Algoritmul CYK *(Cocke–Younger–Kasami)*
(verificare generare cuvânt de către gramatică în F.N.Chomsky)

Am văzut că dacă gramatica avea producţii specifice uneia independente de context, atunci foloseam arbori de derivare şi porneam de la simbolul de start S şi încercam să ajungem la cuvântul dat.

Dar dacă gramatica are producţiile scrise în *forma normală Chomsky*, atunci vom folosi acest algoritm CYK, şi vom porni de la cuvântul dat încercând să ajungem la simbolul de start S.

$V_{i\,j} = \{A \in N \mid \exists B \in V_{i\,k}, \exists C \in V_{i+k,\, j-k}, (A \to BC) \in P, k \in \{1,2,...,j-1\}\}$
Notăm cu $\mathbf{V_{i\,j}}$ mulţimea de neterminale care pot genera acea porţiune din cuvântul de intrare care începe de pe poziţia i şi are lungimea j.

Ideea este de a împărţi cuvântul în două părţi în toate modurile posibile. Dacă prima parte poate fi generată de neterminalul B, a doua parte poate fi generată de neterminalul C şi avem în gramatică producţia "A→BC", atunci înseamnă că întreg cuvântul poate fi generat de neterminalul A.
Observăm că avem împărţiri posibile pentru k=1 (prima literă în prima parte, iar restul în a doua), k=2 (primele două litere în prima parte, iar restul în a doua), ... până la k=j–1 (doar ultima literă în a doua parte, iar restul în prima).

Obs: *Nu reţineţi indicii pe dinafară, ci încercaţi să îi deduceţi logic!*
Valoare pe care o calculăm este $\mathbf{V_{i\,j}}$ pentru cuvântul care începe la poziţia i şi are lungimea j.
Prima parte (generată de neterminalul B) conţine primele k litere din cuvânt. Deci va începe la poziţia i (la fel ca întreg cuvântul) şi va avea lungime k, deci $\mathbf{V_{i\,k}}$.
A doua parte (generată de neterminalul C) conţine restul de litere, adică j–k, şi începe cu k poziţii mai în dreapta faţă de cuvântul total, adică la i+k, deci avem $\mathbf{V_{i+k,\, j-k}}$.

Răspunsul final pe care îl căutăm este $\mathbf{V_{1,\,|w|}}$ adică mulţimea de neterminale din care putem genera cuvântul dat începând de pe prima poziţie şi având lungimea egală cu întreg cuvântul. Dacă în această mulţime se va găsi şi simbolul de start S, înseamnă că cuvântul este generat de gramatică. Dacă nu, atunci nu este generat.

La pasul iniţial vom căuta neterminalele care generează fiecare literă mică (terminal) din cuvânt: $V_{i1} = \{A \in N \mid (A \to x_i) \in P\}$, unde x_i este litera de pe poziţia i din cuvântul dat.

EXEMPLU:

Avem o gramatică în F.N.Chomsky:
$$\begin{aligned} &S \to BA \\ &A \to a \mid BA \mid CA \\ &B \to b \mid AA \mid AB \\ &C \to c \mid BC \mid BB \end{aligned}$$
şi un cuvânt: baabca.

Avem un tabel cu i pe coloane şi j pe linii în care vom completa neterminalele din mulţimile $V_{i\,j}$. Variabilele i şi j iau valori între 1 şi 6 (lungimea cuvântului dat).

Observăm că putem avea **maxim i + j – 1 = |w|** = 6. Deci de exemplu V_{62} nu va fi calculat (6 + 2 – 1 = 7 > 6) pentru că dacă începem pe poziţia 6 (ultima) nu vom putea avea lungimea 2 în cuvânt.

Pentru prima linie, j = 1, vom căuta în gramatică pentru fiecare literă mică de ce neterminale poate fi generată.
Vom avea $V_{21} = V_{31} = V_{61} = \{A\}$ pentru că litera ”a” se află în cuvânt pe poziţiile 2, 3, 6, iar neterminalul A este singurul care generează ”a”.
Vom avea $V_{11} = V_{41} = \{B\}$ pentru că litera ”b” se află în cuvânt pe poziţiile 1, 4, iar neterminalul B este singurul care generează ”b”.
Vom avea $V_{51} = \{C\}$ pentru că litera ”c” se află în cuvânt pe poziţia 5, iar neterminalul C este singurul care generează ”c”.

V_{ij}	i = 1	i = 2	i = 3	i = 4	i = 5	i = 6
j = 1	{B}	{A}	{A}	{B}	{C}	{A}
j = 2						
j = 3						
j = 4						
j = 5						
j = 6						

Pentru a doua linie, j = 2, avem cuvinte de lungime 2, deci nu le putem separa decât într-un singur mod, pentru k = 1.
$V_{12} = V_{11} \times V_{21} = \{B\} \times \{A\} = \{BA\}$
În gramatică apare BA ca membru drept pentru neterminalele S şi A, deci în tabel vom completa cu mulţimea {S, A}.
$V_{22} = V_{21} \times V_{31} = \{A\} \times \{A\} = \{AA\}$
Singurul neterminal care se rescrie în AA este B, deci îl trecem în tabel.
$V_{32} = V_{31} \times V_{41} = \{A\} \times \{B\} = \{AB\}$ rezultă B în tabel.
$V_{42} = V_{41} \times V_{51} = \{B\} \times \{C\} = \{BC\}$ rezultă C în tabel.
$V_{52} = V_{51} \times V_{61} = \{C\} \times \{A\} = \{CA\}$ rezultă A în tabel.

Obs: Pentru a calcula orice valoare de pe linia j=2 se face produsul cartezian între mulţimile aflate la nord şi la nord-est de acea valoare.

V_{ij}	i = 1	i = 2	i = 3	i = 4	i = 5	i = 6
j = 1	{B}	{A}	{A}	{B}	{C}	{A}
j = 2	{S, A}	{B}	{B}	{C}	{A}	
j = 3						
j = 4						
j = 5						
j = 6						

Pentru a treia linie, j = 3, avem cuvinte de lungime 3, deci putem avea k=1 sau k=2 şi reunim cele două cazuri.
$V_{13} = V_{11} \times V_{22} \cup V_{12} \times V_{31} = \{B\} \times \{B\} \cup \{S, A\} \times \{A\} = \{BB, SA, AA\}$
Nu avem SA ca membru drept în gramatică, dar avem BB generat de C, iar AA de B, deci în tabel completăm {B, C}.

$V_{23} = V_{21} \times V_{32} \cup V_{22} \times V_{41} = \{A\} \times \{B\} \cup \{B\} \times \{B\} = \{AB, BB\}$ rezultă {B, C} în tabel.
$V_{33} = V_{31} \times V_{42} \cup V_{32} \times V_{51} = \{A\} \times \{C\} \cup \{B\} \times \{C\} = \{AC, BC\}$ rezultă {C} în tabel.
$V_{43} = V_{41} \times V_{52} \cup V_{42} \times V_{61} = \{B\} \times \{A\} \cup \{C\} \times \{A\} = \{BA, CA\}$ rezultă {S, A} în tabel.

V_{ij}	i = 1	i = 2	i = 3	i = 4	i = 5	i = 6
j = 1	{B}	{A}	{A}	{B}	{C}	{A}
j = 2	{S, A}	{B}	{B}	{C}	{A}	
j = 3	{B, C}	{B, C}	{C}	{S, A}		
j = 4						
j = 5						
j = 6						

Pentru a patra linie, j = 4, avem cuvinte de lungime 4, deci putem avea k=1, k=2 sau k=3.
$V_{14} = (V_{11} \times V_{23}) \cup (V_{12} \times V_{32}) \cup (V_{13} \times V_{41})$
$= \{B\} \times \{B,C\} \cup \{S,A\} \times \{B\} \cup \{B,C\} \times \{B\} = \{BB, BC, SB, AB, CB\}$ Nu avem SB sau CB ca membru drept în gramatică, dar pentru restul avem {B,C} în tabel.
$V_{24} = (V_{21} \times V_{33}) \cup (V_{22} \times V_{42}) \cup (V_{23} \times V_{51})$
$= \{A\} \times \{C\} \cup \{B\} \times \{C\} \cup \{B,C\} \times \{C\} = \{AC, BC, CC\}$
Nu avem AC sau CC, dar pentru BC rezultă {C} în tabel.
$V_{34} = (V_{31} \times V_{43}) \cup (V_{32} \times V_{52}) \cup (V_{33} \times V_{61})$
$= \{A\} \times \{S, A\} \cup \{B\} \times \{A\} \cup \{C\} \times \{A\} = \{AS, AA, BA, CA\}$
Nu avem AS, dar pentru celelalte rezultă {S,A,B} în tabel.

V_{ij}	i = 1	i = 2	i = 3	i = 4	i = 5	i = 6
j = 1	{B}	{A}	{A}	{B}	{C}	{A}
j = 2	{S, A}	{B}	{B}	{C}	{A}	
j = 3	{B, C}	{B, C}	{C}	{S, A}		
j = 4	{B, C}	{C}	{S, A, B}			
j = 5						
j = 6						

Pentru a cincea linie, avem cuvinte de lungime j = 5, iar k ia valori între 1 şi 4.
$V_{15} = (V_{11} \times V_{24}) \cup (V_{12} \times V_{33}) \cup (V_{13} \times V_{42}) \cup (V_{14} \times V_{51})$
$= \{B\} \times \{C\} \cup \{S, A\} \times \{C\} \cup \{B,C\} \times \{C\} \cup \{B,C\} \times \{C\} = \{BC, SC, AC, CC\}$
Nu avem SC, AC sau CC, dar pentru BC rezultă {C} în tabel.
$V_{25} = (V_{21} \times V_{34}) \cup (V_{22} \times V_{43}) \cup (V_{23} \times V_{52}) \cup (V_{24} \times V_{61})$
$= \{A\} \times \{S, A, B\} \cup \{B\} \times \{S, A\} \cup \{B,C\} \times \{A\} \cup \{C\} \times \{A\} = \{AS, AA, AB, BS, BA, CA\}$
Nu avem AS, BS dar pentru AA, AB, BA, CA rezultă {S,A,B} în tabel.

V_{ij}	i = 1	i = 2	i = 3	i = 4	i = 5	i = 6
j = 1	{B}	{A}	{A}	{B}	{C}	{A}
j = 2	{S, A}	{B}	{B}	{C}	{A}	
j = 3	{B, C}	{B, C}	{C}	{S, A}		
j = 4	{B, C}	{C}	{S, A, B}			
j = 5	{C}	{S, A, B}				
j = 6						

Pentru linia a şasea, avem cuvântul total de lungime j=6, variabila k ia valori între 1 şi 5.

$$V_{16} = (V_{11} \times V_{25}) \cup (V_{12} \times V_{34}) \cup (V_{13} \times V_{43}) \cup (V_{14} \times V_{52}) \cup (V_{15} \times V_{61})$$

$$= \{B\} \times \{S, A, B\} \cup \{S, A\} \times \{S, A, B\} \cup \{B, C\} \times \{S, A\} \cup \{B, C\} \times \{A\} \cup \{C\} \times \{A\}$$

$$= \{BS, BA, BB, SS, SA, SB, AS, AA, AB, CS, CA\}$$

Pentru BA, BB, AA, AB, CA avem în tabel {S,A,B,C}.

V_{ij}	i = 1	i = 2	i = 3	i = 4	i = 5	i = 6
j = 1	{B}	{A}	{A}	{B}	{C}	{A}
j = 2	{S, A}	{B}	{B}	{C}	{A}	
j = 3	{B, C}	{B, C}	{C}	{S, A}		
j = 4	{B, C}	{C}	{S, A, B}			
j = 5	{C}	{S, A, B}				
j = 6	{**S**,A,B,C}					

În mulţimea finală obţinută pentru V_{16} există S, deci cuvântul este generat de gramatică.

Obs: Dacă mulţimea finală conţine şi alte elemente, înseamnă că acel cuvânt dat ar putea fi generat şi pornind din acele neterminale (A, B sau C). Dar dacă existau doar ele şi nu exista şi S, acest lucru nu ne ajuta la nimic, cuvântul nu ar fi fost generat de gramatică.

Obs: Unde apare S în tabel înseamnă că gramatica generează şi cuvintele corespunzătoare pentru acele perechi (i, j).

Pentru ij = 12 avem cuvântul ”ba”: $\underline{S} \Rightarrow \underline{B}A \Rightarrow b\underline{A} \Rightarrow ba$.

Pentru ij = 25 avem ”aabca”, pentru ij = 34 avem ”abca”, pentru ij = 43 avem ”bca”.