1. Clasificarea Chomsky a gramaticilor

În funcție de forma producțiilor, lingvistul **Noam Chomsky**, în 1958, a ierarhizat gramaticile astfel:

Tip	**Denumire**	**Forma producțiilor**
0	gramatică generală	oarecare
1	gramatică dependentă de context	$p \to q$ cu $\|p\| \leq \|q\|$ cu p conţinând cel puţin un neterminal
2	gramatică independentă de context	$A \to y$ cu $A \in \boldsymbol{N}$ și $y \in (\boldsymbol{N} \cup \boldsymbol{T})^*$
3	gramatică regulată	$A \to aB$ sau $A \to a$ cu $A, B \in \boldsymbol{N}$ şi $a \in \boldsymbol{T}^*$

Observația 1. Fie $\boldsymbol{\mathcal{G}_i}$ familia gramaticilor de tip $\boldsymbol{i}$ $(\boldsymbol{i} = \overline{0,3})$. Se observă că

$$\boldsymbol{\mathcal{G}_3} \subset \boldsymbol{\mathcal{G}_2} \subset \boldsymbol{\mathcal{G}_1} \subset \boldsymbol{\mathcal{G}_0},$$

adică orice gramatică se numește de tipul ***0***.

Exemplul 1. Fie gramatica $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$. În tabelul de mai jos este indicat tipul gramaticii $\boldsymbol{G}$ în funcție de mulțimea producțiilor $\boldsymbol{P}$:

$\boldsymbol{N} = \{S, X\}$, $\boldsymbol{T} = \{a, b, c\}$	$\boldsymbol{N} = \{S, A, B\}$, $\boldsymbol{T} = \{a, b, c\}$	$\boldsymbol{N} = \{S, A, B\}$, $\boldsymbol{T} = \{a, b\}$	$\boldsymbol{N} = \{X\}$, $\boldsymbol{T} = \{a, b\}$
P			
$S \to abc$ (1) $S \to aSXc$ (2) $cX \to Xc$ (3) $bX \to bb$ (4)	$S \to aaAc$ (1) $aAc \to aAbBc$ (2) $bB \to bBc$ (3) $Bc \to Abc$ (4) $A \to a$ (5)	$A \to a$ (1) $B \to b$ (2) $S \to aA$ (3) $A \to bB$ (4) $B \to bB$ (5)	$X \to aX$ (1) $X \to bX$ (2) $X \to \varepsilon$ (3)
Gramatică generală (tip *0*)	**Gramatică dependentă de context (tip *1*)**	**Gramatică independentă de context (tip *2*)**	**Gramatică regulată (tip *3*)**

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

2. **Clasificarea limbajelor**

Definiția 1.1. Un limbaj $\boldsymbol{L}$ este de tip $\boldsymbol{i}$ $(\boldsymbol{i} = \overline{0,3})$ dacă există o gramatică $\boldsymbol{G}$ de tipul $\boldsymbol{i}$ pentru care $\boldsymbol{L}(\boldsymbol{G}) = \boldsymbol{L}$.

Observația 1.1. Fie $\boldsymbol{\mathcal{L}_i}$ familia limbajelor de tip $\boldsymbol{i}$ $(\boldsymbol{i} = \overline{0,3})$. Se observă că

$$\boldsymbol{\mathcal{L}_3} \subset \boldsymbol{\mathcal{L}_2} \subset \boldsymbol{\mathcal{L}_1} \subset \boldsymbol{\mathcal{L}_0}.$$

Incluziunile sunt stricte:

- Orice limbaj de tip ***i*** + 1 este și de tip $\boldsymbol{i} = \overline{0,2}$
- Există limbaje de tip ***i*** care nu sunt de tip $\boldsymbol{i} + 1, \boldsymbol{i} = \overline{0,2}$.

Propoziția 1.1. Fie $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$ o gramatică de tipul $\boldsymbol{i}$ $(\boldsymbol{i} = \overline{1,3})$. Atunci există o gramatică $\boldsymbol{G'}$ echivalentă cu $\boldsymbol{G}$ și de același tip, cu proprietatea că simbolul inițial S nu apare în membrul drept al producțiilor.

Demonstrație:

Fie S' un simbol nou, adică $S' \notin \boldsymbol{N} \cup \boldsymbol{T}$. Construim gramatica $\boldsymbol{G'} = (\boldsymbol{N} \cup \{\boldsymbol{S'}\}, \boldsymbol{T}, \boldsymbol{S'}, \boldsymbol{P'})$, unde $\boldsymbol{P'} = \boldsymbol{P} \cup \{S' \to \alpha | S \to \alpha\}$. Se observă ușor că $\boldsymbol{L}(\boldsymbol{G}) = \boldsymbol{L}(\boldsymbol{G'})$.

Observația 1.2. Din forma producțiilor pentru gramaticile de tipul $\mathbf{1}, \mathbf{2}$ sau $\mathbf{3}$ rezultă că ε nu aparține limbajului generat de ele. Dacă dorim ca și cuvântul vid ε să aparțină limbajului generat de o gramatică, admitem în mod excepțional producția $S \to \varepsilon$, care nu poate avea alte repercusiuni, conform propoziției 1.1. precedente.

3. Metoda șirului crescător de mulțimi

Fie *A* o mulţime finită şi fie *p* o proprietate definită pe mulţimea submulţimilor lui A.
Fie *X* o submulţime a lui *A* cu *p*(*X*)=1. Definim următorul şir crescător de mulţimi:

$$X_0 = X$$
$$X_{k+1} = X_k \cup \{x \in A \mid p(X_k \cup \{x\}) = 1\}, \forall k>0$$

Evident, $X_0 \subset X_1 \subset ... \subset X_k \subset X_{k+1} \subset ... \subset A$.

Cum *A* este finită, şirul de submulţimi se va stabiliza.

Propoziţia 1.1.1. Dacă $X_{k+1} = X_k$, atunci $X_{k+i} = X_k$, $\forall i \in \mathbb{N}$.

Demonstrație: Se folosește metoda inducţiei matematice după *i*.

Pentru *i* = 1 rezultatul este evident.

Presupunem $X_{k+i} = X_k$ şi demonstrăm că $X_{k+i+1} = X_k$:

X_{k+i+1} = (conform definiţiei şirului de mulţimi)

= $X_{k+i} \cup \{x \in A \mid p(X_{k+i} \cup \{x\}) = 1\}$ = (conform ipotezei de inducţie)
= $X_k \cup \{x \in A \mid p(X_k \cup \{x\}) = 1\}$ = (conform definiţiei şirului de mulţimi)
= X_{k+1} = (conform ipotezei)
= X_k.

Consecința 1.1.1. Se oprește construcţia şirului crescător de mulţimi la primul *k* pentru care se obține $X_k = X_{k+1}$.

4. Problema apartenenței

Teorema 1.2.1. Pentru gramaticile de tipurile $\mathbf{1}, \mathbf{2}$ şi $\mathbf{3}$ este posibil să verificăm apartenenţa unui cuvânt la limbajul generat de ele.

Demonstrație:

Fie $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$ o gramatică de tipul $\mathbf{1}, \mathbf{2}$ sau $\mathbf{3}$ şi fie $w \in \boldsymbol{T}^*$. Dorim să verificăm dacă

$w \in \boldsymbol{L}(\boldsymbol{G})$.

Fie $n = |w|$ şi folosim metoda şirului crescător de mulţimi:

$$\begin{cases} T_0 & = & \{S\} \\ T_{k+1} & = & T_k \cup \left\{\alpha \in (\boldsymbol{N} \cup \boldsymbol{T})^* \middle| \exists \beta \in T_k \text{ cu } \beta \overset{*}{\Rightarrow} \alpha \text{ şi } |\alpha| \leq n\right\} \end{cases}$$

Deoarece $T_0 \subset T_1 \subset \ldots \subset T_k \subset T_{k+1} \subset \ldots \subset F$, unde F este mulţimea finită a cuvintelor de lungime cel mult n formate din simboluri terminale şi neterminale, rezultă că şirul se stabilizează, respectiv $\exists k_0 \in \mathbb{N}$ astfel încât $T_{k_0} = T_{k_0+1} = T_{k_0+2} = \cdots$.

Evident, $w \in \boldsymbol{L}(\boldsymbol{G}) \Leftrightarrow w \in T_{k_0}$.

Exemplul 1.2.1. Fie gramatica regulată $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$, unde

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

$S \to aS$ (1)

$S \to bA$ (2)

$A \to cA$ (3)

$A \to c$ (4)

Cuvântul $w_1 = abc \in \boldsymbol{L}(\boldsymbol{G})$, deoarece se poate obține printr-o derivare din simbolul inițial S astfel:

$$\boldsymbol{S} \underset{(1)}{\overset{*}{\Rightarrow}} a\boldsymbol{S} \underset{(2)}{\overset{*}{\Rightarrow}} ab\boldsymbol{A} \underset{(4)}{\overset{*}{\Rightarrow}} abc = w_1 \in \boldsymbol{L}(\boldsymbol{G})$$

Cuvântul $w_2 = aaabcccc \in \boldsymbol{L(G)}$**,** deoarece se poate obține printr-o derivare din simbolul inițial S astfel:

$$\boldsymbol{S} \underset{(1)}{\overset{*}{\Rightarrow}} a\boldsymbol{S} \underset{(1)}{\overset{*}{\Rightarrow}} aa\boldsymbol{S} \underset{(1)}{\overset{*}{\Rightarrow}} aaa\boldsymbol{S} \underset{(2)}{\overset{*}{\Rightarrow}} aaab\boldsymbol{A} \underset{(3)}{\overset{*}{\Rightarrow}} aaabc\boldsymbol{A} \underset{(3)}{\overset{*}{\Rightarrow}} aaabcc\boldsymbol{A} \underset{(3)}{\overset{*}{\Rightarrow}} aaabccc\boldsymbol{A} \underset{(4)}{\overset{*}{\Rightarrow}} aaabcccc = a^3bc^4.$$

Deci $w_2 \in \boldsymbol{L(G)}$.

Limbajul generat de gramatica $\boldsymbol{G}$ este $\boldsymbol{L(G)} = \{a^n bc^m | n \geq 1, m \geq 1\}$ deoarece se pornește de la simbolul de start *S* și până când *S* "dispare", aplicând de *n* ori producția (1) cuvântul curent are forma $a^n S$.

După dispariția lui *S (*aplicând producția (2)), apare neterminalul *A* care urmează lui *b*. Aplicând de *m-1* ori producția (3), neterminalul *A* este deplasat, de fiecare dată, la dreapta după terminalul *c* și cuvântul curent are forma $a^n b c^{m-1} A$. Rezultă că *A* devine *c,* conform (4), iar cuvântul, format numai din terminale, este $a^n bc^m$.

Exemplul 1.2.2. Fie gramatica regulată $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$, unde

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

$S \to aA$ (1)

$A \to aA \mid aB$ (2)

$B \to bC$ (3)

$C \to cB \mid c$ (4)

Cuvântul $w = aabc \in \boldsymbol{L(G)}$ deoarece:

$$\boldsymbol{S} \underset{(1)}{\overset{*}{\Rightarrow}} a\boldsymbol{A} \underset{(2)}{\overset{*}{\Rightarrow}} aa\boldsymbol{B} \underset{(3)}{\overset{*}{\Rightarrow}} aab\boldsymbol{C} \underset{(4)}{\overset{*}{\Rightarrow}} aabc = w \in \boldsymbol{L(G)}$$

Se poate observa că limbajul generat de gramatica $\boldsymbol{G}$ este

$\boldsymbol{L(G)} = \{a^n (bc)^m | n \geq 2, m \geq 1\}$ (**temă**).

Exemplul 1.2.3. Fie gramatica $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$, $\boldsymbol{N} = \{S\}$**,** $\boldsymbol{T} = \{a, b\}$**,** iar mulțimea producțiilor ***P*** constă din:

$S \to \varepsilon$ (1)

$S \to a$ (2)

$S \to b$ (3)

$S \to aSa$ (4)

$S \to bSb$ (5).

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

Cuvântul $w = aababaa \in \boldsymbol{L(G)}$ deoarece

$$\boldsymbol{S} \underset{(4)}{\overset{*}{\Rightarrow}} a\boldsymbol{S}a \underset{(4)}{\overset{*}{\Rightarrow}} aa\boldsymbol{S}aa \underset{(5)}{\overset{*}{\Rightarrow}} aab\boldsymbol{S}baa \underset{(2)}{\overset{*}{\Rightarrow}} aababaa = w.$$

Se poate observa că limbajul generat de gramatica G este format din toate palindroamele formate din literele a și b.

Exemplul 1.2.4. Fie gramatica $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$, $\boldsymbol{N} = \{S, B, C\}$, $\boldsymbol{T} = \{a, b, c\}$, iar mulțimea producțiilor ***P*** constă din:

$S \rightarrow aSBC$ (1)

$S \rightarrow aBC$ (2)

$CB \rightarrow BC$ (3)

$aB \rightarrow ab$ (4)

$bB \rightarrow bb$ (5)

$bC \rightarrow bc$ (6)

$cC \rightarrow cc$ (7)

Fie cuvântul *w* = *aabbcc* = $a^2b^2c^2$. Cuvântul $w \in \boldsymbol{L(G)}$, deoarece se poate obține printr-o derivare din simbolul inițial S astfel:

$$\boldsymbol{S} \underset{(1)}{\overset{*}{\Rightarrow}} a\boldsymbol{S}BC \underset{(2)}{\overset{*}{\Rightarrow}} a\boldsymbol{aB}CBC \underset{(4)}{\overset{*}{\Rightarrow}} aab\boldsymbol{CB}C \underset{(3)}{\overset{*}{\Rightarrow}} aa\boldsymbol{bB}CC \underset{(5)}{\overset{*}{\Rightarrow}} aab\boldsymbol{bC}C \underset{(6)}{\overset{*}{\Rightarrow}} aabb\boldsymbol{cC} \underset{(7)}{\overset{*}{\Rightarrow}} aabbcc = a^2b^2c^2 .$$

Rezultă că $w \in \boldsymbol{L(G)}$.

Se demonstrează că $\boldsymbol{L(G)}$ = {$a^n b^n c^n$ | *n*>0} astfel:

se pornește de la simbolul de start *S* și până când *S* ”dispare”, cuvântul curent are forma $a^n S \alpha$, unde în α există:

- *n* de *B* și *b*
- *n* de *C* și *c.*

După dispariția lui *S*:

- toate *C*-urile sunt deplasate la dreapta după *b* -uri, conform producției (3)
- toate *B* -urile care urmează lui *a* sau *b* trec în *b*, conform producțiilor (4) și (5)
- toate *C* -urile care urmează lui *b* sau *c* trec în *b*, conform (6) și (7),

și se obține cuvântul $a^n b^n c^n, n > 0$, format numai din terminale.

Temă:

1. Fie gramatica $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$, $\boldsymbol{N} = \{S\}$**,** $\boldsymbol{T} = \{a, b\}$**,** iar mulțimea producțiilor ***P*** constă din:

$$
\begin{array}{ll}
S \rightarrow a & (1) \\
S \rightarrow b & (2) \\
S \rightarrow aSa & (3) \\
S \rightarrow bSb & (4).
\end{array}
$$

   De ce tip este gramatica ***G***? Demonstrați că $w = ababa \in \boldsymbol{L}(\boldsymbol{G})$.

2. Fie gramatica $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$, $\boldsymbol{N} = \{S\}$**,** $\boldsymbol{T} = \{(,)\}$**,** iar mulțimea producțiilor ***P*** constă din:

$$
\begin{array}{ll}
S \rightarrow () & (1) \\
S \rightarrow (S) & (2) \\
S \rightarrow SS & (3).
\end{array}
$$

   De ce tip este gramatica ***G***?
   Demonstrați că

$$w_1 = ((\ \ ))() \in \boldsymbol{L}(\boldsymbol{G}), \text{iar } w_2 = ())(() \notin \boldsymbol{L}(\boldsymbol{G}) .$$

3. Fie gramatica $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$, $\boldsymbol{N} = \{S, A, B\}$**,** $\boldsymbol{T} = \{a, b\}$**,** iar mulțimea producțiilor ***P*** este indicată în tabelul de mai jos. Completați pe ultima linie a tabelului tipul fiecărei gramatici.

P			
$A \rightarrow a$ $A \rightarrow b$ $S \rightarrow aA$ $A \rightarrow bB$ $B \rightarrow A$	$A \rightarrow a$ $B \rightarrow b$ $S \rightarrow aA$ $A \rightarrow aB$ $B \rightarrow bA$	$A \rightarrow a$ $B \rightarrow b$ $S \rightarrow aAa$ $A \rightarrow bBb$ $B \rightarrow A$	$A \rightarrow a$ $B \rightarrow b$ $S \rightarrow aA$ $aA \rightarrow bbbB$ $bbB \rightarrow A$

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