# Počáteční gramatika

$S \to ABC \mid aSb$

$A \to BB \mid \lambda$

$B \to C \mid \lambda$

$C \to A \mid cC \mid D$

$D \to dD$

$E \to eS$

# Ukončitelné neterminály

**Info**

Hledáme množinu $N_t$, tj. neterminály, které mohou odvodit řetězec z terminálů (příp. $\lambda$). Množinu získáme iterativně:

1. Inicializace: $N_t^0 = \emptyset$
2. Přidáme neterminály s pravidlem obsahujícím pouze terminály nebo $\lambda$:
$$N_t^1 = N_t^0 \cup \{A \mid A \Rightarrow \alpha, \alpha \in T^* \cup \{\lambda\}\}$$
3. Přidáme neterminály, které mají pravidlo složené z terminálů a neterminálů z předchozí množiny:
$$N_t^2 = N_t^1 \cup \left\{A \mid A \Rightarrow \alpha, \alpha \in \left(T \cup N_t^1\right)^*\right\}$$
4. Opakujeme krok 3 dokud množina roste, výsledkem je $N_t$.
5. Odstraníme neterminály mimo $N_t$.

$N_t^0 = \emptyset$

$N_t^1 = \{A, B\}$

$N_t^2 = \{S, A, B, C\}$

$N_t^3 = \{S, A, B, C, E\}$

$N_t^4 = \{S, A, B, C, E\}$

$S \to ABC \mid aSb$

$A \to BB \mid \lambda$

$B \to C \mid \lambda$

$C \to A \mid cC$

$E \to eS$

# Dosažitelné symboly

**Info**

Hledáme množinu $V$, tj. symboly dosažitelné z počátečního symbolu $S$. Množinu získáme iterativně:

1. Inicializace: $V_0 = \{S\}$
2. Přidáme symboly z pravých stran pravidel pro $X \in V_0$
3. Opakujeme krok 2 dokud množina roste, výsledkem je $V$.
4. Odstraníme symboly mimo $V$ a pravidla, která je obsahují.

$$V_0 = \{S\}$$

$$V_1 = \{S, A, B, C, a, b, \lambda\}$$

$$V_2 = \{S, A, B, C, a, b, c, \lambda\}$$

$$V_3 = \{S, A, B, C, a, b, c, \lambda\}$$

$$S \to ABC \mid aSb$$

$$A \to BB \mid \lambda$$

$$B \to C \mid \lambda$$

$$C \to A \mid cC$$

# Odstranění $\lambda$-pravidel

**Info**

1. Sestrojíme množinu $N$, obsahující všechny neterminály, které jsou "nullable" (vede od nich nějaká cesta k $\lambda$, která neobsahuje žádné terminály):
   - neterminály s pravidly druhu $X \to \lambda$ jsou nullable
   - neterminály s pravidly druhu $X \to A_1 A_2 ... A_k$ jsou nullable právě když všechny neterminály $A_1, A_2 ... A_k$ jsou též nullable
2. Ke všem pravidlům přidáme pravidla vzniklá nahrazením nullable neterminálů lambdami (kromě $X \to \lambda$)
3. musí-li gramatika též akceptovat $\lambda$, zavedeme nový startovací terminál $S_0$ s pravidlem:

$$S_0 \to S \mid \lambda$$

$N_\lambda^0 = \{A, B, C\}$

$N_\lambda^1 = \{S, A, B, C\}$

$N_\lambda^2 = \{S, A, B, C\}$

$S_0 \to S \mid \lambda$

$S \to ABC \mid AB \mid BC \mid AC \mid A \mid B \mid C \mid aSb \mid ab$

$A \to BB \mid B$

$B \to C$

$C \to A \mid cC \mid c$

# Odstranění jednoduchých pravidel

**Info**

- Pro každý neterminál X:
  1. najdeme množinu $U$ všech neterminálů, ke kterým lze dojít po jednoduchých pravidlech
     I. přidáme do množiny každý neterminál dosažitelný jednoduchými pravidly pro $X$:
     $$U_0 = \{Y : X \to Y\}$$
     II. přidáme do množiny každý neterminál dosažitelný pravidly pro neterminály v předchozí množině:
     $$U_{i+1} = U_i \cup \{Z : Y \to Z, Y \in U_i\}$$
     III. opakujeme krok III, dokud množina roste
  2. přidáme k neterminálu $X$ všechny produkce neterminálů v $U$
- nakonec odebereme všechna jednoduchá pravidla (duplicity můžeme vynechat)

pozn.: Nemá-li $X$ žádná jednoduchá pravidla, můžeme ho přeskočit

$S_0 \to S \mid \lambda$
$S \to ABC \mid AB \mid BC \mid AC \mid A \mid B \mid C \mid aSb \mid ab$
$A \to BB \mid B$
$B \to C$
$C \to A \mid cC \mid c$

## Terminál $S_0$

$U_0 = \{S\}$
$U_1 = \{S, A, B, C\}$
$U_2 = \{A, B, C\}$
$S_0 \to S \mid ABC \mid AB \mid BC \mid AC \mid aSb \mid ab \mid BB \mid cC \mid c \mid \lambda$
$S \to ABC \mid AB \mid BC \mid AC \mid A \mid B \mid C \mid aSb \mid ab$
$A \to BB \mid B$
$B \to C$
$C \to A \mid cC \mid c$

## Terminál $S$

$U_0 = \{A, B, C\}$
$U_1 = \{A, B, C\}$

$S_0 \to S \mid ABC \mid AB \mid BC \mid AC \mid aSb \mid ab \mid BB \mid cC \mid c \mid \lambda$
$S \to ABC \mid AB \mid BC \mid AC \mid A \mid B \mid C \mid aSb \mid ab \mid BB \mid cC \mid c$
$A \to BB \mid B$
$B \to C$
$C \to A \mid cC \mid c$

## Terminál $A$

$U_0 = \{B\}$
$U_1 = \{B, C\}$
$U_2 = \{A, B, C\}$
$U_3 = \{A, B, C\}$

$S_0 \to S \mid ABC \mid AB \mid BC \mid AC \mid aSb \mid ab \mid BB \mid cC \mid c \mid \lambda$
$S \to ABC \mid AB \mid BC \mid AC \mid A \mid B \mid C \mid aSb \mid ab \mid BB \mid cC \mid c$
$A \to BB \mid B \mid cC \mid c$
$B \to C$
$C \to A \mid cC \mid c$

## Terminál $B$

$U_0 = \{C\}$
$U_1 = \{A, C\}$
$U_2 = \{A, B, C\}$
$U_3 = \{A, B, C\}$

$S_0 \to S \mid ABC \mid AB \mid BC \mid AC \mid aSb \mid ab \mid BB \mid cC \mid c \mid \lambda$
$S \to ABC \mid AB \mid BC \mid AC \mid A \mid B \mid C \mid aSb \mid ab \mid BB \mid cC \mid c$
$A \to BB \mid B \mid cC \mid c$
$B \to C \mid BB \mid cC \mid c$
$C \to A \mid cC \mid c$

## Terminál $C$

$U_0 = \{A\}$
$U_1 = \{A, B\}$
$U_2 = \{A, B, C\}$
$U_3 = \{A, B, C\}$

$S_0 \to S \mid ABC \mid AB \mid BC \mid AC \mid aSb \mid ab \mid BB \mid cC \mid c \mid \lambda$
$S \to ABC \mid AB \mid BC \mid AC \mid A \mid B \mid C \mid aSb \mid ab \mid BB \mid cC \mid c$
$A \to BB \mid B \mid cC \mid c$
$B \to C \mid BB \mid cC \mid c$
$C \to A \mid BB \mid cC \mid c$

## Finální odstranění jednoduchých pravidel a doplnění počáteční lambdy

$S_0 \to ABC \mid AB \mid BC \mid AC \mid aSb \mid ab \mid BB \mid cC \mid c \mid \lambda$
$S \to ABC \mid AB \mid BC \mid AC \mid aSb \mid ab \mid BB \mid cC \mid c$
$A \to BB \mid cC \mid c$
$B \to BB \mid cC \mid c$
$C \to BB \mid cC \mid c$

# Chomského normální forma (CNF)

**Info**

- Gramatika je v Chomského normální formě, obsahuje-li pouze pravidla převádějící jeden neterminál buď na dva neterminály nebo jeden terminál:

$$A \to BC, \qquad A \to a$$
$$A, B, C \in N, a \in \Sigma$$

- Nevyhovující pravidla nahrazujeme následovně:
  - vyskytuje-li se terminál $a$ na pravé straně pravidla s dalšími symboly (např. $A \to aBC$), nahrazujeme ho novým neterminálem $a'$ s produkcí $a' \to a$:

$$A \to a'BC, a' \to a$$
$$a' \in N$$

  - je-li pravidlo složeno ze tří a více neterminálů, rozdělíme ji pomocí nového neterminálu na dvojice (příklad z předchozího kroku):

$$A \to XBC \Rightarrow A \to \langle XB \rangle C; \langle XB \rangle \Rightarrow XB$$
$$\langle XB \rangle \in N$$

  - pokračujeme, dokud existují nevyhovující pravidla.
- Výsledná gramatika je v CNF.

## Nová pravidla

$S_0 \to ABC \Rightarrow S_0 - \langle AB \rangle C; \langle AB \rangle \to AB$

$S_0 \to aSb \Rightarrow S_0 \to a'Sb'; a' \to a; b' \to b$

$S_0 \to ab \Rightarrow S_0 \to a'b'$

$S_0 \to cC \Rightarrow S_0 \to c'C; c' \to c$

$S_0 \to a'Sb' \Rightarrow S_0 \to \langle a'S \rangle b'; \langle a'S \rangle \to a'S$

$S \to ABC \Rightarrow S \to \langle AB \rangle$

$S \to aSb \Rightarrow a'Sb'$

$S \to ab \Rightarrow S \to a'b'$

$S \to cC \Rightarrow S \to c'C$

$S \to a'Sb' \Rightarrow S \to \langle a'S \rangle b'$

$A \to cC \Rightarrow A \to c'C$

$B \to cC \Rightarrow B \to c'C$

$C \to cC \Rightarrow C \to c'C$

## Výsledná gramatika v CNF

$S_0 \to AB \mid BC \mid AC \mid BB \mid c \mid \langle AB \rangle C \mid a'b' \mid c'C \mid \langle a'S \rangle b' \mid \lambda$

$S \to AB \mid BC \mid AC \mid BB \mid c \mid \langle AB \rangle C \mid a'b' \mid c'C \mid \langle a'S \rangle b'$

$A \to BB \mid c'C \mid c$

$B \to BB \mid c'C \mid c$

$C \to BB \mid c'C \mid c$

$\langle AB \rangle \to AB$

$\langle a'S \rangle \to a'S$

$a' \to a$

$b' \to b$

$c' \to c$

# Greibachové normální forma (GNF)

## krok zpět (převod z CNF do GNF by byl bolestivý)

$S_0 \to ABC \mid AB \mid BC \mid AC \mid aSb \mid ab \mid BB \mid cC \mid c \mid \lambda$

$S \to ABC \mid AB \mid BC \mid AC \mid aSb \mid ab \mid BB \mid cC \mid c$

$A \to BB \mid cC \mid c$

$B \to BB \mid cC \mid c$

$C \to BB \mid cC \mid c$

## Odstranění levé rekurze

**Info**

- existuje-li pravidlo tvaru:

$$A \to A\alpha_1 \mid \ldots \mid A\alpha_n \mid \beta_1 \mid \ldots \mid \beta_m$$

  kde:
  - $\alpha_i$ jsou neprázdné řetězce symbolů,
  - $\beta_i$ jsou řetězce symbolů, které nezačínají $A$,

  nahradíme jej následujícími pravidly:

$$A \to \beta_1 A' \mid \ldots \mid \beta_m A'$$
$$A' \to \alpha_1 A' \mid \ldots \mid \alpha_n A' \mid \lambda$$

- výsledná gramatika neobsahuje levou rekurzi

$B \to BB \mid cC \mid c \Rightarrow B \to cCB' \mid cB'; B' = BB'$

### Gramatika bez levé rekurze

$S_0 \to ABC \mid AB \mid BC \mid AC \mid aSb \mid ab \mid BB \mid cC \mid c \mid \lambda$

$S \to ABC \mid AB \mid BC \mid AC \mid aSb \mid ab \mid BB \mid cC \mid c$

$A \to BB \mid cC \mid c$

$B \to BB' \mid cCB \mid cB$

$C \to BB \mid cC \mid c$

$B' \to BB'$

## Převod do GNF

**Info**

- Gramatika je v GNF, obsahuje-li pouze pravidla tvaru $A \to a\beta$, kde $\beta$ je libovolný řetězec neterminálů
- Obsahuje-li gramatika pravidlo začínající netereminálem, nahrazujeme ho jeho produkcemi:

$$A \to B\gamma \Rightarrow A \to b\beta\gamma$$
$$\text{pro všechna } B \to b\beta$$
$$A, B \in N; b \in \Sigma$$

- Obsahuje-li pravidlo terminál za prvním symbolem, nahradíme ho neterminálem, který daný terminál produkuje:

$$A \to aAbc \Rightarrow S \to aAb'c', b' \to b, c' \to c$$

- Opakujeme předchozí pravidlo, dokud se neobjeví na začátku terminál
- výsledná gramatika je v GNF

**Pozor**

obsahuje-li původní jazyk prázdný řetězec, musíme ponechat pravidlo $S \to \lambda$, nevyskytuje-li se $S$ na pravé straně jiného pravidla (S zastupuje počáteční neterminál, v našem případě $S_0$). Ostatní pravidla s $\lambda$ se při převodu odstraňují.

$S_0 \to ABC \Rightarrow S_0 \to cCBC \mid cBC$

$S_0 \to AB \Rightarrow S_0 \to cCB \mid cB$

$S_0 \to BC \Rightarrow S_0 \to cCBC \mid cBC$

$S_0 \to AC \Rightarrow S_0 \to cCC \mid cC$

$S_0 \to aSb \Rightarrow S_0 \to aSb'; b' \to b$

$S_0 \to ab \Rightarrow S_0 \to ab' S_0 \to BB \Rightarrow S_0 \to cCBB \mid cBB$

$S \to ABC \Rightarrow S \to cCBC \mid cBC$

$S \to AB \Rightarrow S \to cCB \mid cB$

$S \to BC \Rightarrow S \to cCBC \mid cBC$

$S \to AC \Rightarrow S \to cCC \mid cC$

$S \to aSb \Rightarrow S \to aSb'$

$S \to ab \Rightarrow S \to ab'$

$S \to BB \Rightarrow S \to cCBB \mid cBB$

$A \to BB \Rightarrow A \to cCBB \mid cBB$

$B \to BB' \Rightarrow B \to cCBB' \mid cBB'$

$C \to BB \Rightarrow C \to cCBB'B \mid cBB'B \mid cCBB \mid cBB$

$B' \to BB' \Rightarrow B' \to cCBB'B' \mid cBB'B' \mid cCBB' \mid cBB'$

# Gramatika v GNF

$S_0 \to cCBC \mid cBC \mid cCB \mid cB \mid cCBC \mid cBC \mid cCC \mid cC \mid aSb' \mid ab' \mid cCBB \mid cBB \mid cC \mid c \mid \lambda$

$S \to cCBC \mid cBC \mid cCB \mid cB \mid cCBC \mid cBC \mid cCC \mid cC \mid aSb' \mid ab' \mid cCBB \mid cBB \mid cC \mid c$

$A \to cCBB \mid cBB \mid cC \mid c$

$B \to cCBB' \mid cBB' \mid cCB \mid cB$

$C \to cCBB'B \mid cBB'B \mid cCBB \mid cBB \mid cC \mid c$

$B' \to cCBB'B' \mid cBB'B' \mid cCBB' \mid cBB'$

$b' \to b$