1.3.1 Нормална форма на Чомски

Една граматика $G' = (V, \Sigma, P, S)$ е в нормална форма на Чомски, ако
$P \subseteq V \times \Sigma \cup V \times VV$.

- ε-елиминиране
- елиминиране на (единичните) правила от вида $A \longrightarrow B$
- елиминиране на правилата с дълга дясна част

ε-елиминация

$\forall G = (V, \Sigma, P, S)$ с $P \subseteq V \times (V \cup \Sigma)^*$:

$\exists G' : L(G) = L(G'), G'$ е от тип 2

ε елиминация за $G = (V, \Sigma, P, S)$

$V_\varepsilon := \{\}$

while $\exists X \to \alpha \in P : X \notin V_\varepsilon \land \alpha \in V_\varepsilon^*$ do $V_\varepsilon := V_\varepsilon \cup \{X\}$

assert $V_\varepsilon = \left\{ X \in V : X \stackrel{*}{\Rightarrow} \varepsilon \right\}$

while $\exists X \to \alpha Y \beta \in P : Y \in V_\varepsilon \land X \to \alpha\beta \notin P$ do

$\quad P := P \cup X \to \alpha\beta$ // invariant : $L(G)$

$P := P \setminus (V \times \{\varepsilon\})$ // invariant : $L(G) \setminus \{\varepsilon\}$

if $S \in V_\varepsilon$ then return $(V \cup \{S'\}, \Sigma, P \cup \{S' \to \varepsilon, S' \to S\}, S')$

else return (V, Σ, P, S)

Упражнение: Линейно време за откриване на V_ε.

$(\mathcal{O}(|V| + \sum_{X \to r \in P} |r|))$

Упражнение a^*b^*

$$G = (\{S, A, B\}, \{a, b\}, P, S),$$
$$P = \{S \longrightarrow AB, A \longrightarrow aA, A \longrightarrow \varepsilon, B \longrightarrow bB, B \longrightarrow \varepsilon\}$$

while $\exists X \longrightarrow \alpha \in P : X \notin V_\varepsilon \wedge \alpha \in V_\varepsilon^*$ do $V_\varepsilon := V_\varepsilon \cup \{X\}$

$$V_\varepsilon : \{\} \rightsquigarrow \{A\} \rightsquigarrow \{A, B\} \rightsquigarrow \{A, B, S\}$$

Упражнение a^*b^*

$$G = (\{S, A, B\}, \{a, b\}, P, S),$$
$$P = \{S \to AB, A \to aA, A \to \varepsilon, B \to bB, B \to \varepsilon\}$$
$$V_\varepsilon = \{A, B, S\}$$

while $\exists X \to \alpha Y \beta \in P : Y \in V_\varepsilon \wedge X \to \alpha\beta \notin P$ do $P := P \cup X \to \alpha\beta$

$A \to aA, A \in V_\varepsilon \rightsquigarrow A \to a$

$B \to bB, B \in V_\varepsilon \rightsquigarrow B \to b$

$S \to AB, A \in V_\varepsilon \rightsquigarrow S \to B$

$S \to AB, B \in V_\varepsilon \rightsquigarrow S \to A$

$S \to A, A \in V_\varepsilon \rightsquigarrow S \to \varepsilon$

$S \to B, B \in V_\varepsilon$ but $S \to \varepsilon \in P$

$$P = \{S \to AB, A \to aA, A \to \varepsilon, B \to bB, B \to \varepsilon,$$
$$A \to a, B \to b, S \to B, S \to A, S \to \varepsilon\}$$

Упражнение a^*b^*

$$G = (\{S,A,B\},\{a,b\},P,S),$$
$$P = \{S \to AB, A \to aA, A \to \varepsilon, B \to bB, B \to \varepsilon\}$$
$$V_\varepsilon = \{A,B,S\}$$
$$P := \{S \to AB, A \to aA, A \to \varepsilon, B \to bB, B \to \varepsilon,$$
$$A \to a, B \to b, S \to B, S \to A, S \to \varepsilon\}$$

$$P := P \setminus (V \times \{\varepsilon\})$$

$$P := \{S \to AB, A \to aA, B \to bB,$$
$$A \to a, B \to b, S \to B, S \to A\}$$

Return $(\{S', S, A, B\},\{a,b\},P,S')$,

$$P := \{S' \to \varepsilon, S' \to S, S \to AB, A \to aA, B \to bB,$$
$$A \to a, B \to b, S \to B, S \to A\}$$

Доказателство за коректност

- $X \in V_\varepsilon \longrightarrow X \stackrel{*}{\Rightarrow} \varepsilon$: даден извод
- $X \stackrel{*}{\Rightarrow} \varepsilon \longrightarrow X \in V_\varepsilon$: индукция по дължината на извода
- Инварианта на цикъла
- Завършване !
- Елиминация на ε-правилата не променя $L(G) \setminus \{\varepsilon\}$: Индукция по дължината на извода. Заменяме правилото $\gamma X \delta \stackrel{X \to \alpha Y \beta}{\Rightarrow} \gamma \alpha Y \beta \delta \stackrel{Y \to \varepsilon}{\Rightarrow} \gamma \alpha \beta \delta$ с $\gamma X \delta \stackrel{X \to \alpha \beta}{\Rightarrow} \gamma \alpha \beta \delta$.

Доказателство за коректност — завършване

assert $V_\varepsilon = \left\{X \in V : X \overset{*}{\Rightarrow} \varepsilon\right\}$

while $\exists X \to \alpha Y \beta \in P : Y \in V_\varepsilon \wedge X \to \alpha\beta \notin P$ do

$\quad P := P \cup \{X \to \alpha\beta\}$

Нека $k := \max\{|r| : X \to r \in P\}$.

Наблюдение: Има нови правила $X \to w \in P$, $|w| < k$.

Но са само краен брой правила с ограничена дължина.

Елиминация на цикличните единични правила

$G = (V, \Sigma, P, S)$ контекстно-свободна без ε-правила
Да разгледаме графа $U = (V, P \cap V \times V)$ на единичните правила.

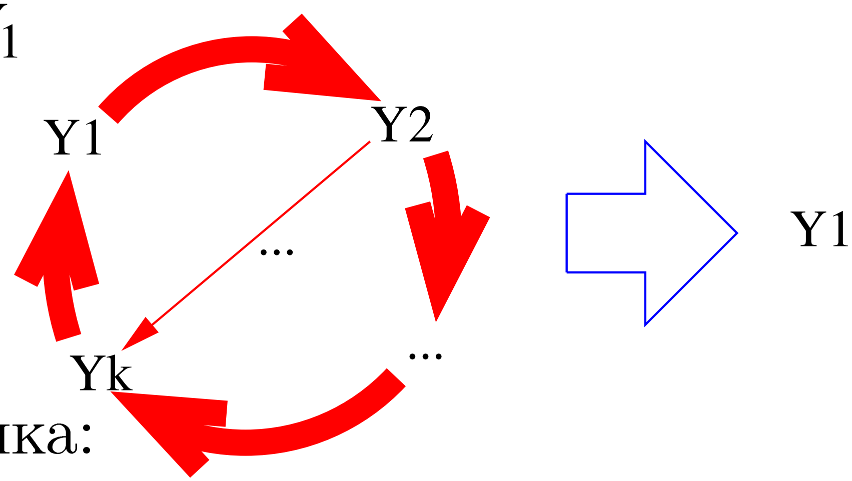
while $\exists$cycle $\{Y_1, \dots, Y_k\}$ in U, $k \geq 1$ do

  invariant : $L(G)$

  replace $Y_2, \dots, Y_k$ in G с Y_1

  $P := P \setminus \{Y_1 \to Y_1\}$

assert U is cycle-free



Граф-теоретична гледна точка:
контракция(редуциране) на силно свързаните компоненти

Елиминация на нецикличните единични правила

invariant : Графът на единичните правила U няма цикли

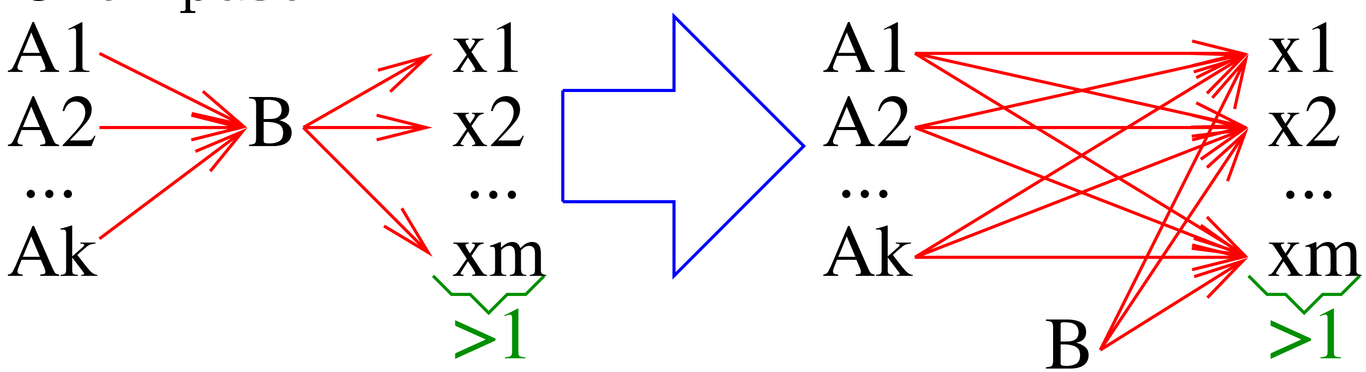
while $\exists X \rightarrow Y \in P \cap V \times V : P \cap \{Y\} \times V = \emptyset$ do

invariant : $L(G)$

$P := P \cup \{X \rightarrow x : X \rightarrow Y \in P \wedge Y \rightarrow x \in P\}$

$P := P \setminus V \times \{Y\}$

assert U е празен



Граф-теоретична гледна точка: Нареждаме ги в обратна на топологична сортировка ред.

Нормална форма на Чомски

Една граматика $G' = (V, \Sigma, P, S)$ е в нормална форма на Чомски, ако
$P \subseteq V \times \Sigma \cup V \times VV$.

Твърдение: За всяка контекстно-свободна гарматика G с $\varepsilon \notin L(G)$, $\exists G'$ в нормална форма на Чомски, такава че $L(G) = L(G')$.

Д-во на твърдението: Стъпка по стъпка променяме граматиката G.
Инвариант: $L(G)$ остава непроменен.

1. ε елиминиране
2. елиминиране на единичните правила.

Елиминация на смесените десни страни

foreach $a \in \Sigma$ do

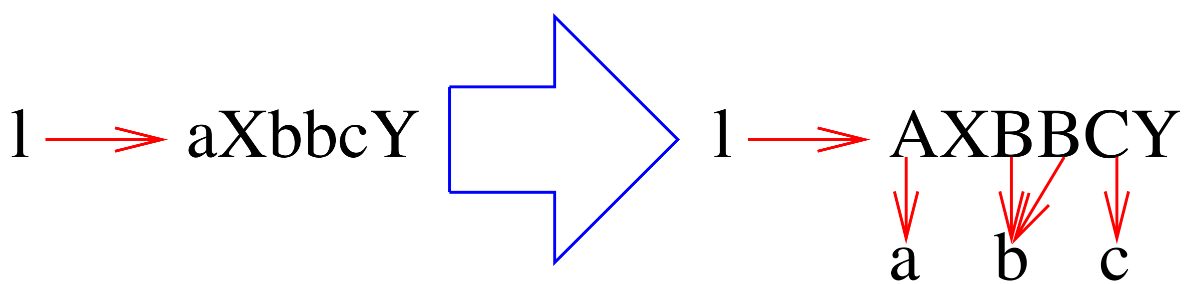
  $V_a :=$ new variable

  $P := P \cup \{V_a \to a\}$

  foreach $\ell \to r \in P$ do

    if $a \in r \wedge |r| \geq 2$ then replace a с V_a in $V \to r$

assert $P \subseteq V \times \Sigma \cup V \times V^*$



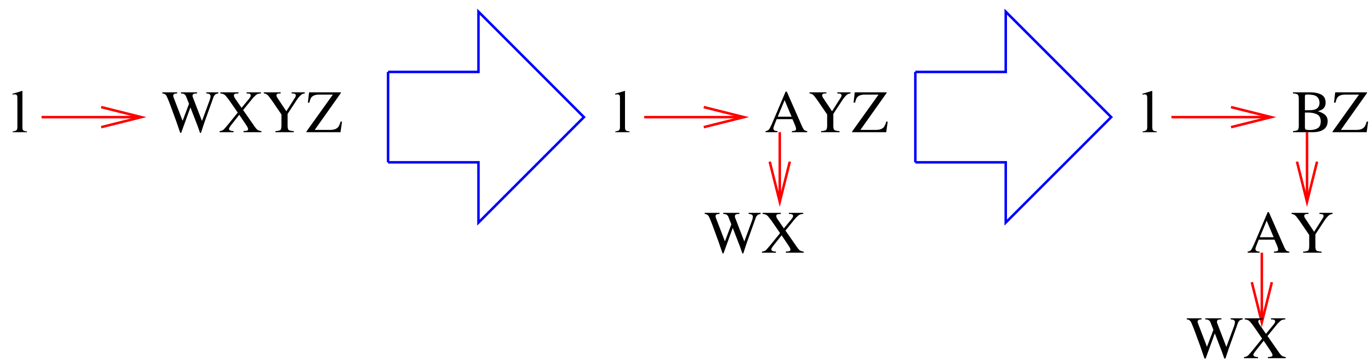
Елиминация на дългите десни страни

while $\exists X \to Y_1 Y_2 Y_3 \cdots Y_k \in P$ с $k \geq 3$ do

$C :=$ new variable

$P := P \cup \{C \to Y_1 Y_2, X \to C Y_3 \cdots Y_k\} \setminus \{X \to Y_1 Y_2 \cdots Y_k\}$

Цикълът завършва, тъй като $\sum_{\ell \to r \in P} \max(0, |r| - 2)$ намалява.



Упражнение

$$\{E \to E+E, E \to E*E, E \to (E), E \to a\}$$

$\rightsquigarrow$

$$\{E \to EV_+E, E \to EV_*E, E \to V_(EV_), E \to a,$$
$$V_+ \to +, V_* \to *, V_(\to (, V_) \to)\}$$

$\rightsquigarrow$

$$\begin{aligned}
\{&C \to EV_+, E \to CE,\\
&D \to EV_*, E \to DE,\\
&F \to V_(E, E \to FV_),\\
&E \to a,\\
&V_+ \to +, V_* \to *, V_(\to (, V_) \to)\}
\end{aligned}$$