{M_2} \cap \boldsymbol{N}, \text{a î.} Y \rightarrow \alpha X\beta \in \boldsymbol{P} \text{ cu } \alpha, \beta \in (\boldsymbol{N} \cup \boldsymbol{T})^*\} =$

$= \{a, b, S, A, B, C, D\}$.

Deoarece $\boldsymbol{M_3} = \boldsymbol{M_2}$ şirul de mulţimi s-a stabilizat şi $\boldsymbol{M_3}$ este mulţimea simbolurilor accesibile. Rezultă că mulţimea simbolurilor inaccesibile este $(\boldsymbol{N} \cup \boldsymbol{T}) \setminus \boldsymbol{M_3} = \{E\}$.

Deci simbolul *E* poate fi eliminat, împreună cu producțiile în care apare *E*, simplificând astfel gramatica independentă de context $\boldsymbol{G}$ şi obţinând o gramatică independentă de context $\boldsymbol{G^a} = (\boldsymbol{N^a}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P^a})$ echivalentă cu $\boldsymbol{G}$, unde $\boldsymbol{N^a} = \{\boldsymbol{S}, \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}\}$, iar mulţimea producţiilor $\boldsymbol{P^a}$ este alcătuită din producțiile:

$S \rightarrow aB | AC | A$ (1)

$A \rightarrow ASC | BC | aD | a | AS | B$ (2)

$B \rightarrow bS | b$ (3)

$C \rightarrow BA$ (4)

$D \rightarrow abC | ab$ (5)

Etapa 4. Eliminarea redenumirilor din gramatica $\boldsymbol{G^a}$.

În această etapă aplicând algoritmul de eliminare a redenumirilor, se determină toate redenumirile din $\boldsymbol{P}$, precum și producțiile noi necesare, în următorul tabel:

Redenumiri	**Producții noi**
$S \rightarrow A$	$S \rightarrow ASC$
	$S \rightarrow BC$
	$S \rightarrow aD$
	$S \rightarrow a$
	$S \rightarrow AS$

Lect. univ. dr. TATIANA-CORINA DOSESCU
Facultatea de Informatică
Universitatea Titu Maiorescu

$A \rightarrow B$	$A \rightarrow bS$ $A \rightarrow b$
$S \overset{*}{\Rightarrow} B$	$S \rightarrow bS$ $S \rightarrow b$

Gramatica independentă de context echivalentă cu $\boldsymbol{G^a}$ și fără redenumiri este

$\boldsymbol{G^0} = (\boldsymbol{N^a}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P_0})$, unde mulţimea producţiilor $\boldsymbol{P_0}$ este următoarea:

$$S \rightarrow aB|\ AC|\ ASC|BC|\ aD|AS|\ bS|\ a|\ b \quad (1)$$
$$A \rightarrow ASC|\ BC|\ aD|\ a|\ AS|\ bS|\ b \quad (2)$$
$$B \rightarrow bS|\ b \quad (3)$$
$$C \rightarrow BA \quad (4)$$
$$D \rightarrow abC|\ ab \quad (5)$$

Etapa 5. Aducerea gramaticii $\boldsymbol{G^0} = (\boldsymbol{N^a}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P_0})$ la forma normală Chomsky.

Mai întâi, asupra gramaticii $\boldsymbol{G^0}$ se aplică propoziția 1, obținându-se o gramatică echivalentă notată prin $\boldsymbol{G_2} = (\boldsymbol{N_2}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P_2})$, unde:

$$\boldsymbol{N_2} = \boldsymbol{N^a} \cup \widetilde{\boldsymbol{N^a} \setminus \{S\}} = \{S, A, \tilde{A}, B, \tilde{B}, C, D\}$$

și producțiile care alcătuiesc mulțimea $\boldsymbol{P_2}$ sunt:

$$\tilde{A} \rightarrow a \quad (1)$$
$$\tilde{B} \rightarrow b \quad (2)$$
$$S \rightarrow \tilde{A}B|\ AC|\ ASC|\ BC|\ \tilde{A}D|\ AS|\ \tilde{B}S|\ a|\ b \quad (3)$$
$$A \rightarrow ASC|\ BC|\ \tilde{A}D|\ a|\ AS|\ \tilde{B}S|\ b \quad (4)$$
$$B \rightarrow \tilde{B}S|b \quad (5)$$
$$C \rightarrow BA \quad (6)$$
$$D \rightarrow \tilde{A}\tilde{B}C|\ ab \quad (7)$$

Acum, deoarece există producții de forma de forma $A \rightarrow B_1B_2 \ldots B_m$ cu $m > 2$, asupra gramaticii $\boldsymbol{G_2}$ se aplică teorema de mai sus, obţinându-se o gramatică independentă de context $\boldsymbol{G^{(n)}} = \left(\boldsymbol{N^{(n)}}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P^{(n)}}\right)$ în forma normală Chomsky echivalentă cu $\boldsymbol{G_2}$, unde:

- $\boldsymbol{N^{(n)}} = \{S, A, B, \tilde{A}, \tilde{B}, C, D, X_1, X_2\}$;
- mulțimea producțiilor $\boldsymbol{P^{(n)}}$ este alcătuită din:

$$\tilde{A} \rightarrow a \quad (1)$$
$$\tilde{B} \rightarrow b \quad (2)$$

$$S \rightarrow \tilde{A}B \mid AC \mid AX_1 \mid AS \mid BC \mid \tilde{A}D \mid \tilde{B}S \mid a \mid b \quad (3)$$
$$X_1 \rightarrow SC \quad (4)$$
$$A \rightarrow AX_1 \mid BC \mid \tilde{A}D \mid a \mid AS \mid \tilde{B}S \mid b \quad (5)$$
$$B \rightarrow \tilde{B}S \mid b \quad (6)$$
$$C \rightarrow BA \quad (7)$$
$$D \rightarrow \tilde{A}X_2 \mid \tilde{A}\tilde{B} \quad (8)$$
$$X_2 \rightarrow \tilde{B}C \quad (9).$$

Lect. univ. dr. TATIANA-CORINA DOSESCU
Facultatea de Informatică
Universitatea Titu Maiorescu