

Algoritmul Cocke-Younger-Kasami (1965)

Acest algoritm abreviat **CYK** este utilizat pentru a stabili dacă un cuvânt aparține sau nu limbajului generat de o gramatică independentă de context în forma normală Chomsky.

Algoritmul, folosind metoda programării dinamice, descompune cuvântul supus analizei în subșiruri din ce în ce mai lungi și determină simbolurile neterminale din care pot fi obținute, complexitatea algoritmului fiind de tipul $O(n^3)$.

Fie o gramatică independentă de context $G = (N, T, S, P)$ în forma normală Chomsky și fie un cuvânt $w = a_1 a_2 \dots a_n \in T^*$.

Subșirul din cuvântul w care începe din poziția $i \in \mathbb{N}^*$, $i \leq n$ și are lungimea $j \in \mathbb{N}^*$, $j \leq n+1-i$, se notează prin $w_{i,j} = a_i a_{i+1} \dots a_{i+j-1}$ și este numit **subcuvânt** al lui w . Se observă că nu au sens subcuvintele a căror lungime depășesc valoarea $n + 1 - i$.

Pentru fiecare pereche i și j pentru care există subcuvântul $w_{i,j}$ se definește mulțimea $A_{i,j}$ formată din neterminalele gramaticii din care poate fi derivat $w_{i,j}$, adică

$$A_{i,j} = \{X \in N \mid X \xRightarrow{*} w_{i,j}\}.$$

În funcție de producțiile gramaticii unele dintre aceste mulțimi pot fi vide.

Aplicația 1. Pentru $n = 5$ să se indice mulțimile $A_{i,j}$ care au sens.

Răspuns.

Pentru $j = 1$ au sens mulțimile:

$$A_{1,1}, A_{2,1}, A_{3,1}, A_{4,1}, A_{5,1};$$

Pentru $j = 2$ au sens mulțimile:

$$A_{1,2}, A_{2,2}, A_{3,2}, A_{4,2};$$

Pentru $j = 3$ au sens mulțimile:

$$A_{1,3}, A_{2,3}, A_{3,3};$$

Pentru $j = 4$ au sens mulțimile:

$$A_{1,4}, A_{2,4};$$

Pentru $j = 5$ are sens mulțimea:

$$A_{1,5};$$

Unele dintre mulțimile care au sens pot fi vide.

Algoritmul **CYK** are următorii pași:

Pasul 1. Se determină pentru fiecare $i = \overline{1, n}$:

$$A_{i,1} = \{X \in N \mid X \rightarrow a_i \text{ este din } P\};$$

Pasul 2. Pentru fiecare pereche i și j , $j \geq 2$, pentru care există subcuvântul $w_{i,j}$ se calculează:

$$A_{i,j} = \bigcup_{1 \leq k \leq j-1} \{X \in N \mid \exists Y \in A_{i,k}, \exists Z \in A_{i+k,j-k} \text{ a.î. } X \rightarrow YZ \text{ este din } P\};$$

Pasul 3. Dacă $S \in A_{1,n}$, atunci $w \in L(G)$ și STOP.

Dacă $S \notin A_{1,n}$, atunci $w \notin L(G)$ și STOP.

Observația 1. Calculul mulțimilor $A_{i,j}$ folosește metoda programării dinamice, cu scopul de a simplifica acest calcul. Totuși, aceste mulțimi se pot calcula și direct, identificând toate derivările posibile de forma $X \Rightarrow^* w_{i,j}$.

Observația 2. Au loc implicațiile:

$$S \Rightarrow^* w \Leftrightarrow S \in A_{1,n}.$$

Exemplul 1. Se consideră gramatica independentă de context $G = (N, T, S, P)$, unde $N = \{S, D\}$, $T = \{a, b\}$ și mulțimea P formată din următoarele producții:

$$S \rightarrow aSb \mid aDb \quad (1)$$

$$D \rightarrow aD \mid a \quad (2)$$

Pentru a aplica algoritmul **CYK** se determină, mai întâi, gramatica independentă de context $G^{(n)} = (N^{(n)}, T, S, P^{(n)})$ în forma normală Chomsky, echivalentă cu gramatica G , unde $N^{(n)} = \{S, A, B, C, D\}$, $T = \{a, b\}$ și următoarea mulțime a producțiilor P :

$$S \rightarrow AB \mid AC \quad (1)$$

$$A \rightarrow a \quad (2)$$

$$B \rightarrow SB \mid b \quad (3)$$

$$C \rightarrow DB \quad (4)$$

$$D \rightarrow AD \mid a \quad (5)$$

Fie cuvântul $w_1 = aaaabb$, folosind algoritmul **CYK** să se stabilească dacă $w_1 \in L(G)$.

Răspunsul se află aplicând algoritmul gramaticii $G^{(n)}$ echivalentă cu gramatica G .

Pasul 1.

Deoarece $n = |w_1| = 6$, se determină mulțimile care au sens

$$A_{i,1} = \{X \in N \mid X \rightarrow a_i \text{ este din } P\}, \text{ pentru } i = \overline{1,6}:$$

- cum $A_{1,1}$ se referă la subcuvântul $w_{1,1} = a$ se obține

$$A_{1,1} = \{A, D\}, \text{ deoarece } P \text{ conține producțiile } A \rightarrow a \text{ și } D \rightarrow a;$$

- cum $A_{2,1}$ se referă la subcuvântul $w_{2,1} = a$ se obține

$$A_{2,1} = \{A, D\}, \text{ deoarece } P \text{ conține producțiile } A \rightarrow a \text{ și } D \rightarrow a;$$

- cum $A_{3,1}$ se referă la subcuvântul $w_{3,1} = a$ se obține

$$A_{3,1} = \{A, D\}, \text{ deoarece } P \text{ conține producțiile } A \rightarrow a \text{ și } D \rightarrow a;$$

- cum $A_{4,1}$ se referă la subcuvântul $w_{4,1} = a$ se obține

$$A_{4,1} = \{A, D\}, \text{ deoarece } P \text{ conține producțiile } A \rightarrow a \text{ și } D \rightarrow a;$$

- cum $A_{5,1}$ se referă la subcuvântul $w_{5,1} = b$ se obține

$$A_{5,1} = \{B\}, \text{ deoarece } P \text{ conține producția } B \rightarrow b;$$

- cum $A_{6,1}$ se referă la subcuvântul $w_{6,1} = b$ se obține

$$A_{6,1} = \{B\}, \text{ deoarece } P \text{ conține producția } B \rightarrow b.$$

Pasul 2.

Conform algoritmului pasul 2 se obține astfel:

- **pentru $j = 2$** se determină mulțimile $A_{i,2}$, $i = \overline{1,5}$, folosind mulțimile determinate la pasul 1 sau de mai sus, $A_{i,1}$ și $A_{i+1,1}$:

- cum $A_{1,2}$ se referă la subcuvântul $w_{1,2} = aa$, se obține $A_{1,2} = \{D\}$, deoarece $A_{1,1} = \{A, D\} = A_{2,2-1}$ și se caută neterminalul X care satisface producția $X \rightarrow AD$ din P , acesta fiind $X = D$. Celelalte variante $X \rightarrow AA$, $X \rightarrow DD$ și $X \rightarrow DA$ nu există în P .
- cum $A_{2,2}$ se referă la subcuvântul $w_{2,2} = aa$, se obține $A_{2,2} = \{D\}$, deoarece $A_{2,1} = \{A, D\} = A_{3,2-1}$ se caută neterminalul X care satisface producția $X \rightarrow AD$ din P , acesta fiind $X = D$. Celelalte variante $X \rightarrow AA$, $X \rightarrow DD$ și $X \rightarrow DA$ nu există în P .
- cum $A_{3,2}$ se referă la subcuvântul $w_{3,2} = aa$, se obține $A_{3,2} = \{D\}$, deoarece $A_{3,1} = \{A, D\} = A_{4,2-1}$ se caută neterminalul X care satisface producția $X \rightarrow AD$ din P , acesta fiind $X = D$. Celelalte variante $X \rightarrow AA$, $X \rightarrow DD$ și $X \rightarrow DA$ nu există în P .
- cum $A_{4,2}$ se referă la subcuvântul $w_{4,2} = ab$, se obține $A_{4,2} = \{S, C\}$, deoarece $A_{4,1} = \{A, D\}$, $A_{5,2-1} = \{B\}$ se caută neterminalul X care satisface producția $X \rightarrow AB$ sau producția $X \rightarrow DB$ din P , acesta fiind $X = S$, respectiv $X = C$.
- cum $A_{5,2}$ se referă la subcuvântul $w_{5,2} = bb$, se obține $A_{5,2} = \emptyset$, deoarece $A_{5,1} = \{B\}$ și $A_{6,2-1} = \{B\}$ se caută neterminalul X care satisface producția $X \rightarrow BB$, care nu există în P .

- **pentru $j = 3$** se determină mulțimile $A_{i,3}$, $i = \overline{1,4}$, folosind mulțimile determinate la pasul 1 sau de mai sus, $A_{i,1}$ și $A_{i+1,2}$ sau $A_{i,2}$ și $A_{i+2,1}$:

- cum $A_{1,3}$ se referă la subcuvântul $w_{1,3} = aaa$, se obține

$$A_{1,3} = \{D\} \cup \emptyset = \{D\}, \text{ deoarece:}$$

- *cazul 1:* $A_{1,1} = \{A, D\}$ și $A_{2,3-1} = \{D\}$ se caută neterminalul X care satisface producția $X \rightarrow AD$ din P , acesta fiind $X = D$. Cealaltă variantă $X \rightarrow DD$ nu există în P ;

- *cazul 2:* $A_{1,2} = \{D\}$, $A_{3,3-2} = \{A, D\}$ nu există neterminalul X care să satisfacă vreo producție $X \rightarrow DA$ sau $X \rightarrow DD$ în P .

- cum $A_{2,3}$ se referă la subcuvântul $w_{2,3} = aaa$, se obține

$$A_{2,3} = \{D\} \cup \emptyset = \{D\}, \text{ deoarece:}$$

- *cazul 1:* $A_{2,1} = \{A, D\}$ și $A_{3,3-1} = \{D\}$ se caută neterminalul X care satisface producția $X \rightarrow AD$ din P , acesta fiind $X = D$. Cealaltă variantă $X \rightarrow DD$ nu există în P ;

- *cazul 2:* $A_{2,2} = \{D\}$ și $A_{4,3-2} = \{A, D\}$ nu există neterminalul X care să satisfacă vreo producție $X \rightarrow DA$ sau $X \rightarrow DD$ în P .

- cum $A_{3,3}$ se referă la subcuvântul $w_{3,3} = aab$, se obține

$A_{3,3} = \{S\} \cup \{C\} = \{S, C\}$, deoarece:

- *cazul 1*: $A_{3,1} = \{A, D\}$ și $A_{4,3-1} = \{S, C\}$ se caută neterminalul X care satisface producția $X \rightarrow AC$ din P , acesta fiind $X = S$. Celelalte variante $X \rightarrow AS \mid DS \mid DC$ nu există în P ;

- *cazul 2*: $A_{3,2} = \{D\}$ și $A_{5,3-2} = \{B\}$ se caută neterminalul X care satisface producția $X \rightarrow DB$ din P , acesta fiind $X = C$.

- cum $A_{4,3}$ se referă la subcuvântul $w_{4,3} = \mathbf{abb}$, se obține

$A_{4,3} = \emptyset \cup \{B\} = \{B\}$, deoarece:

- *cazul 1*: $A_{4,1} = \{A, D\}$ și $A_{5,3-1} = \emptyset$, nu există neterminalul X care să satisfacă vreo producție din P ;

- *cazul 2*: $A_{4,2} = \{S, C\}$ și $A_{5,3-2} = \{B\}$ se caută neterminalul X care satisface producția $X \rightarrow SB$ din P , acesta fiind $X = B$. Cealaltă variantă $X \rightarrow CB$ nu există în P .

- **pentru $j = 4$** se determină mulțimile $A_{i,4}$, $i = \overline{1,3}$, folosind mulțimile determinate la pasul 1 sau de mai sus, respectiv $A_{i,1}$ și $A_{i+1,3}$ sau $A_{i,2}$ și $A_{i+2,2}$ sau $A_{i,3}$ și $A_{i+3,1}$:

- cum $A_{1,4}$ se referă la subcuvântul $w_{1,4} = \mathbf{aaaa}$, se obține

$A_{1,4} = \{D\} \cup \emptyset \cup \emptyset = \{D\}$, deoarece:

- *cazul 1*: $A_{1,1} = \{A, D\}$ și $A_{2,4-1} = \{D\}$ se caută neterminalul X care satisface producția $X \rightarrow AD$ din P , acesta fiind $X = D$. Cealaltă variantă $X \rightarrow DD$ nu există în P ;

- *cazul 2*: $A_{1,2} = \{D\}$ și $A_{3,4-2} = \{D\}$, nu există neterminalul X care să satisfacă vreo producție $X \rightarrow DD$ în P .

- *cazul 3*: $A_{1,3} = \{D\}$ și $A_{4,4-3} = \{A, D\}$ nu există neterminalul X care să satisfacă vreo producție $X \rightarrow DA$ sau $X \rightarrow DD$ în P .

- cum $A_{2,4}$ se referă la subcuvântul $w_{2,4} = \mathbf{aaab}$, se obține

$A_{2,4} = \{S\} \cup \emptyset \cup \{C\} = \{C, S\}$, deoarece:

- *cazul 1*: $A_{2,1} = \{A, D\}$ și $A_{3,4-1} = \{S, C\}$ se caută neterminalul X care satisface producția $X \rightarrow AC$ din P , acesta fiind $X = S$. Celelalte variante $X \rightarrow AS \mid DS \mid DC$ nu pot exista în P .

- *cazul 2*: $A_{2,2} = \{D\}$ și $A_{4,4-2} = \{S, C\}$, nu există neterminalul X care să satisfacă vreo producție $X \rightarrow DS$ sau $X \rightarrow DC$ din P ;

- *cazul 3*: $A_{2,3} = \{D\}$ și $A_{5,4-3} = \{B\}$ se caută neterminalul X care satisface producția $X \rightarrow DB$ din P , acesta fiind $X = C$.

- cum $A_{3,4}$ se referă la subcuvântul $w_{3,4} = \mathbf{aabb}$, se obține

$A_{3,4} = \{C, S\} \cup \emptyset \cup \{B\} = \{B, C, S\}$, deoarece:

- *cazul 1*: $A_{3,1} = \{A, D\}$ și $A_{4,4-1} = \{B\}$ se caută neterminalul X care satisface producția $X \rightarrow AB$ sau producția $X \rightarrow DB$ din P , acesta fiind $X \in \{S, C\}$.

- *cazul 2*: $A_{3,2} = \{D\}$ și $A_{5,4-2} = \emptyset$, nu există neterminalul X care să satisfacă vreo producție din P ;

- *cazul 3*: $A_{3,3} = \{S, C\}$ și $A_{6,4-3} = \{B\}$ se caută neterminalul X care satisface producția $X \rightarrow SB$ din P , acesta fiind $X = B$. Cealaltă variantă $X \rightarrow CB$ nu există în P .

- **pentru $j = 5$** se determină mulțimile $A_{i,5}$, $i = \overline{1,2}$, folosind mulțimile determinate la pasul 1 sau de mai sus, respectiv $A_{i,1}$ și $A_{i+1,4}$ sau $A_{i,2}$ și $A_{i+2,3}$ sau $A_{i,3}$ și $A_{i+3,2}$ sau $A_{i,4}$ și $A_{i+4,1}$:

- cum $A_{1,5}$ se referă la subcuvântul $w_{1,5} = \mathbf{aaaab}$, se obține

$$A_{1,5} = \{S\} \cup \emptyset \cup \emptyset \cup \{C\} = \{C, S\}, \text{ deoarece:}$$

- *cazul 1:* $A_{1,1} = \{A, D\}$ și $A_{2,5-1} = \{C, S\}$ se caută neterminalul X care satisface producția $X \rightarrow AC$ din P , acesta fiind $X \in \{S\}$. Celelalte variante $X \rightarrow AS \mid DC \mid DS$ nu pot exista în P .
- *cazul 2:* $A_{1,2} = \{D\}$ și $A_{3,5-2} = \{C, S\}$, nu există neterminalul X care să satisfacă vreo producție $X \rightarrow DS$ sau $X \rightarrow DC$ din P ;
- *cazul 3:* $A_{1,3} = \{D\}$ și $A_{4,5-3} = \{S, C\}$, nu există neterminalul X care să satisfacă vreo producție $X \rightarrow DS$ sau $X \rightarrow DC$ din P ;
- *cazul 4:* $A_{1,4} = \{D\}$ și $A_{5,5-4} = \{B\}$ se caută neterminalul X care satisface producția $X \rightarrow DB$ din P , acesta fiind $X = C$.

- cum $A_{2,5}$ se referă la subcuvântul $w_{2,5} = \mathbf{aaabb}$, se obține

$$A_{2,5} = \{S, C\} \cup \{C\} \cup \emptyset \cup \{B\} = \{B, C, S\}, \text{ deoarece:}$$

- *cazul 1:* $A_{2,1} = \{A, D\}$ și $A_{3,5-1} = \{B, C, S\}$ se caută neterminalul X care satisface producția $X \rightarrow AC$ sau producția $X \rightarrow AB$ sau producția $X \rightarrow DB$ din P , acesta fiind $X \in \{S, C\}$. Celelalte variante $X \rightarrow AS \mid DC \mid DS$ nu pot exista în P .
- *cazul 2:* $A_{2,2} = \{D\}$ și $A_{4,5-2} = \{B\}$ se caută neterminalul X care satisface producția $X \rightarrow DB$ din P , acesta fiind $X = C$.
- *cazul 3:* $A_{2,3} = \{D\}$ și $A_{5,5-3} = \emptyset$, nu există neterminalul X care să satisfacă vreo producție din P ;
- *cazul 4:* $A_{2,4} = \{C, S\}$ și $A_{6,5-4} = \{B\}$ se caută neterminalul X care satisface producția $X \rightarrow SB$ din P , acesta fiind $X = B$. Cealaltă variantă $X \rightarrow CB$ nu există în P .

- **pentru $j = 6$** se determină mulțimea $A_{1,6}$, folosind mulțimile determinate la pasul 1 sau de mai sus, respectiv $A_{1,1}$ și $A_{2,5}$ sau $A_{1,2}$ și $A_{3,4}$ sau $A_{1,3}$ și $A_{4,3}$ sau $A_{1,4}$ și $A_{5,2}$ sau $A_{1,5}$ și $A_{6,1}$:

- cum $A_{1,6}$ se referă la subcuvântul $w_{1,6} = \mathbf{aaaabb}$, se obține

$$A_{1,6} = \{S, C\} \cup \{C\} \cup \{C\} \cup \emptyset \cup \{B\} = \{B, C, S\}, \text{ deoarece:}$$

- *cazul 1:* $A_{1,1} = \{A, D\}$ și $A_{2,6-1} = \{B, C, S\}$ se caută neterminalul X care satisface producția $X \rightarrow AC$ sau producția $X \rightarrow AB$ sau producția $X \rightarrow DB$ din P , acesta fiind $X \in \{S, C\}$. Celelalte variante $X \rightarrow AS \mid DC \mid DS$ nu există în P .
- *cazul 2:* $A_{1,2} = \{D\}$ și $A_{3,6-2} = \{B, C, S\}$ se caută neterminalul X care satisface producția $X \rightarrow DB$ din P , acesta fiind $X = C$. Celelalte variante $X \rightarrow DC \mid DS$ nu există în P ;
- *cazul 3:* $A_{1,3} = \{D\}$ și $A_{4,6-3} = \{B\}$ se caută neterminalul X care satisface producția $X \rightarrow DB$ din P , acesta fiind $X = C$.
- *cazul 4:* $A_{1,4} = \{D\}$ și $A_{5,6-4} = \emptyset$, nu există neterminalul X care să satisfacă vreo producție din P ;
- *cazul 5:* $A_{1,5} = \{C, S\}$ și $A_{6,6-5} = \{B\}$ se caută neterminalul X care satisface producția $X \rightarrow SB$ din P , acesta fiind $X = B$. Cealaltă variantă $X \rightarrow CB$ nu există în P .

Rezultatele obținute se pot sistematiza sub forma următorului tabel cu 6 linii și 6 coloane:

$i \backslash j$	1	2	3	4	5	6
1	$\{D, A\}$	$\{D, A\}$	$\{D, A\}$	$\{D, A\}$	$\{B\}$	$\{B\}$
2	$\{D\}$	$\{D\}$	$\{D\}$	$\{S, C\}$	\emptyset	
3	$\{D\}$	$\{D\}$	$\{S, C\}$	$\{B\}$		
4	$\{D\}$	$\{S, C\}$	$\{S, B, C\}$			
5	$\{S, C\}$	$\{S, B, C\}$				
6	$\{S, B, C\}$					

Pasul 3. Deoarece $S \in A_{1,6}$, rezultă $w_1 \in L(G)$.

Fie cuvântul $w_2 = aabaabbba$, folosind algoritmul **CYK** să se stabilească dacă $w_2 \in L(G)$. Răspunsul se află aplicând algoritmul gramaticii $G^{(n)}$ echivalentă cu gramatica G . Deoarece $n = |w_2| = 9$, în urma aplicării algoritmului **CYK** s-au obținut rezultatele din următorul tabel cu 9 linii și 9 coloane:

$i \backslash j$	1	2	3	4	5	6	7	8	9
1	$\{D, A\}$	$\{D, A\}$	$\{B\}$	$\{D, A\}$	$\{D, A\}$	$\{B\}$	$\{B\}$	$\{B\}$	$\{D, A\}$
2	$\{D\}$	$\{S, C\}$	\emptyset	$\{D\}$	$\{S, C\}$	\emptyset	\emptyset	\emptyset	
3	$\{S, C\}$	\emptyset	\emptyset	$\{S, C\}$	$\{B\}$	\emptyset	\emptyset		
4	\emptyset	\emptyset	\emptyset	$\{S, B, C\}$	\emptyset	\emptyset			
5	\emptyset	\emptyset	\emptyset	$\{B\}$	\emptyset				
6	\emptyset	$\{B\}$	\emptyset	\emptyset					
7	$\{S, B, C\}$	$\{B\}$	\emptyset						
8	$\{S, B, C\}$	\emptyset							
9	\emptyset								

Deoarece $S \notin A_{1,9} = \emptyset$ cuvântul $w_2 \notin L(G)$.