

Curs – sapt.3

- Gramatici. Clasificarea Chomsky
 - Gramatici regulare
- Limbaje regulare. Echivalente
 - Expresii regulare
- Proprietati ale limbajelor regulare

Gramatica

O gramatica este un cvadruplu $\mathbf{G} = (\mathbf{N}, \Sigma, P, S)$

- \mathbf{N} este un alfabet de simboluri ***neterminale***
- Σ este un alfabet de simboluri ***terminale***
- $\mathbf{N} \cap \Sigma = \emptyset$
- $P \subseteq (\mathbf{N} \cup \Sigma)^* \mathbf{N} (\mathbf{N} \cup \Sigma)^* \times (\mathbf{N} \cup \Sigma)^*$
 - P multime finită (multimea regulilor de productie)
 - $S \in \mathbf{N}$ (simbolul de start - simbolul initial)

Notatie:

$(\alpha, \beta) \in P$ se noteaza: $\alpha \rightarrow \beta$
(α se înlocuieste cu β)

Notatii

- la nivel abstract (exemple matematice, specificari)
 - Σ : a,b,... litere mici de la inceputul alfabetului
 - N: A,B,.. litere mari de la inceputul alfabetului
 - Σ sau N: X,Y,...litere mari de la sfarsitul alfabetului
 - Σ^* : x,y,... litere mici de la sfarsitul alfabetului
 - $(\Sigma \cup N)^*$: α, β, \dots litere grecesti
- nu se folosesc spatii cand avem nevoie de mai multe caractere pentru a specifica un simbol (terminal sau neterminat)

Relatii de derivare

relatii binare peste $(\Sigma \cup N)^*$ adica $(\Sigma \cup N)^* \times (\Sigma \cup N)^*$

- derivare directă

$$\gamma \Rightarrow \delta \Leftrightarrow \exists \gamma_1, \gamma_2, \alpha, \beta \in (N \cup \Sigma)^*$$

a.i. $\gamma = \gamma_1 \alpha \gamma_2$, $\delta = \gamma_1 \beta \gamma_2$, iar $(\alpha \rightarrow \beta) \in P$

- k-derivare

$\overset{k}{\Rightarrow}$

(o succesiune de k derivări directe)

- + derivare

$\overset{+}{\Rightarrow}$

dacă $\exists k > 0$ a.i. cele 2 secvențe să fie într-o relație de "k derivare"

- * derivare

$\overset{*}{\Rightarrow}$

dacă fie cele 2 secvențe sunt egale, fie între ele există o relație de +derivare

Limbaj generat de o gramatica

- Limbaj generat gramatica $G=(N, \Sigma, P, S)$
$$L(G) = \{ w \in \Sigma^* \mid S \xrightarrow{*} w \}$$
- Forma propositională
– $\alpha \in (N \cup \Sigma)^*$ a.i. $S \xrightarrow{*} \alpha$
- Propozitie (cuvant)
– un element din $L(G)$
- Gramatici echivalente
daca genereaza acelasi limbaj

Tipuri de gramatici

- Gramatica monotonă
 - $\forall \alpha \rightarrow \beta \in P: |\alpha| \leq |\beta| \quad \alpha, \beta \in (N \cup \Sigma)^*$
 - caz special: $S \rightarrow \epsilon$ poate să aparțină lui P. În acest caz S nu apare în membrul drept al nici unei reguli de producție.
- Gramatica dependenta de context
reguli de producție sunt de forma:
$$\alpha A \beta \rightarrow \alpha \gamma \beta \quad A \in N$$
$$\alpha, \beta, \gamma \in (N \cup \Sigma)^*, \gamma \neq \epsilon$$
 - caz special: $S \rightarrow \epsilon$ poate să aparțină lui P. În acest caz S nu apare în membrul drept al nici unei reguli de producție.

Tipuri de gramatici

- Gramatica regulara:

reg. prod. sunt de forma

- $A \rightarrow aB$

- $A \rightarrow b$

unde $A, B \in N$ si $a, b \in \Sigma$

caz special: $S \rightarrow \epsilon$ poate $\in P$. In acest caz S nu apare în membrul drept al nici unei reguli de productie.

- Gramatica independenta de context:

reg. productie sunt de forma $A \rightarrow \alpha$, $A \in N$, $\alpha \in (N \cup \Sigma)^*$

Clasificarea Chomsky

- Gramatici de tip **0**
nici o restrictie (*suplimentara*) referitoare la forma regulilor de productie
- Gramaticile de tip **1**
dependente de context \Leftrightarrow *gramatici monotone*
(*monotonic, non-contracting*)
- Gramaticile de tip **2**
gramatici independente de context
- Gramaticile de tip **3**
gramatici regulate
=> Limbaje independente de context
=> Limbaje regulate

Ierarhia Chomsky

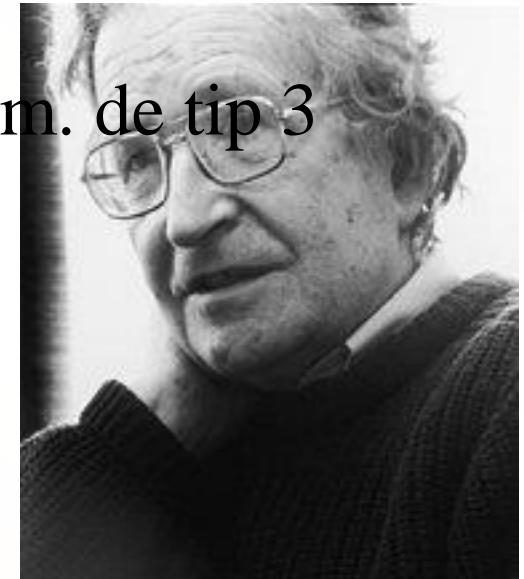
Fie

~ 1959-1963

- \mathcal{L}_0 - multimea limbajelor generate de gram. de tip 0
- \mathcal{L}_1 - multimea limbajelor generate de gram. de tip 1
- \mathcal{L}_2 - multimea limbajelor generate de gram. de tip 2
- \mathcal{L}_3 - multimea limbajelor generate de gram. de tip 3

Are loc:

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



Ierarhia Chomsky: observatii

Teorema:

Fiecare dintre familiile de limbaje:

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

este inchisa fata de operatia de reuniiune

Limbaje regulare. Echivalente

- Limbaj regular
 - = limbaj generat de o gramatica regulara
- putere de exprimare
 - AF: AFN \leftrightarrow AFD
 - AF \leftrightarrow gr.regulare
 - AF \leftrightarrow (m.regulare \leftrightarrow expr.reg.)

Multimi regulare

Fie Σ un alfabet.

Multimile regulare peste Σ se definesc recursiv astfel:

1. Φ este o m. reg. peste Σ
2. $\{\varepsilon\}$
3. $\{a\}$ daca: $a \in \Sigma$
4. RUS daca R, S – multim regulare peste Σ +
5. RS daca R, S – multim regulare peste Σ
6. R^* daca R – multime regulara peste Σ
7. Orice alta multime regulara se obtine aplicand de un numar finit de ori reg. 1-6

Multimi regulare si expresii regulare

- Expresii regulare

1.	Φ	expr. reg. coresp. m.reg.	Φ
2.	ϵ		$\{\epsilon\}$
3.	a	daca: $a \in \Sigma$	$\{a\}$
4.	$r+s$	daca r,s – expresii regulare	$r s$
5.	rs	daca r,s – expresii regulare	RS
6.	r^*	daca r – expresie regulara	R^*
7.	Orice alta expr. reg. se obtine aplicand de un numar finit de ori reg. 1-6		

- Expresii regulare echivalente:

- mult. regulare reprezentate de acestea sunt egale

Expresii regulare

- expresiile regulare – secv. obtinute prin concatenarea de simb. din
$$\Sigma \cup \{\Phi, \varepsilon, +, *, (,)\} \quad (\dots \text{prioritate} \dots)$$
- multimile regulare asociate expresiilor regulare sunt limbaje regulare

Deci: *Orice expresie regulară peste Σ descrie un limbaj regular peste Σ*

Proprietati de inchidere ale limbajelor regulare

Teorema:

Daca

L_1, L_2 sunt limbaje regulare peste alfabetul Σ

atunci:

$L_1 \cup L_2, L_1 \cap L_2, L_1 L_2, L_1^*, \text{complement}(L_1)$
sunt limbaje regulare peste alfabetul Σ

Lema de pompare pt. limbaje regulare

- Daca L este un limbaj regular,
- atunci $\exists p \in \mathbf{N}^*$ (fix pt. un limbaj dat)
(oricat de mare)
- astfel incat:
 $\forall w \in L$ de lungime cel putin p
exista o descompunere de forma $w=xyz$,
unde $0 < |y| \leq p$
cu proprietatea ca: $xy^i z \in L, \forall i \in \mathbf{N}$

Lema de pompare pt. limbaje regulare

$(\forall L \subseteq \Sigma^*)$

$(\text{regular}(L) \Rightarrow$

$((\exists p \geq 1)((\forall w \in L)((|w| \geq p) \Rightarrow$

$((\exists x, y, z \in \Sigma^*)(w = xyz \wedge (|y| \geq 1 \wedge |y| \leq p \wedge (\forall n \geq 0)(xy^n z \in L)))))))$

(enunt formal al teoremei)

Lema de pompare pt. limbaje regulare

(o alta versiune, mai “puternica”)

Daca L este un limbaj regular,

- atunci $\exists p \in \mathbb{N}^*$ (fix pt. un limbaj dat)
(oricat de mare)
- astfel incat:
 $\forall w \in L$ de lungime cel putin p
exista o descompunere de forma $w=xyz$ astfel incat
 - $0 < |y|$
 - $|x y| \leq p$
 - $xy^i z \in L, \forall i \in \mathbb{N}$

Lema de pompare pt. limbaje regulare

Observatii:

- Lema da o conditie necesara dar nu suficiente
- daca un limbaj satisface conditiile lemei nu inseamna ca este regular
- folosim negatia lemei de pompare pt. a dem. ca un limbaj nu este regular

Lema de pompare pt. limbaje regulare

De ce se intampla asa:

- Daca L – limb. reg.

\Rightarrow exista G – gram. reg. a.i. $L(G) = L$ (def.)

\Rightarrow exista M – AF a.i. $L(M) = L$ (teorema)

- Fie p – nr. de stari ale lui M
- daca $|w| \geq p$ si w – acceptat

$\Rightarrow \exists$ un drum in graful asociat lui M a.i. etichetele arcelor sunt simboluri din w

\Rightarrow drumul este de lungimea p ; adica trece prin $p + 1$ noduri din graf

$\Rightarrow \exists$ un nod prin care se trece de cel putin 2 ori

\Rightarrow ciclu/bucla – care se poate repeta de oricate ori !!

\Rightarrow se poate repeta sirul etichetelor arcelor din bucla !!

(de 0 sau mai multe ori)

Exemplu:

Fie L - limbajul regular corespunzator expresiei regulare:

aa^*b^*

1) fie $w = ab$;

Puteti identifica o descompunere $w=xyz$ a.i. xy^iz in L ?

2) fie $w = aa$;

Puteti identifica o descompunere $w=xyz$ a.i. xy^iz in L ?

Analog pt.: $a(ba)^*$

si $w = aba$

Analog pt.: $L=\{a,b\}$ si $w = a$

Proprietati de inchidere ale limbajelor regulate

$L_1 \cap L_2$

- $M_1 = (Q_1, \Sigma, \delta_1, q_{01}, F_1)$
- $M_2 = (Q_2, \Sigma, \delta_2, q_{02}, F_2)$
- ? $M = (Q, \Sigma, \delta, q_0, F)$

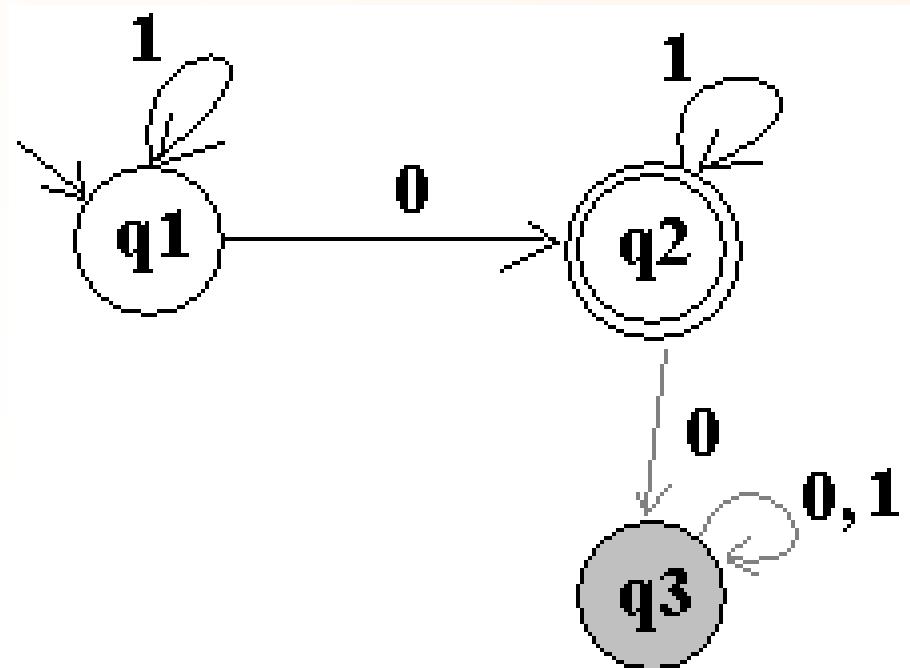
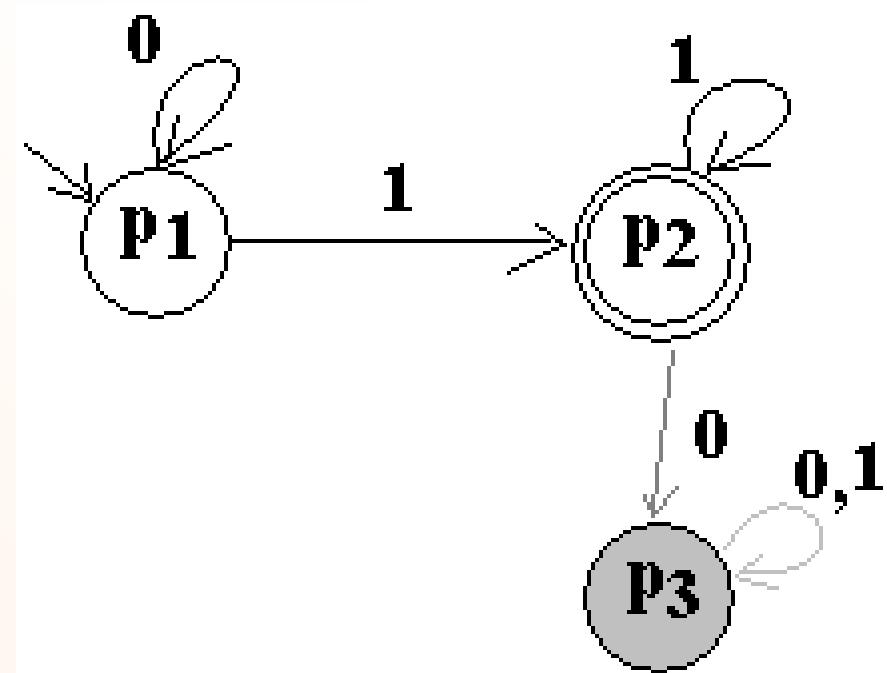
Obs.:

A fost facut exemplu
pentru intersectie in sapt.14

PP. ca aut. M_1 si M_2 sunt deterministe, complet definite !

(alg. de constr. !!)

- $M = (Q_1 \times Q_2, \Sigma, \delta, (q_{01}, q_{02}), F_1 \times F_2)$
- $\delta((q_1, q_2), a) = (\delta_1(q_1, a), \delta_2(q_2, a))$



Proprietati de inchidere ale limbajelor regulate

complement(L_1)

- $M_1 = (Q_1, \Sigma, \delta_1, q_{01}, F_1)$
- $?M = (Q, \Sigma, \delta, q_0, F)$

PP. ca aut. M_1 este determinist complet definit !
(alg. de constr.)

- $M_1 = (Q_1, \Sigma, \delta_1, q_{01}, Q_1 - F_1)$