Kapitola XII.
Normální formy a vlastnosti bezkontextových jazyků

Chomského normální forma (CNF)

Definice: Nechť $G = (N, T, P, S)$ je BKG.
G je v *Chomského normální formě*, pokud každé pravidlo z P má jeden ze tvarů:

- $A \rightarrow BC$, kde $A, B, C \in N$;
- $A \rightarrow a$, kde $A \in N$, $a \in T$;

Příklad:

$G = (N, T, P, S)$, kde $N = \{A, B, C, S\}$, $T = \{a, b\}$,
$P = \{S \rightarrow CB, C \rightarrow AS, S \rightarrow AB, A \rightarrow a, B \rightarrow b\}$,
je v Chomského normální formě.

Pozn.: $L(G) = \{a^n b^n : n \geq 1\}$

Greibachové normální forma (GNF)

Definice: Nechť $G = (N, T, P, S)$ je BKG.
G je v *Greibachové normalní formě*, pokud každé pravidlo z P má následující tvar:

- $\boldsymbol{A \rightarrow ax}$, kde $A \in N$, $a \in T$, $x \in N^*$

Příklad:

$G = (N, T, P, S)$, kde $N = \{\boldsymbol{B}, \boldsymbol{S}\}$, $T = \{\boldsymbol{a}, \boldsymbol{b}\}$,
$P = \{\mathbf{S \rightarrow aSB}, \mathbf{S \rightarrow aB}, \mathbf{B \rightarrow b}\}$
je v Greibachové normální formě.

Pozn.: $L(G) = \{a^n b^n : n \geq 1\}$

Generativní síla normálních forem

Tvrzení: Pro každou BKG existuje ekvivalentní gramatika *G'* v Chomského normální formě.

Důkaz: Viz str. 348 v knize [Meduna: Automata and Languages]

Tvrzení: Pro každou BKG existuje ekvivalentní gramatika *G'* v Greibachové normální formě.

Důkaz: Viz str. 376 v knize [Meduna: Automata and Languages]

Pozn.: Základní vlastnosti CNF a GNF:

CNF: pokud $S \Rightarrow^{n} w$; $w \in T^*$ **potom** $n = 2|w| - 1$

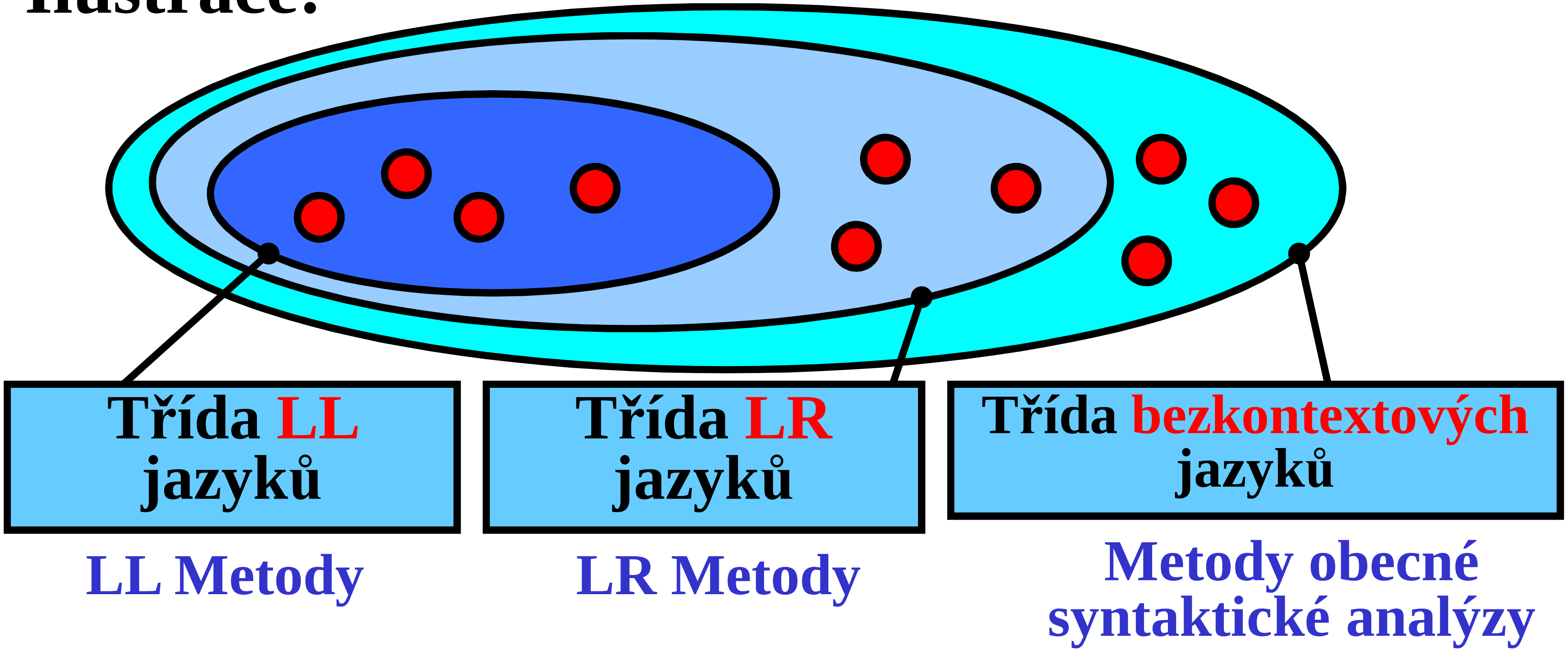
GNF: pokud $S \Rightarrow^{n} w$; $w \in T^*$ **potom** $n = |w|$

Metody obecné syntaktické analýzy

• **Metody obecné syntaktické analýzy** mohou být použity pro libovolný bezkontextový jazyk

---

Ilustrace:

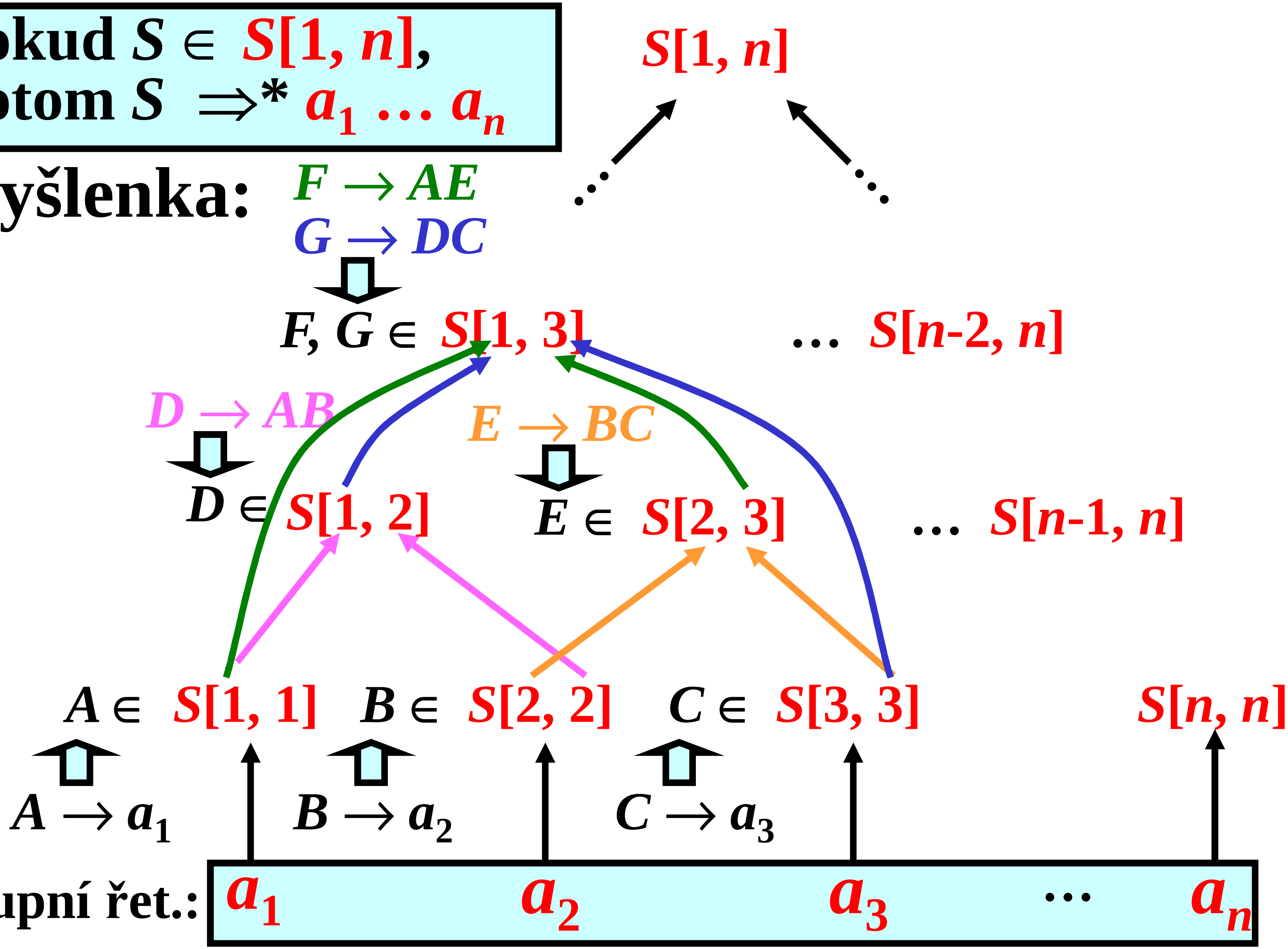


• **Pozn.:** Třída **LR** jazyků =
třída **deterministických jazyků**

Obecná SA založená na CNF

pokud $S \in S[1, n]$,
potom $S \Rightarrow^* a_1 \dots a_n$

Myšlenka:



• **Vstupní řet.:**

Algoritmus: Obecná SA založená na CNF

- **Vstup:** $G = (N, T, P, S)$ v CNF, $\boldsymbol{w} = \boldsymbol{a_1 \ldots a_n}$
- **Výstup:** **ANO**, pokud $w \in L(G)$
  NE, pokud $w \notin L(G)$

- **Metoda:**
- pro každé $\boldsymbol{a_i}$, kde $i = 1, \ldots, n$:
  $\boldsymbol{S}[\boldsymbol{i}, \boldsymbol{i}] := \{\boldsymbol{A} : \boldsymbol{A} \rightarrow \boldsymbol{a_i} \in \boldsymbol{P}\}$
- Aplikuj následující pravidlo, dokud žádná z množin $S[i, k]$ nemůže být změněna:
  <u>if</u> $\boldsymbol{A} \rightarrow \boldsymbol{BC} \in \boldsymbol{P}$, $\boldsymbol{B} \in \boldsymbol{S}\,[\boldsymbol{i}, \boldsymbol{j}]$, $\boldsymbol{C} \in \boldsymbol{S}\,[\boldsymbol{j}+1, \boldsymbol{k}]$,
  kde $1 \leq \boldsymbol{i} \leq \boldsymbol{j} < \boldsymbol{k} \leq \boldsymbol{n}$ **<u>then</u>** přidej $\boldsymbol{A}$ do $\boldsymbol{S}[\boldsymbol{i}, \boldsymbol{k}]$
- **<u>if</u>** $\boldsymbol{S} \in \boldsymbol{S}[1, \boldsymbol{n}]$ **<u>then</u>** napiš('**ANO**')
  <u>else</u> napiš('**NE**')

Obecná SA založená na CNF: Příklad 1/5

$G = (N, T, P, S)$, kde $N = \{A, B, C, S\}$, $T = \{a, b, c\}$,
$P = \{S \to AC, C \to SB, A \to a, B \to b, S \to c\}$

Otázka: $aacbb \in L(G)$**?**

$S[1, 1]=\{A\}$ $S[2, 2]=\{A\}$ $S[3, 3]=\{S\}$ $S[4, 4]=\{B\}$ $S[5, 5]=\{B\}$

$A \to a$ $A \to a$ $S \to c$ $B \to b$ $B \to b$

a *a* *c* *b* *b*

Obecná SA založená na CNF: Příklad 2/5

$G = (N, T, P, S)$, kde $N = \{A, B, C, S\}$, $T = \{a, b, c\}$,

$P = \{S \rightarrow AC, C \rightarrow SB, A \rightarrow a, B \rightarrow b, S \rightarrow c\}$

Otázka: $aacbb \in L(G)$**?**

$? \cancel{\rightarrow} AA$ $\quad ? \cancel{\rightarrow} AS$ $\quad C \rightarrow SB$ $\quad ? \cancel{\rightarrow} BB$

$S[1, 2]=\emptyset$ $\quad S[2, 3]=\emptyset$ $\quad S[3, 4]=\{C\}$ $\quad S[4, 5]=\emptyset$

$S[1, 1]=\{A\}$ $\quad S[2, 2]=\{A\}$ $\quad S[3, 3]=\{S\}$ $\quad S[4, 4]=\{B\}$ $\quad S[5, 5]=\{B\}$

a	*a*	*c*	*b*	*b*

Obecná SA založená na CNF: Příklad 3/5

$G = (N, T, P, S)$, kde $N = \{A, B, C, S\}$, $T = \{a, b, c\}$,
$P = \{S \rightarrow AC, C \rightarrow SB, A \rightarrow a, B \rightarrow b, S \rightarrow c\}$

Otázka: $aacbb \in L(G)$**?**

$S \rightarrow AC$

$S[1, 3]=\emptyset$ $S[2, 4]=\{S\}$ $S[3, 5]= \emptyset$

$S[1, 2]=\emptyset$ $S[2, 3]=\emptyset$ $S[3, 4]=\{C\}$ $S[4, 5]=\emptyset$

$S[1, 1]=\{A\}$ $S[2, 2]=\{A\}$ $S[3, 3]=\{S\}$ $S[4, 4]=\{B\}$ $S[5, 5]=\{B\}$

a	*a*	*c*	*b*	*b*

Obecná SA založená na CNF: Příklad 4/5

$G = (N, T, P, S)$, kde $N = \{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{S}\}$, $T = \{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}\}$,

$P = \{\boldsymbol{S} \rightarrow \boldsymbol{AC}, \boldsymbol{C} \rightarrow \boldsymbol{SB}, \boldsymbol{A} \rightarrow \boldsymbol{a}, \boldsymbol{B} \rightarrow \boldsymbol{b}, \boldsymbol{S} \rightarrow \boldsymbol{c}\}$

Otázka: $\boldsymbol{aacbb} \in L(G)$**?**

$\boldsymbol{C \rightarrow SB}$

$\boldsymbol{S}$**[1, 4]=**$\emptyset$ $\boldsymbol{S}$**[2, 5]=**$\{\boldsymbol{C}\}$

$\boldsymbol{S}$**[1, 3]=**$\emptyset$ $\boldsymbol{S}$**[2, 4]=**$\{\boldsymbol{S}\}$ $\boldsymbol{S}$**[3, 5]=** $\emptyset$

$\boldsymbol{S}$**[1, 2]=**$\emptyset$ $\boldsymbol{S}$**[2, 3]=**$\emptyset$ $\boldsymbol{S}$**[3, 4]=**$\{\boldsymbol{C}\}$ $\boldsymbol{S}$**[4, 5]=**$\emptyset$

$\boldsymbol{S}$**[1, 1]=**$\{\boldsymbol{A}\}$ $\boldsymbol{S}$**[2, 2]=**$\{\boldsymbol{A}\}$ $\boldsymbol{S}$**[3, 3]=**$\{\boldsymbol{S}\}$ $\boldsymbol{S}$**[4, 4]=**$\{\boldsymbol{B}\}$ $\boldsymbol{S}$**[5, 5]=**$\{\boldsymbol{B}\}$

a	***a***	***c***	***b***	***b***

Obecná SA založená na CNF: Příklad 5/5

$G = (N, T, P, S)$, kde $N = \{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{S}\}$, $T = \{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}\}$,
$P = \{\boldsymbol{S} \rightarrow \boldsymbol{AC}, \boldsymbol{C} \rightarrow \boldsymbol{SB}, \boldsymbol{A} \rightarrow \boldsymbol{a}, \boldsymbol{B} \rightarrow \boldsymbol{b}, \boldsymbol{S} \rightarrow \boldsymbol{c}\}$

Otázka: $\boldsymbol{aacbb} \in L(G)$?

$S \rightarrow AC$

$S[1, 5] = \{S\}$ $\quad$ $S \in S[1, 5]$ ⇒ **ANO**

$S[1, 4] = \emptyset$ $\quad$ $S[2, 5] = \{C\}$

$S[1, 3] = \emptyset$ $\quad$ $S[2, 4] = \{S\}$ $\quad$ $S[3, 5] = \emptyset$

$S[1, 2] = \emptyset$ $\quad$ $S[2, 3] = \emptyset$ $\quad$ $S[3, 4] = \{C\}$ $\quad$ $S[4, 5] = \emptyset$

$S[1, 1] = \{A\}$ $\quad$ $S[2, 2] = \{A\}$ $\quad$ $S[3, 3] = \{S\}$ $\quad$ $S[4, 4] = \{B\}$ $\quad$ $S[5, 5] = \{B\}$

a	a	c	b	b

Pumping lemma pro BKJ

• Nechť L je BKJ. Potom existuje $k \geq 1$ takové, že:
pokud $z \in L$ a $|z| \geq k$, pak existuje u, v, w, x, y tak, že $z = uvwxy$, přičemž dále platí:
1) $vx \neq \varepsilon$ **2)** $|vwx| \leq k$ **3)** pro každé $m \geq 0$: $uv^m wx^m y \in L$

Příklad:

$G = (\{S, A\}, \{a, b, c\}, \{S \rightarrow aAa, A \rightarrow bAb, A \rightarrow c\}, S)$
generuje $L(G) = \{ab^n cb^n a : n \geq 0\}$, tedy $L(G)$ je BKJ.

Existuje $k = 5$ takové, že **1)**, **2)** and **3)** platí:

• pro $z = abcba$: $z \in L(G)$ a $|z| \geq 5$:

$$\begin{matrix} a & b & c & b & a \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ u & v & w & x & y \end{matrix}$$

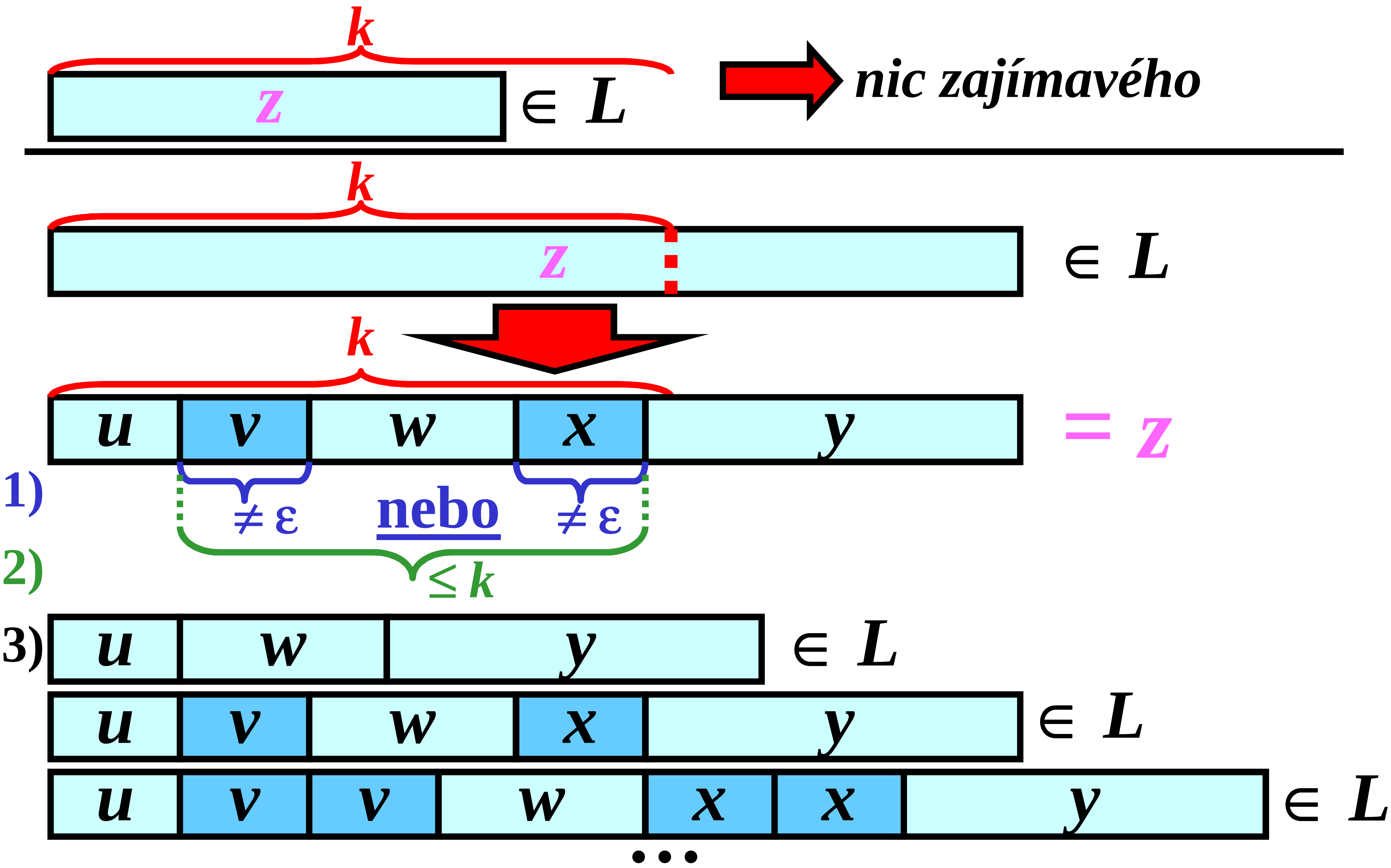
$vx = bb \neq \varepsilon$
$|vwx| = 3$: $1 \leq 3 \leq 5$

$uv^0wx^0y = ab^0cb^0a = aca \in L(G)$
$uv^1wx^1y = ab^1cb^1a = abcba \in L(G)$
$uv^2wx^2y = ab^2cb^2a = abbcbba \in L(G)$
$\vdots$

• pro $z = abbcbba$: $z \in L(G)$ a $|z| \geq 5$:
$\vdots$

Pumping lemma: Ilustrace

- L = libovolný bezkontextový jazyk:



Pumping lemma: Aplikace

- Pomocí pumping lemmy pro BKJ často provádíme důkaz sporem, že daný jazyk **<u>není</u>** bezkontextový:

Předpokládejme, že L je bezkontextový

Uvažujme PL konstantu $\boldsymbol{k}$ a vyberme $\boldsymbol{z} \in L$, jehož délka je závislá na $\boldsymbol{k}$ tak, že $|\boldsymbol{z}| \geq \boldsymbol{k}$ je vždy pravdivé

Pro <u>všechny</u> dekompozice $\boldsymbol{z}$ na $\boldsymbol{uvwxy}$: $vx \neq \varepsilon$, $|vwx| \leq k$, ukážeme:

existuje $m \geq 0$ pro které $\boldsymbol{uv^m wx^m y} \notin \boldsymbol{L}$;

ale podle PL platí vztah: $\boldsymbol{uv^m wx^m y} \in \boldsymbol{L}$

SPOR

špatný předpoklad

Proto

***L* není bezkontextový**

Pumping lemma: Příklad 1/2

Dokažme, že $L = \{a^n b^n c^n : n \geq 1\}$ není BKJ.

1) Předpokládejme, že L je BKJ. Nechť $k \geq 1$ je konstanta z pumping lemmy pro daný jazyk L.

2) Nechť $z = a^k b^k c^k$: $a^k b^k c^k \in L$, $|z| = |a^k b^k c^k| = 3k \geq k$

3) Všechny dekompozice z na $uvwxy$; $vx \neq \varepsilon$, $|vwx| \leq k$:

$$\overbrace{aaaaa\ldots aa}^{k}\overbrace{bb\ldots bb\ldots bb}^{k}\overbrace{cc\ldots ccccc}^{k}$$

a) $vwx \in \{a\}^*\{b\}^*$, $vx \neq \varepsilon$

b) $vwx \in \{b\}^*\{c\}^*$, $vx \neq \varepsilon$

Pumping lemma: Příklad 2/2

a) $vwx \in \{a\}^*\{b\}^*$:

- Pumping lemma: $uv^0wx^0y \in L$

$$\underbrace{a\ldots a}_{u}\overbrace{\underbrace{a\ldots aabb\ldots b}_{vwx}}^{\leq k}\underbrace{b\ldots bcc\ldots cc}_{y}$$

- $uv^0wx^0y = uwy = \underbrace{a\ldots a}_{u}\underbrace{\ldots aabb\ldots}_{w}\underbrace{b\ldots bcc\ldots cc}_{y} \notin L$

Pozn.: *uwy* obsahuje „k“ symbolů c, ale méně než „k“ symbolů a nebo b

b) $vwx \in \{b\}^*\{c\}^*$:

- Pumping lemma: $uv^0wx^0y \in L$

$$\underbrace{aa\ldots ab\ldots b}_{u}\overbrace{\underbrace{b\ldots bbcc\ldots c}_{vwx}}^{\leq k}\underbrace{c\ldots c}_{y}$$

- $uv^0wx^0y = uwy = \underbrace{aa\ldots ab\ldots b}_{u}\underbrace{\ldots bbcc\ldots}_{w}\underbrace{c\ldots c}_{y} \notin L$

Pozn.: *uwy* obsahuje „k“ symbolů a, ale méně než „k“ symbolů b nebo c

Všechny dekompozice vedou ke sporu!

4) Proto L není bezkontextový jazyk.

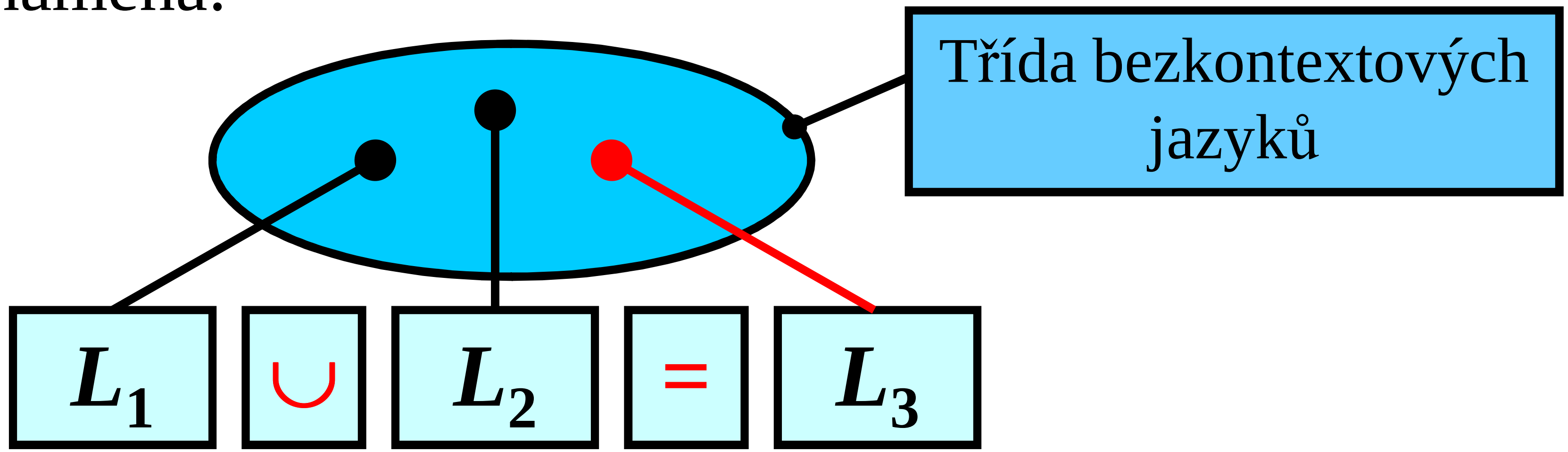
Uzávěrové vlastnosti BKJ

Definice: Třída bezkontextových jazyků je uzavřená vůči operaci ***o***, pokud výsledek operace ***o*** na libovolné bezkontextové jazyky je opět bezkontextový jazyk.

Ilustrace:

• Třída bezkontextových jazyků je uzavřená vůči ***sjednocení***.

To znamená:



Algoritmus: BKG pro sjednocení

- **Vstup:** $G_1 = (N_1, T, P_1, S_1)$ a $G_2 = (N_2, T, P_2, S_2)$;
- **Výstup:** Gramatika $G_u = (N, T, P, S)$ taková, že:

$$L(G_u) = L(G_1) \cup L(G_2)$$

- **Metoda:**
- Nechť $S \notin N_1 \cup N_2$, dále nechť $N_1 \cap N_2 = \emptyset$:
  - $N := \{S\} \cup N_1 \cup N_2$;
  - $P := \{S \rightarrow S_1, S \rightarrow S_2\} \cup P_1 \cup P_2$;

Algoritmus: BKG pro konkatenaci

- **Vstup:** $G_1 = (N_1, T, P_1, S_1)$ a $G_2 = (N_2, T, P_2, S_2)$;
- **Výstup:** $G_c = (N, T, P, S)$ taková, že:

$$L(G_c) = L(G_1) \,.\, L(G_2)$$

- **Metoda:**
- Nechť $S \notin N_1 \cup N_2$, dále nechť $N_1 \cap N_2 = \emptyset$:
  - $N := \{S\} \cup N_1 \cup N_2$;
  - $P := \{S \rightarrow S_1S_2\} \cup P_1 \cup P_2$;

Algoritmus: BKG pro iteraci

- **Vstup:** $G_1 = (N_1, T, P_1, S_1)$
- **Výstup:** $G_i = (N, T, P, S)$ taková, že: $L(G_i) = L(G_1)^*$

- **Metoda:**
- Nechť $S \notin N_1$:
  - $N := \{S\} \cup N_1$;
  - $P := \{S \to S_1S, S \to \varepsilon\} \cup P_1$;

Uzávěrové vlastnosti

Tvrzení: Třída BKJ je uzavřená vůči:
sjednocení, **konkatenaci**, **iteraci**.

Důkaz:

- Nechť L_1, L_2 jsou dva **bezkontextové jazyky**.
- Potom existují dvě BKG G_1, G_2, pro které platí: $L(G_1) = L_1$, $L(G_2) = L_2$;
- Sestrojme gramatiky pomocí předchozích algoritmů:
  - G_u, pro kterou platí: $L(G_u) = L(G_1) \cup L(G_2)$
  - G_c, pro kterou platí: $L(G_c) = L(G_1) \,.\, L(G_2)$
  - G_i, pro kterou platí: $L(G_i) = L(G_1)^*$
- Každá BKG definuje bezkontextový jazyk, tedy: $L_1 L_2$, $L_1 \cup L_2$, L_1^* jsou **bezkontextové jazyky**.

Průnik: Není uzavřeno

Tvrzení: Třída bezkontextových jazyků **není** uzavřená vůči **průniku**.

Důkaz:

- Průnik nějakých dvou BKJ nesmí být tedy BKJ:

- $L_1 = \{a^m b^n c^n : m, n \geq 1\}$ je BKJ
- $L_2 = \{a^n b^n c^m : m, n \geq 1\}$ je BKJ
- $L_1 \cap L_2 = \{a^n b^n c^n : n \geq 1\}$ není BKJ
  (Důkaz je založen na pumping lemma viz dříve)

CBD

Doplněk: Není uzavřeno

Tvrzení: Třída bezkontextových jazyků **není** uzavřená vůči **doplňku**.

Důkaz sporem:

- Předpokládejme, že třída bezkontextových jazyků je uzavřená vůči doplňku:
- $L_1 = \{a^m b^n c^n : m, n \geq 1\}$ je **BKJ**
- $L_2 = \{a^n b^n c^m : m, n \geq 1\}$ je **BKJ**
- $\overline{L_1}, \overline{L_2}$ jsou tedy **BKJ**
- $\overline{L_1} \cup \overline{L_2}$ je **BKJ** (třída BKJ je uzavřená vůči sjednocení)
- $\overline{\overline{L_1} \cup \overline{L_2}}$ je **BKJ** (předpoklad)
- De-Morganovy zákony říkají: $L_1 \cap L_2 = \{a^n b^n c^n : n \geq 1\}$ je **BKJ**
- $\{a^n b^n c^n : n \geq 1\}$ ale není **BKJ** $\Rightarrow$ **Spor**

Hlavní rozhodnutelné problémy

1. Problém členství:

- **Instance:** BKG G, $w \in T^*$; **Otázka:** $w \in L(G)$**?**

2. Problém prázdnosti:

- **Instance:** BKG G; **Otázka:** $L(G) = \emptyset$**?**

3. Problém konečnosti:

- **Instance:** BKG G; **Otázka:** Je $L(G)$ konečný**?**

Algoritmus: Problém členství

- **Vstup:** BKG $G = (N, T, P, S)$ v CNF, $w \in T^+$
- **Výstup:** **ANO**, pokud $w \in L(G)$

  NE, pokud $w \notin L(G)$

- **Metoda I:**
- **<u>if</u>** $S \Rightarrow^{\boldsymbol{n}} w$, kde $1 \le \boldsymbol{n} \le 2|w| - 1$, **<u>then</u>** napiš('**ANO**')

  <u>else</u> napiš('**NE**')
- **Metoda II:**
- Viz Obecná metoda SA založená na CNF

Celkově:

Problém členství je pro BKJ rozhodnutelný

Dostupné symboly

Myšlenka: Symbol X je *dostupný*, pokud $S \Rightarrow^* \ldots X \ldots$, kde S je počáteční neterminál.

Definice: Nechť $G = (N, T, P, S)$ je BKG. Symbol $X \in N \cup T$ je *dostupný*, pokud existuje $u, v \in (N \cup T)^*$, takové, že: $S \Rightarrow^* uXv$. Jinak je X *nedostupný*.

Pozn.: Každý nedostupný symbol může být odstraněn z BKG

Příklad:

$G = (\{S, A, B\}, \{a, b\}, \{S \rightarrow SB, S \rightarrow a, A \rightarrow ab, B \rightarrow aB\}, S)$

S - dostupný: pro $u = \varepsilon, v = \varepsilon$: $S \Rightarrow^0 S$

A - **nedostupný**: neexistuje $u, v \in \Sigma^*$ takové, že: $S \Rightarrow^* uAv$

B - dostupný: pro $u = S, v = \varepsilon$: $S \Rightarrow^1 SB$

a - dostupný: pro $u = \varepsilon, v = \varepsilon$: $S \Rightarrow^1 a$

b - **nedostupný**: neexistuje $u, v \in \Sigma^*$ takové, že: $S \Rightarrow^* ubv$

Ukončující symboly

Myšlenka: Symbol *X* je *ukončující*, pokud *X derivuje řetězec terminálů.*

Definice: Nechť $G = (N, T, P, S)$ je BKG. Symbol $X \in N \cup T$ je *ukončující*, pokud existuje řetězec $w \in T^*$, pro který platí: $X \Rightarrow^* w$. Jinak je *X neukončující*.

Pozn.: Každý neukončující symbol může být odstraněn z BKG

Příklad:

$G = (\{S, A, B\}, \{a, b\}, \{S \to SB, S \to a, A \to ab, B \to aB\}, S)$

Symbol S - ukončující: pro $w = a$: $S \Rightarrow^1 a$

Symbol A - ukončující: pro $w = ab$: $A \Rightarrow^1 ab$

Symbol B - **neukončující**: neexistuje $w \in T^*$ takové, že: $B \Rightarrow^* w$

Symbol a - ukončující: pro $w = a$: $a \Rightarrow^0 a$

Symbol b - ukončující: pro $w = b$: $b \Rightarrow^0 b$

Algoritmus: Problém prázdnosti

- **Vstup:** BKG $G = (N, T, P, S)$;
- **Výstup:** **ANO**, pokud $L(G) = \emptyset$
  NE, pokud $L(G) \neq \emptyset$

- **Metoda:**
- **if** S je neukončující **then** napiš('**ANO**')
  else napiš('**NE**')

Celkově:

Problém prázdnosti je pro BKJ rozhodnutelný

Algoritmus: Problém konečnosti

- **Vstup:** BKG $G = (N, T, P, S)$ v CNF;
- **Výstup:** **ANO**, pokud $L(G)$ je konečný
  NE, pokud $L(G)$ je nekonečný

- **Metoda:**
- Nechť $k = 2^{\text{card}(N)}$
- **if** existuje $z \in L(G)$, $k \leq |z| < 2k$ **then** napiš('**NE**')
  else napiš('**ANO**')

Celkově:

Problém konečnosti je pro BKJ rozhodnutelný

Hlavní nerozhodnutelné problémy

1. Problém ekvivalence:

- **Instance:** BKG G_1, G_2; **Otázka:** $L(G_1) = L(G_2)$**?**

2. Problém jednoznačnosti:

- **Instance:** BKG G; **Otázka:** Je G jednoznačná**?**

Poznámka:

Je matematicky dokázáno, že neexistují žádné algoritmy, které by tyto problémy vyřešily v konečném čase.