Cocke-Younger-Kasami (CYK) algorithm

Input: $G = \langle T, N, S, P\rangle$ context free grammar in Chomsky normal form a word $u \in T^*$.

Output: $u \overset{?}{\in} L(G)$

For $u = \lambda$ $u \in L(G) \Leftrightarrow S \to \lambda \in P$.
Let $u = t_1 \dots t_n,\ t_i \in T, n \geq 1$. Let A_i be the left hand side and β_i be the right hand side of $P_i \in P$ $(A_i \in N,\ \beta_i \in T \cup N^2.)$

CYK algorithm defines the sets $H_{i,j},\ 1 \leq i \leq j \leq n$ recursively by incrementing $(j - i)$ as follows.

$$H_{i,i} := \{A_j \mid \beta_j = t_i\}$$

$$H_{i,j} := \{A_k \mid \beta_k \in \bigcup_{h=i}^{j-1} H_{i,h} H_{h+1,j}\} \quad (i < j)$$

$u \in L(G) \iff S \in H_{1,n}$.

It can be proved, that $H_{i,j} = \{X \in N \mid X \Rightarrow_G^* t_i \cdots t_j\}$.

$$\begin{array}{ccccccc}
 & & & H_{1,n} & & & \\
 & & H_{1,n-1} & & H_{2,n} & & \\
 & & & \vdots & & & \\
 & H_{1,3} & H_{2,4} & & & & \\
H_{1,2} & H_{2,3} & & & \cdots & H_{n-1,n} & \\
H_{1,1} & H_{2,2} & & & & \cdots & H_{n,n} \\
\hline
t_1 & t_2 & & \cdots & & & t_n
\end{array}$$

$\Uparrow$ bottom-up
row by row

To prove the correctness of the algorithm it is enough to prove $H_{i,j} = \{X \in N \mid X \Rightarrow_G^* t_i \cdots t_j\}$ for all $1 \leq i \leq j \leq n$.

Induction for $(j - i)$.

For $j - i = 0$ it is obvious.

Suppose that the statement holds for all $1 \leq k \leq \ell \leq n$ having the property $\ell - k < j - i$. Let $1 \leq i < j \leq n$.

Consider a derivation $X \Rightarrow^* t_i \cdots t_j$. Since the word to be derived has length at least 2 the first step of the derivation should be $X \Rightarrow YZ$ for some $Y, Z \in N$. Then there exists $i \leq h < j$ with the properties $Y \Rightarrow^* t_i \cdots t_h$ and $Z \Rightarrow^* t_{h+1} \cdots t_j$. Since $h - i < j - i$ and $j - (h + 1) < j - i$ we have $Y \in H_{i,h}, Z \in H_{h+1,j}$ by the induction implying $X \in H_{i,j}$.

On the contrary $X \in H_{i,j}$ $(j > i)$ implies the existence of an $i \leq h < j$ and $Y \in H_{i,h}, Z \in H_{h+1,j}$ with the property $X \rightarrow YZ \in P$. So $Y \Rightarrow^* t_i \cdots t_h$ and $Z \Rightarrow^* t_{h+1} \cdots t_j$. Then $X \Rightarrow YZ \Rightarrow^* t_i \cdots t_h Z \Rightarrow^* t_i \cdots t_j$.

Example: let $G = \langle\{S, A, B, C, U, V, W, X, Y, Z\}, \{a, b, c\}, S, P\rangle$ be a grammar in Chomsky NF with production rules P below. Decide by CYK algorithm whether $aabbcc$ can be generated in G or not.

$S \to AB \,|\, BC$
$A \to XA \,|\, a$
$X \to a$
$C \to YC \,|\, c$
$Y \to c$
$B \to UV \,|\, VW \,|\, XS$
$U \to XX$
$W \to YY \,|\, XS$
$V \to ZZ$
$Z \to b$

Calculating the 1st row:

$S \to AB \,|\, BC$
$A \to XA \,|\, a$
$X \to a$
$C \to YC \,|\, c$
$Y \to c$
$B \to UV \,|\, VW \,|\, XS$
$U \to XX$
$W \to YY \,|\, XS$
$V \to ZZ$
$Z \to b$

				$\{Y, C\}$	
a	a	b	b	c	c

Calculating the 2nd row:

$S \to AB \mid BC$
$A \to XA \mid a$
$X \to a$
$C \to YC \mid c$
$Y \to c$
$B \to UV \mid VW \mid XS$
$U \to XX$
$W \to YY \mid XS$
$V \to ZZ$
$Z \to b$

$$\begin{array}{cccccc}
\multicolumn{2}{c}{\{A, U\}} & & & & \\
\{A, X\} & \{A, X\} & \{Z\} & \{Z\} & \{Y, C\} & \{Y, C\} \\
\hline
a & a & b & b & c & c
\end{array}$$

$$\{A, X\}\{A, X\} = \{AA, AX, XA, XX\}$$

Calculating the 3rd row:

$S \to AB \mid BC$
$A \to XA \mid a$
$X \to a$
$C \to YC \mid c$
$Y \to c$
$B \to UV \mid VW \mid XS$
$U \to XX$
$W \to YY \mid XS$
$V \to ZZ$
$Z \to b$

$$\begin{array}{cccccccccccc}
 & & & \multicolumn{2}{c}{\{\ \}} & & & & & & & \\
 & \multicolumn{2}{c}{\{A, U\}} & \multicolumn{2}{c}{\{\ \}} & \multicolumn{2}{c}{\{V\}} & \multicolumn{2}{c}{\{\ \}} & \multicolumn{2}{c}{\{C, W\}} & \\
\multicolumn{2}{c}{\{A, X\}} & \multicolumn{2}{c}{\{A, X\}} & \multicolumn{2}{c}{\{Z\}} & \multicolumn{2}{c}{\{Z\}} & \multicolumn{2}{c}{\{Y, C\}} & \multicolumn{2}{c}{\{Y, C\}} \\
\hline
\multicolumn{2}{c}{a} & \multicolumn{2}{c}{a} & \multicolumn{2}{c}{b} & \multicolumn{2}{c}{b} & \multicolumn{2}{c}{c} & \multicolumn{2}{c}{c}
\end{array}$$

$$\{A, X\}\{\ \} = \{\ \} \quad \{A, U\}\{Z\} = \{AZ, UZ\}$$

Calculating the 4th row:

$S \to AB \mid BC$
$A \to XA \mid a$
$X \to a$
$C \to YC \mid c$
$Y \to c$
$B \to UV \mid VW \mid XS$
$U \to XX$
$W \to YY \mid XS$
$V \to ZZ$
$Z \to b$

$\{B\}$

$\{\ \}\quad \{\ \}\quad \{\ \}\quad \{\ \}$

$\{A, U\}\quad \{\ \}\quad \{V\}\quad \{\ \}\quad \{C, W\}$

$\{A, X\}\quad \{A, X\}\quad \{Z\}\quad \{Z\}\quad \{Y, C\}\quad \{Y, C\}$

$a\quad a\quad b\quad b\quad c\quad c$

$\{Z\}\{\ \} = \{\ \}$ $\{V\}\{C, W\} = \{VC, VW\}$ $\{\ \}\{Y, C\} = \{\ \}$

Calculating the 5th row:

$S \to AB \mid BC$
$A \to XA \mid a$
$X \to a$
$C \to YC \mid c$
$Y \to c$
$B \to UV \mid VW \mid XS$
$U \to XX$
$W \to YY \mid XS$
$V \to ZZ$
$Z \to b$

$\{S\}$

$\{B\}\quad \{\ \}\quad \{B\}$

$\{\ \}\quad \{\ \}\quad \{\ \}\quad \{\ \}$

$\{A, U\}\quad \{\ \}\quad \{V\}\quad \{\ \}\quad \{C, W\}$

$\{A, X\}\quad \{A, X\}\quad \{Z\}\quad \{Z\}\quad \{Y, C\}\quad \{Y, C\}$

$a\quad a\quad b\quad b\quad c\quad c$

$\{\ AB, XB\ \}$ $\{\ \}$ $\{\ \}$ $\{\ \}$

Calculating the 6th row:

$$
\begin{array}{l}
S \to AB \mid BC \\
A \to XA \mid a \\
X \to a \\
C \to YC \mid c \\
Y \to c \\
B \to UV \mid VW \mid XS \\
U \to XX \\
W \to YY \mid XS \\
V \to ZZ \\
Z \to b
\end{array}
$$

$$
\begin{array}{cccccccccccc}
 & & & & & \{S, B, W\} & & & & & \\
 & & & & \{S\} & & \{S\} & & & & \\
 & & & \{B\} & & \{\,\} & & \{B\} & & & \\
 & & \{\,\} & & \{\,\} & & \{\,\} & & \{\,\} & & \\
 & \{A, U\} & & \{\,\} & & \{V\} & & \{\,\} & & \{C, W\} & \\
\{A, X\} & & \{A, X\} & & \{Z\} & & \{Z\} & & \{Y, C\} & & \{Y, C\} \\
\hline
a & & a & & b & & b & & c & & c
\end{array}
$$

$$
\{\, AS, XS\} \quad \{\, AB, UB \,\} \quad \{\,\} \quad \{\, BC, BW \,\} \quad \{\, SY, SC \,\}
$$

The answer:

$$
\begin{array}{l}
S \to AB \mid BC \\
A \to XA \mid a \\
X \to a \\
C \to YC \mid c \\
Y \to c \\
B \to UV \mid VW \mid XS \\
U \to XX \\
W \to YY \mid XS \\
V \to ZZ \\
Z \to b
\end{array}
$$

$$
\begin{array}{cccccccccccc}
 & & & & & \{S, B, W\} & & & & & \\
 & & & & \{S\} & & \{S\} & & & & \\
 & & & \{B\} & & \{\,\} & & \{B\} & & & \\
 & & \{\,\} & & \{\,\} & & \{\,\} & & \{\,\} & & \\
 & \{A, U\} & & \{\,\} & & \{V\} & & \{\,\} & & \{C, W\} & \\
\{A, X\} & & \{A, X\} & & \{Z\} & & \{Z\} & & \{Y, C\} & & \{Y, C\} \\
\hline
a & & a & & b & & b & & c & & c
\end{array}
$$

Since $S \in H_{1,6}$ we have $aabbcc \in L(G)$.

A derivation:

$$S \Rightarrow AB \Rightarrow XAB \Rightarrow XAVW \Rightarrow XAZZW \Rightarrow XAZZYY \Rightarrow^{*} aabbcc.$$

Exercises for practice (nonterminals: capital letters, initial symbol: S)

Exercise 1:

$$
\begin{aligned}
S &\to AA \,|\, YK \\
A &\to CY \,|\, YK \,|\, CS \\
C &\to AY \,|\, c \\
Y &\to YA \,|\, y \\
K &\to k
\end{aligned}
$$

Word: $cykcyk$

Exercise 2:

$$
\begin{aligned}
S &\to AA \,|\, CD \\
A &\to BC \,|\, CD \,|\, BS \\
B &\to AC \,|\, a \\
C &\to CA \,|\, b \\
D &\to DA \,|\, c
\end{aligned}
$$

Word: $abcabc$