

~ Seminar 5 ~

Gramatici

$G = (N, T, S, P)$

$N = \{A, B, C, \dots\}$ mulțimea de simboluri neterminale

$T = \{a, b, c, \dots\}$ mulțimea de simboluri terminale

$S \in N$ simbolul de start

P mulțimea de producții (sau reguli de producție)

Gramatici independente de context (GIC) / Context-free grammars (CFG)

$$P \subseteq N \times (N \cup T)^*$$

Forma Normală Chomsky

O gramatică este (scrisă) în **forma normală Chomsky** dacă are doar producții care au în membrul stâng un neterminal, iar în membrul drept fie un terminal, fie două neterminale.

$A \rightarrow a$

$A \rightarrow BC$, unde $A, B, C \in N$, $a \in T$

Transformarea GIC \rightarrow gramatică în F.N. Chomsky

Obs: Pentru fiecare pas detaliat sub formă de algoritm și aplicat pe exemple

\rightarrow **[vezi CURS 8 + 9]**

Pas 1: Se aplică **algoritmul de reducere** (se elimină simbolurile și producțiile nefolositoare):

[Curs 8, pag 2 – Algoritmul de la Teorema 1]

- a) Se elimină simbolurile și producțiile “**neutilizabile**”. **[pasi 1-3 din alg. \Rightarrow mulț. N_1]**

Def: $X \in (N \cup T)$ este un simbol “neutilizabil” dacă nu există nicio derivare

$$S \overset{*}{\Rightarrow} \alpha X \gamma \overset{*}{\Rightarrow} \alpha \beta \gamma \text{ cu } \alpha, \beta, \gamma \in T^*.$$

(Adică plecând de la un cuvânt care conține neterminalul X și aplicând oricât de multe producții, nu putem ajunge să generăm un cuvânt care să nu conțină neterminale.)

Ideea intuitivă: Găsim mulțimea simbolurilor din N care sunt **utilizabile**:

-- Plecăm de la cele care au cel puțin o producție cu membrul drept (de tip “caz de oprire”) în T^* .

-- Iar apoi tot căutăm alte neterminale care au cel puțin o producție cu membrul drept în $(N \cup T)^*$, dar în care toate neterminalele fac deja parte din mulțimea celor utilizabile.

- b) Se elimină simbolurile și producțiile “**inaccesibile**”. **[pasi 4-5 din alg. \Rightarrow mulț. N_2]**

Def: $X \in (N \cup T)$ este un simbol “inaccesibil” dacă nu există nicio derivare

$$S \overset{*}{\Rightarrow} \alpha X \gamma, \text{ cu } \alpha, \gamma \in (N \cup T)^*.$$

(Adică plecând din S și aplicând oricât de multe producții, nu putem ajunge să generăm un cuvânt care să conțină neterminalul X .)

Ideea intuitivă: Găsim mulțimea simbolurilor din N care sunt **accesibile**:

-- Plecăm de la neterminalele care apar în membrul drept al producțiilor lui S .

-- Apoi tot căutăm alte neterminale care apar în membrul drept al producțiilor acelor neterminale care fac deja parte din mulțimea celor accesibile.

Pas 2: Se elimină λ -producțiile. [Curs 8, pag 5 – Alg. de la Teorema 2 \Rightarrow mulțimile N_a și P_1]

Obs: Dacă $\lambda \in L(G)$ (cuvântul vid trebuie să fie generat de gramatică), atunci avem voie să avem unica λ -producție a neterminalului de start, dar acesta nu are voie să mai apară în gramatică nicăieri (în membrul drept al producțiilor).

\rightarrow Se adaugă un nou simbol de start S' și producțiile $S' \rightarrow S \mid \lambda$.

Vrem să nu mai avem în gramatică producții care au ca membru drept cuvântul vid.
Pot exista două situații:

- (i) Dacă neterminalul care are o λ -producție NU are și alte producții, atunci
 - va fi eliminată producția lui
 - toate producțiile care au ca membru drept un cuvânt de lungime minim 2 în care apare acest neterminal \rightarrow vor fi înlocuite prin eliminarea neterminalului din cuvinte
 - dacă membrul drept avea lungime 1 (adică era doar acest neterminal), atunci se înlocuiește cu λ (dar și această producție va fi ulterior eliminată tot la „pasul 2”)
- (ii) Dacă neterminalul care are o λ -producție are și alte producții, atunci
 - va fi eliminată doar λ -producția lui.
 - toate producțiile care au ca membru drept un cuvânt de lungime minim 2 în care apare acest neterminal \rightarrow vor fi înlocuite atât de varianta în care cuvântul conține neterminalul, cât și de varianta în care neterminalul este eliminat din cuvânt.
 - dacă membrul drept avea lungime 1 (adică era doar acest neterminal), atunci această producție este o redenumire și va fi eliminată la „pasul 3”.

Pas 3: Se elimină redenumirile. [Curs 8, pag 10 – Alg. de la Teorema 3 \Rightarrow mulțimea P']

Vrem să nu mai avem în gramatică producții care au ca membru drept un neterminal.

Înlocuim neterminalul din dreapta cu toate cuvintele care sunt membru drept în producțiile sale.

Verificăm dacă acest neterminal din dreapta mai apare în dreapta în cadrul altor producții

- Dacă nu, atunci eliminăm toate producțiile pe care le avea.
- Dacă mai apare în cuvinte de lungime minim 2, atunci trebuie să i le păstrăm

Dacă apare în cuvânt de lungime 1, înseamnă că este tot o redenumire și va fi eliminată tot la „pasul 3”.

Pas 4: Se aplică din nou *algoritmul de reducere* (vezi Pas 1).

Pt. pasii 5 și 6 \rightarrow [Curs 9, pag 4 – Algoritmul din teorema]

Pas 5: Se adaugă neterminale noi pentru terminalele din cuvinte de lungime >1 .

[pas I din alg. curs – ($A_y \rightarrow y$) din curs corespunde cu ($X_y \rightarrow y$) din exemplu seminar pag 4]

Vrem ca terminalele să apară doar singure în membrul drept. De aceea, peste tot unde apar în componența unui cuvânt de lungime minim 2, le vom înlocui cu un neterminal nou și vom adăuga producția de la neterminalul nou la terminalul pe care l-a înlocuit.

Pas 6: Se adaugă neterminale noi pentru „spargerea” cuvintelor de lungime >2 .

[pas II din alg. curs – neterminalele noi Z_i din curs corespund cu Y_i din ex. seminar pag 4]

Vrem ca în dreapta să avem cuvinte formate din exact două neterminale. De aceea, unde avem cuvinte mai lungi, păstrăm doar primul neterminal din cuvânt și îi alipim un neterminal nou, iar noul neterminal va avea o producție cu membrul drept cuvântul pe care l-a înlocuit.

Reluăm procedeul până când toate cuvintele ajung la lungimea 2.

Obs: Fiecare producție care avea un cuvânt de lungime k va fi înlocuită de $k-1$ producții cu cuvinte de lungime 2.

Exemplu:

Avem următoarea gramatică independentă de context:

$$S \rightarrow a^{(1)}B^{(2)}c^{(3)}DeF \mid HF \mid HBc$$

$$B \rightarrow b^{(4)} \mid \lambda^{(5)}$$

$$D \rightarrow d^{(6)} \mid \lambda^{(7)}$$

$$F \rightarrow G^{(8)}$$

$$G \rightarrow f^{(9)} \mid g^{(10)}$$

$$H \rightarrow \lambda^{(11)}$$

→ Aplicare pas 2 (i)

Neterminalul H este singurul de acest gen, având doar λ -producția (11), care va fi eliminată. În plus, producția (2) va fi înlocuită cu „ $S \rightarrow F$ ” (12), iar producția (3) va fi înlocuită cu „ $S \rightarrow Bc$ ” (13).

$$S \rightarrow a^{(1)}B^{(12)}c^{(13)}DeF \mid F \mid Bc$$

$$B \rightarrow b^{(4)} \mid \lambda^{(5)}$$

$$D \rightarrow d^{(6)} \mid \lambda^{(7)}$$

$$F \rightarrow G^{(8)}$$

$$G \rightarrow f^{(9)} \mid g^{(10)}$$

→ Aplicare pas 2 (ii)

Neterminalele B și D au mai multe producții printre care și câte o λ -producție. Deci (5) și (7) vor fi eliminate.

În producția (13) apare doar B, deci va fi înlocuită cu „ $S \rightarrow Bc$ ” (14) și „ $S \rightarrow c$ ” (15).

În producția (1) apar și B și D, deci va fi înlocuită de toate variațiile posibile:

-- B și D prezente: „ $S \rightarrow aBcDeF$ ” (16)

-- B lipsă, D prezent: „ $S \rightarrow acDeF$ ” (17)

-- B prezent, D lipsă: „ $S \rightarrow aBceF$ ” (18)

-- B și D lipsă: „ $S \rightarrow aceF$ ” (19)

$$S \rightarrow a^{(16)}B^{(17)}c^{(18)}DeF \mid ac^{(19)}DeF \mid a^{(11)}B^{(14)}c^{(15)}eF \mid F \mid Bc \mid c$$

$$B \rightarrow b^{(4)} \quad ; \quad D \rightarrow d^{(6)} \quad ; \quad F \rightarrow G^{(8)} \quad ; \quad G \rightarrow f^{(9)} \mid g^{(10)}$$

→ Aplicare pas 3

Avem două șiruri de redenumiri:

$$F \rightarrow G^{(8)} \rightarrow f^{(9)} \mid g^{(10)}$$

Producția (8) va fi înlocuită cu „ $F \rightarrow f$ ” (20) și „ $F \rightarrow g$ ” (21).

Neterminalul G nu mai apare nicăieri în dreapta, deci vom elimina (9) și (10).

$$S \rightarrow F^{(11)} \rightarrow f^{(20)} \mid g^{(21)}$$

Producția (11) va fi înlocuită cu „ $S \rightarrow f$ ” (22) și „ $S \rightarrow g$ ” (23).

Dar neterminalul F mai apare în dreapta în producțiile 16–19, deci NU vom elimina producțiile lui F.

$$S \rightarrow a^{(16)}B^{(17)}c^{(18)}DeF \mid ac^{(19)}DeF \mid a^{(22)}B^{(23)}c^{(14)}eF \mid f \mid g \mid Bc \mid c$$

$$B \rightarrow b^{(4)} \quad ; \quad D \rightarrow d^{(6)} \quad ; \quad F \rightarrow f^{(20)} \mid g^{(21)}$$

→ Aplicare pas 5

Producțiile 22, 23, 15, 4, 6, 20, 21 au deja forma dorită.

Producția (16) va fi înlocuită cu „ $S \rightarrow X_a B X_c D X_e F$ ” (24).

Producția (17) va fi înlocuită cu „ $S \rightarrow X_a X_c D X_e F$ ” (25).

Producția (18) va fi înlocuită cu „ $S \rightarrow X_a B X_c X_e F$ ” (26).

Producția (19) va fi înlocuită cu „ $S \rightarrow X_a X_c X_e F$ ” (27).

Producția (14) va fi înlocuită cu „ $S \rightarrow B X_c$ ” (28).

În plus, vom avea „ $X_a \rightarrow a$ ” (29); „ $X_c \rightarrow c$ ” (30); „ $X_e \rightarrow e$ ” (31).

$$S \rightarrow X_a B X_c D X_e F \mid X_a X_c D X_e F \mid X_a B X_c X_e F \mid X_a X_c X_e F \mid f \mid g \mid B X_c \mid c$$

$$B \xrightarrow{(4)} b \quad ; \quad D \xrightarrow{(6)} d \quad ; \quad F \xrightarrow{(20)} f \mid g \xrightarrow{(21)}$$

$$X_a \xrightarrow{(29)} a \quad ; \quad X_c \xrightarrow{(30)} c \quad ; \quad X_e \xrightarrow{(31)} e$$

→ Aplicare pas 6

Producția (24) va deveni „ $S \rightarrow X_a Y_1$ ” (32), iar $Y_1 \sim B X_c D X_e F$.

Apoi „ $Y_1 \rightarrow B Y_2$ ” (33), iar $Y_2 \sim X_c D X_e F$.

Apoi „ $Y_2 \rightarrow X_c Y_3$ ” (34), iar $Y_3 \sim D X_e F$.

Apoi „ $Y_3 \rightarrow D Y_4$ ” (35), iar $Y_4 \sim X_e F$.

În final „ $Y_4 \rightarrow X_e F$ ” (36).

Observăm că putem avea sufixe comune ale cuvintelor și atunci am putea folosi neaterminale deja adăugate anterior.

Producția (25) va deveni „ $S \rightarrow X_a Y_2$ ” (37), pentru că aveam $Y_2 \sim X_c D X_e F$.

Producția (26) va deveni „ $S \rightarrow X_a Y_5$ ” (38), iar $Y_5 \sim B X_c X_e F$.

Apoi „ $Y_5 \rightarrow B Y_6$ ” (39), iar $Y_6 \sim X_c X_e F$.

Apoi „ $Y_6 \rightarrow X_c Y_4$ ” (40), pentru că aveam $Y_4 \sim X_e F$.

Producția (27) va deveni „ $S \rightarrow X_a Y_6$ ” (41), pentru că aveam $Y_6 \sim X_c X_e F$.

$$S \rightarrow X_a Y_1 \mid X_a Y_2 \mid X_a Y_5 \mid X_a Y_6 \mid f \mid g \mid B X_c \mid c$$

$$B \xrightarrow{(4)} b \quad ; \quad D \xrightarrow{(6)} d \quad ; \quad F \xrightarrow{(20)} f \mid g \xrightarrow{(21)}$$

$$X_a \xrightarrow{(29)} a \quad ; \quad X_c \xrightarrow{(30)} c \quad ; \quad X_e \xrightarrow{(31)} e$$

$$Y_1 \xrightarrow{(33)} B Y_2 \quad ; \quad Y_2 \xrightarrow{(34)} X_c Y_3 \quad ; \quad Y_3 \xrightarrow{(35)} D Y_4 \quad ; \quad Y_4 \xrightarrow{(36)} X_e F \quad ; \quad Y_5 \xrightarrow{(39)} B Y_6 \quad ; \quad Y_6 \xrightarrow{(40)} X_c Y_4$$

EXERCITII:

EX_1: Aplicați algoritmul de reducere (pasul 1) pe gramatica:

$$S \rightarrow A \mid B$$

$$A \rightarrow aB \mid bS \mid b$$

$$B \rightarrow AB \mid Ba$$

$$C \rightarrow AS \mid b$$

EX_2: Aplicați tot algoritmul de transformare în F.N.Chomsky pe gramatica:

$$S \rightarrow aABa \mid CD \mid bbAC$$

$$A \rightarrow bc \mid d \mid \lambda$$

$$B \rightarrow \lambda \mid E$$

$$C \rightarrow A \mid dcabb \mid S$$

$$D \rightarrow BAd \mid B$$

$$E \rightarrow Ea \mid bbE$$

$$F \rightarrow abc$$