LFA CURS 10

LEMA DE POMPARE PENTRU LIMBAJE INDEPENDENTE DE CONTEXT

În cele ce urmează vom folosi notația $\mathcal{L}_{CF}$ pentru familia limbajelor independente de context.

Întocmai ca în cazul limbajelor regulate, există o lemă de pompare pentru limbajele independente de context.

Informal, pentru fiecare $L \in \mathcal{L}_{CF}$, fiecare șir suficient de lung din L poate fi descompus în 5 subșiruri astfel încât cele 3 segmente de la mijloc au o lungime limitată (de o constantă care depinde de L), al 2-lea și al 4-lea subșir nu sunt ambele nule, și, oricâte cópii ale șirurilor al 2-lea și al 4-lea am „pompa", șirul obținut rămâne în L.

Diferența față de limbajele regulate este că în cazul lemei de pompare pentru limbajele independente de context se „pompează" simultan 2 subșiruri în loc de unul singur.

<u>Teoremă</u> Pentru fiecare limbaj $L \in \mathcal{L}_{CF}$ există $k \geq 1$ astfel încât fiecare $z \in L$, $|z| \geq k$, poate fi descompus în 5 subșiruri astfel:

(i) $z = uvwxy$

(ii) $vx \neq \lambda$ (sau $|vx| \geq 1$)

(iii) $|vwx| \leq k$

(iv) $uv^t wx^t y \in L$, $\forall\, t \geq 0$

<u>Demonstrație</u> Vom considera o gramatică independentă de context în forma normală Chomsky, $G = (N, \Sigma, S, P)$, care generează L; $L(G) = L$ și dacă $\lambda \in L(G)$, atunci în G există producția $S \to \lambda$ (unica λ-producție) astfel încât în membrul drept al niciunei producții din P nu apare S.

Toate producțiile lui G vor fi de forma $A \to BC$ sau $A \to a$, $B, C, A \in N$, $a \in \Sigma$, cu excepția (eventual) a producției $S \to \lambda$ dacă $\lambda \in L(G)$.

Atunci, în orice arbore de derivare din G numărul de simboluri de la un anumit nivel este de cel mult de 2 ori numărul de simboluri de la nivelul precedent.

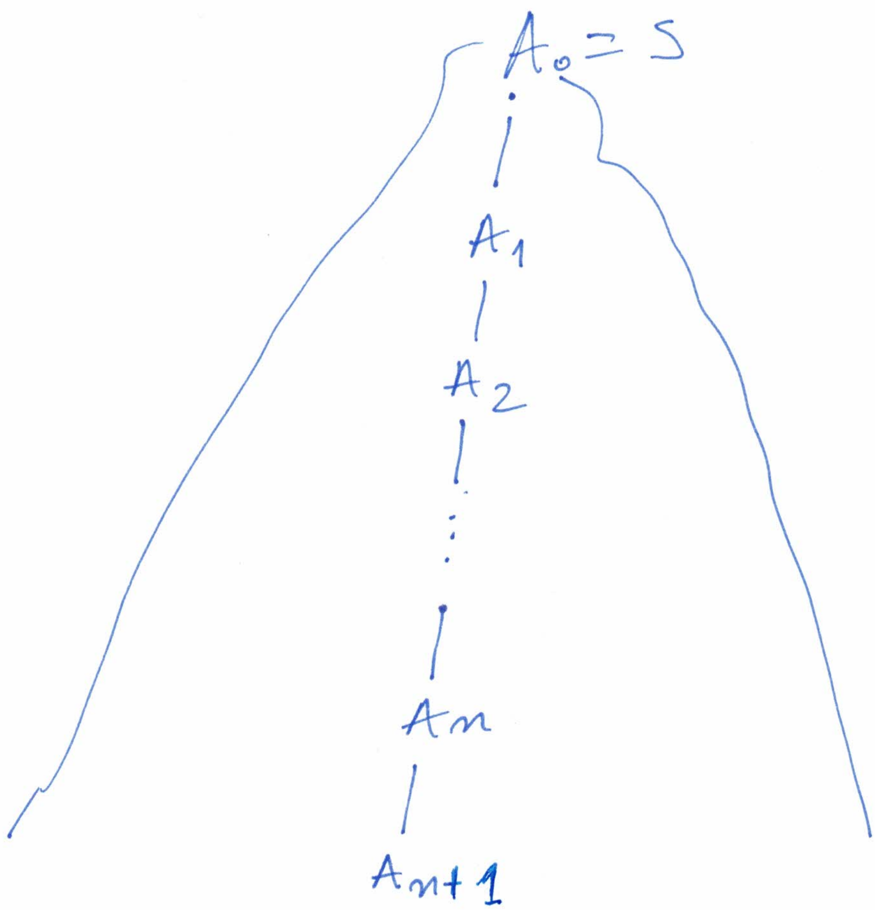
Astfel, pe nivelul 0 avem rădăcina arborelui etichetată cu S, pe nivelul 1 există cel mult 2 simboluri, pe nivelul 2 avem cel mult 4 simboluri, ..., iar pe nivelul i avem cel mult 2^i simboluri (terminale sau neterminale).

Totodată, pentru a avea cel puțin 2^n simboluri pe un nivel, arborele trebuie să aibă adâncimea de cel puțin n, deci are cel puțin $n+1$ nivele.

Vom considera $k = 2^{n+1}$, unde n este numărul neterminalelor lui G.

Fie $z \in L$, $|z| \geq k$. Rezultă că orice arbore de derivare pentru z are cel puțin adâncimea $n+1$. Considerăm cel mai lung drum din arbore. Acest drum are cel puțin lungimea $n+1$, de aceea conține cel puțin $n+2$ noduri, etichetate cu $A_0, A_1, \ldots, A_n, A_{n+1}$, unde $A_0 = S$, $A_0, A_1, \ldots, A_n \in N$. Ultimul nod al acestui drum va fi etichetat cu un simbol terminal, deoarece aparține frontierei arborelui de derivare al lui $z \in \Sigma^*$.

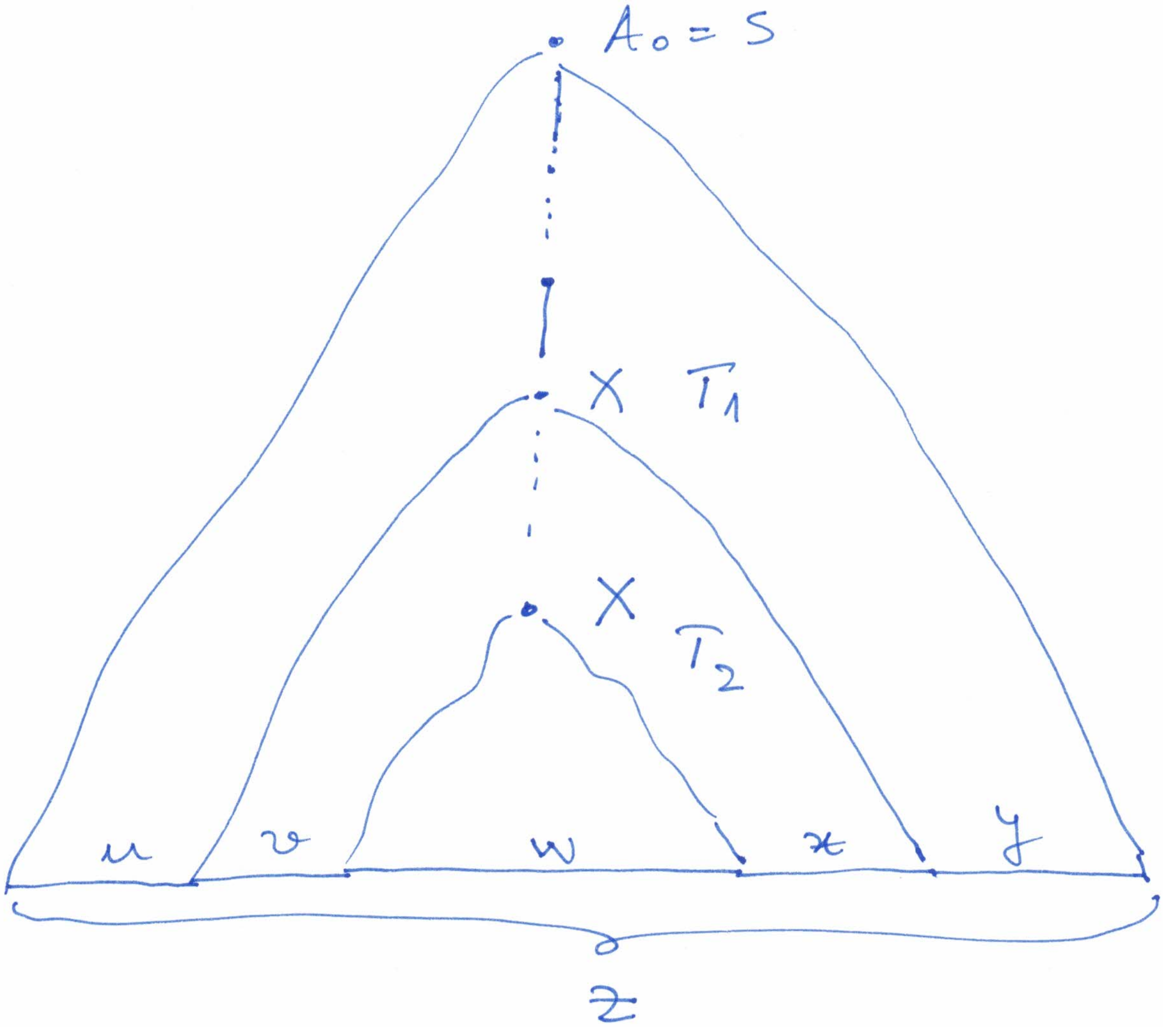
$= 4z$



Deoarece G are n neterminali, rezultă că există cel puțin un neterminal care apare de cel puțin 2 ori în acest drum.

Considerăm prima pereche de noduri de pe acest drum etichetate cu același neterminal, să îl notăm cu $X \in N$, de jos în sus.

Fie T_1 subarborele de rădăcină etichetată cu prima apariție a lui X și T_2 subarborele de rădăcină etichetată cu a doua apariție a lui X, ca în figura:



Vom considera u, v, w, x, y ca mai sus, $z = uvwxy$.
Aceasta înseamnă că în G avem derivările:

$$S \overset{*}{\Rightarrow} uXy \overset{+}{\Rightarrow} uvXxy \overset{+}{\Rightarrow} uvwxy,$$

unde

$$X \overset{+}{\Rightarrow} vXx \text{ şi } X \overset{+}{\Rightarrow} w$$

Observăm că $vx \neq \lambda$. Într-adevăr, dacă $v = x = \lambda$ rezultă că în G avem derivarea $X \overset{+}{\Rightarrow} X$, ceea ce nu este posibil, deoarece G este în formă normală Chomsky. Deci $vx \neq \lambda$.
Totodată, $|vwx| \leq k = 2^{n+1}$, deoarece X

etichetează prima pereche de noduri, de jos în sus, pe drumul ales (care este cel mai lung, de lungime cel puțin k), etichetate cu același neterminal (X).

De asemenea, în G avem derivările:

$$S \stackrel{*}{\Rightarrow} u\cancel{X}y \stackrel{*}{\Rightarrow} uwy \quad (t=0 \text{ din condiția iv})$$

$$S \stackrel{*}{\Rightarrow} uXy \stackrel{*}{\Rightarrow} uvXxy \stackrel{*}{\Rightarrow} \dots \stackrel{*}{\Rightarrow} uv^tXx^ty$$

$$\stackrel{*}{\Rightarrow} uv^twx^ty \ , \ t \geq 1$$

Cu aceasta, demonstrația este încheiată.

<u>Observații</u>

1) Teorema anterioară mai este cunoscută sub numele de: lema de pompare pentru limbajele independente de context, lema Bar-Hillel (după numele celor care au demonstrat-o), teorema u-v-w-x-y.

2) Teorema de mai sus ne furnizează o condiție <u>necesară</u>, dar nu și suficientă pentru ca un limbaj să fie independent de context.

3) Există o formă mai „tare" a acestei teoreme, cunoscută ca Lema lui Ogden.

4) Pentru a arăta că un limbaj L nu este independent de context este suficient să arătăm că pentru orice $k \geq 1$ există $z \in L$, $|z| \geq k$, astfel încât $\forall u, v, w, x, y$ pentru care $z = uvwxy$, $vx \neq \lambda$, $|vwx| \leq k$, există $t \geq 0$ astfel încât $uv^t w x^t y \notin L$.

EXEMPLE

1) $L_1 = \{a^n b^n a^n \mid n \geq 0\} \notin \mathcal{L}_{CF}$

Demonstrație Presupunem că $L_1 \in \mathcal{L}_{CF}$

Fie $k \geq 1$ și $z = a^k b^k a^k$, $|z| = 3k \geq k$.

Fie u, v, w, x, y astfel încât $z = uvwxy$, $vx \neq \lambda$, $|vwx| \leq k$.

Să considerăm șirul $z_2 = uv^2wx^2y$.

Avem cazurile:

- Dacă v sau x conțin cel puțin un 'a' și cel puțin un 'b', atunci $z_2 = uv^2wx^2y$ nu are forma $a^* b^* a^*$, deci $z_2 \notin L$
- Dacă v și x conțin doar 'a'-uri, atunci $|z_2|_b = |z|_b$, $|z|_a = 2|z|_b$, $|z_2|_a = |z|_a + |vx|_a = 2|z_2|_b + |vx|_a \geq 2|z_2|_b + 1$, deci $z_2 \notin L$.

- Dacă v și x conțin doar 'b'-uri, avem:

$|z|_a = 2|z|_b$, $|z_2|_a = |z|_a$, $|z_2|_b = |z|_b + |vx|_b$.

Atunci:

$$\left.\begin{array}{l} |z_2|_a = |z|_a = 2|z|_b \\ 2|z_2|_b = 2|z|_b + 2|vx|_b \end{array}\right\} \Rightarrow |z_2|_a \neq 2|z_2|_b \Rightarrow$$

$$\Rightarrow z_2 \notin L$$

- Dacă v sau x conține doar 'a'-uri, iar celălalt conține doar 'b' uri, atunci uv^2wx^2y nu poate fi de forma $a^n b^m a^m$, deci $z_2 \notin L$.

În concluzie, $L_1 \notin \mathcal{L}_{CF}$

2) Să se arate că $L_2 = \{a^{n^2} \mid n \geq 1\}$ nu este independent de context.

<u>Dem</u> Presupunem că $L_2 \in \mathcal{L}_{CF}$.

Fie $k \geq 1$ și $z = a^{k^2}$, $|z| = k^2 \geq k$.

Fie u, v, w, x, y astfel încât $z = uvwxy$, $|vx| \geq 1$, $|vwx| \leq k$. Fie $|vx| = m$, $1 \leq m \leq k$.

Avem:

$|uv^2wx^2y| = |uvwxy| + |vx| = k^2 + m$

Dar $k^2 < k^2 + m \leq k^2 + k < (k+1)^2$, deci $uv^2wx^2y \notin L_2$;

rezultă că $L_2 \notin \mathcal{L}_{CF}$

<u>Exerciții</u> Să se arate că următoarele limbaje nu sunt independente de context:

$L_3 = \{a^n b^n c^n \mid n \geq 1\}$

$L_4 = \{a^{2^n} \mid n \geq 1\}$

$L_5 = \{a^p \mid p \geq 2,\ p \text{ prim}\}$

$L_6 = \{ww \mid w \in \{a, b\}^*\}$

$L_7 = \{a^n b^n a^m \mid 1 \leq m \leq n\}$

$L_8 = \{w \in \{a, b, c\}^* \mid |w|_a = |w|_b = |w|_c\}$

<u>Obs.</u> În exercițiul de mai sus puteți folosi și rezultatul următor, pe care îl vom demonstra ulterior:

$\mathcal{L}_{CF}$ este închisă la intersecția cu limbaje regulate.

PROPRIETĂȚI DE ÎNCHIDERE PENTRU $\mathcal{L}_{CF}$

Teorema 2 $\mathcal{L}_{CF}$ este închisă la $\cup, \cdot, *$, dar nu este închisă la $\cap$ și $^-$ (complementare)

Dem. Fie $L_1, L_2 \in \mathcal{L}_{CF}$ și $G_1 = (N_1, \Sigma_1, S_1, P_1)$, $G_2 = (N_2, \Sigma_2, S_2, P_2)$ gramatici independente de context astfel încât $L(G_1) = L_1$, $L(G_2) = L_2$, $N_1 \cap N_2 = \phi$.

1. Închiderea la reuniune.

Construim $G = (N_1 \cup N_2 \cup \{S\}, \Sigma_1 \cup \Sigma_2, S, P_1 \cup P_2 \cup \cup \{S \to S_1 | S_2\})$.

Evident, $L(G) = L(G_1) \cup L(G_2) = L_1 \cup L_2$

2. Închiderea la concatenare.

Fie gramatica:

$G = (N_1 \cup N_2 \cup \{S\}, \Sigma_1 \cup \Sigma_2, S, P_1 \cup P_2 \cup \{S \to S_1 S_2\})$

Avem $L(G) = L(G_1) \cdot L(G_2) = L_1 \cdot L_2$.

3) Închiderea la iterația Kleene:

Fie $G = (N_1 \cup \{S_1\}, \Sigma_1, S, P \cup \{S \to S_1 S | \lambda\})$

Avem $L(G) = (L(G_1))^* = L_1^*$.

4. $\mathcal{L}_{CF}$ nu este închisă la intersecție

Fie $L_1 = \{a^i b^i a^j \mid i, j \geq 0\}$,

$L_2 = \{a^i b^j a^j \mid i, j \geq 0\}$

care pot fi generate cu gramaticile

$G_1 = (\{S_1, A, B\}, \{a, b\}, S_1, \{S_1 \to AB, A \to aAb \mid \lambda, B \to aB \mid \lambda\})$

$G_2 = (\{S_2, C, D\}, \{a, b\}, S_2, \{S_2 \to CD, C \to aC \mid \lambda, D \to bDa \mid \lambda\})$,

deci $L_1, L_2 \in \mathcal{L}_{CF}$.

Însă, $L_1 \cap L_2 = \{a^k b^k a^k \mid k \geq 0\}$, care am văzut că nu este independent de context

5. $\mathcal{L}_{CF}$ nu este închisă la complementară

Notăm cu $\overline{L}$ complementara lui L față de Σ^*, unde $L \subseteq \Sigma^*$.

Fie $L_1, L_2 \in \mathcal{L}_{CF}$. Presupunem că $\mathcal{L}_{CF}$ închisă la complementară.

Atunci:

$L = L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$, deci $L \in \mathcal{L}_{CF}$,

adică $\mathcal{L}_{CF}$ închisă la intersecție, contradicție