

Kapitola XII.

Normální formy a vlastnosti bezkontextových jazyků

Chomského normální forma (CNF)

Definice: Necht' $G = (N, T, P, S)$ 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: Necht' $G = (N, T, P, S)$ je BKG.
 G je v *Greibachové normální formě*, pokud
 každé pravidlo z P má následující tvar:

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

Příklad:

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

$P = \{S \rightarrow aSB, S \rightarrow aB, B \rightarrow b\}$

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 Greibachově normální formě.

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

Pozn.: Základní vlastnosti CNF a GNF:

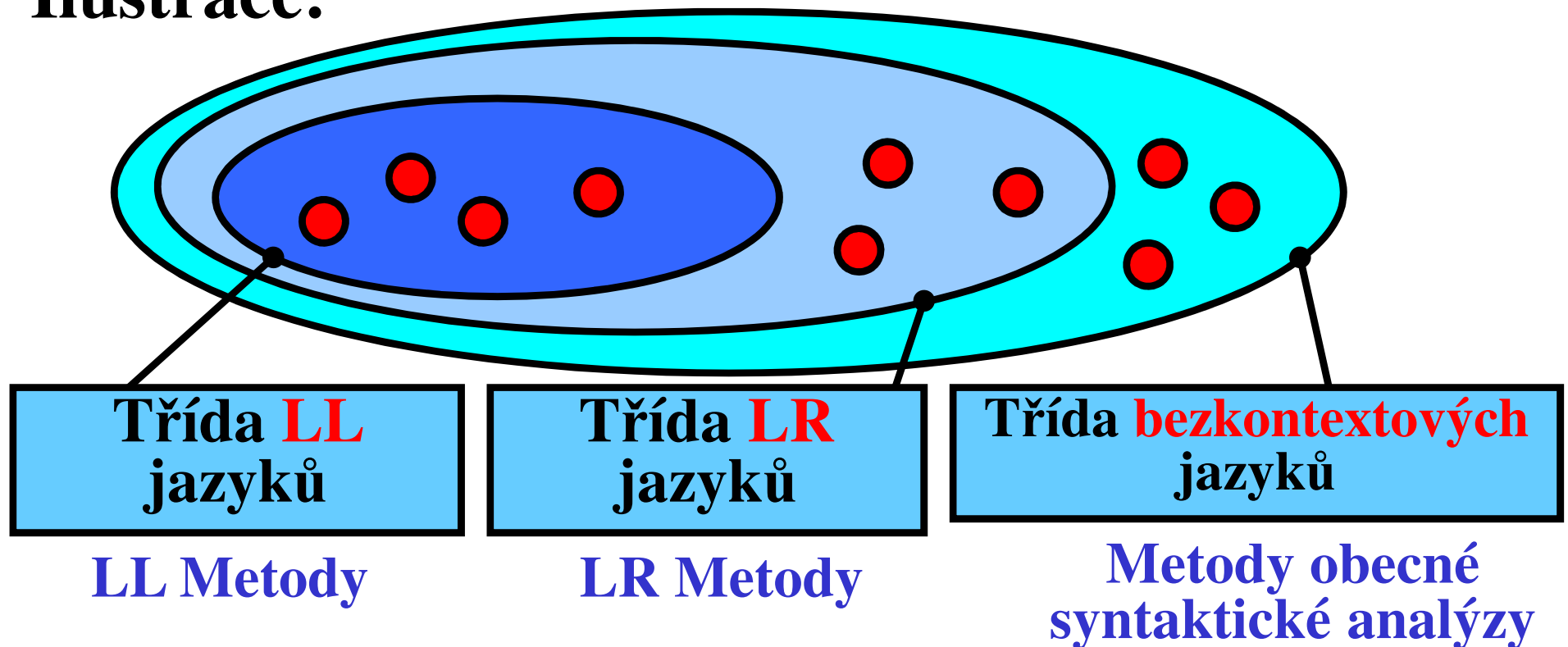
CNF: pokud $S \Rightarrow^n w$; $w \in T^*$ potom $n = 2|w| - 1$

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:

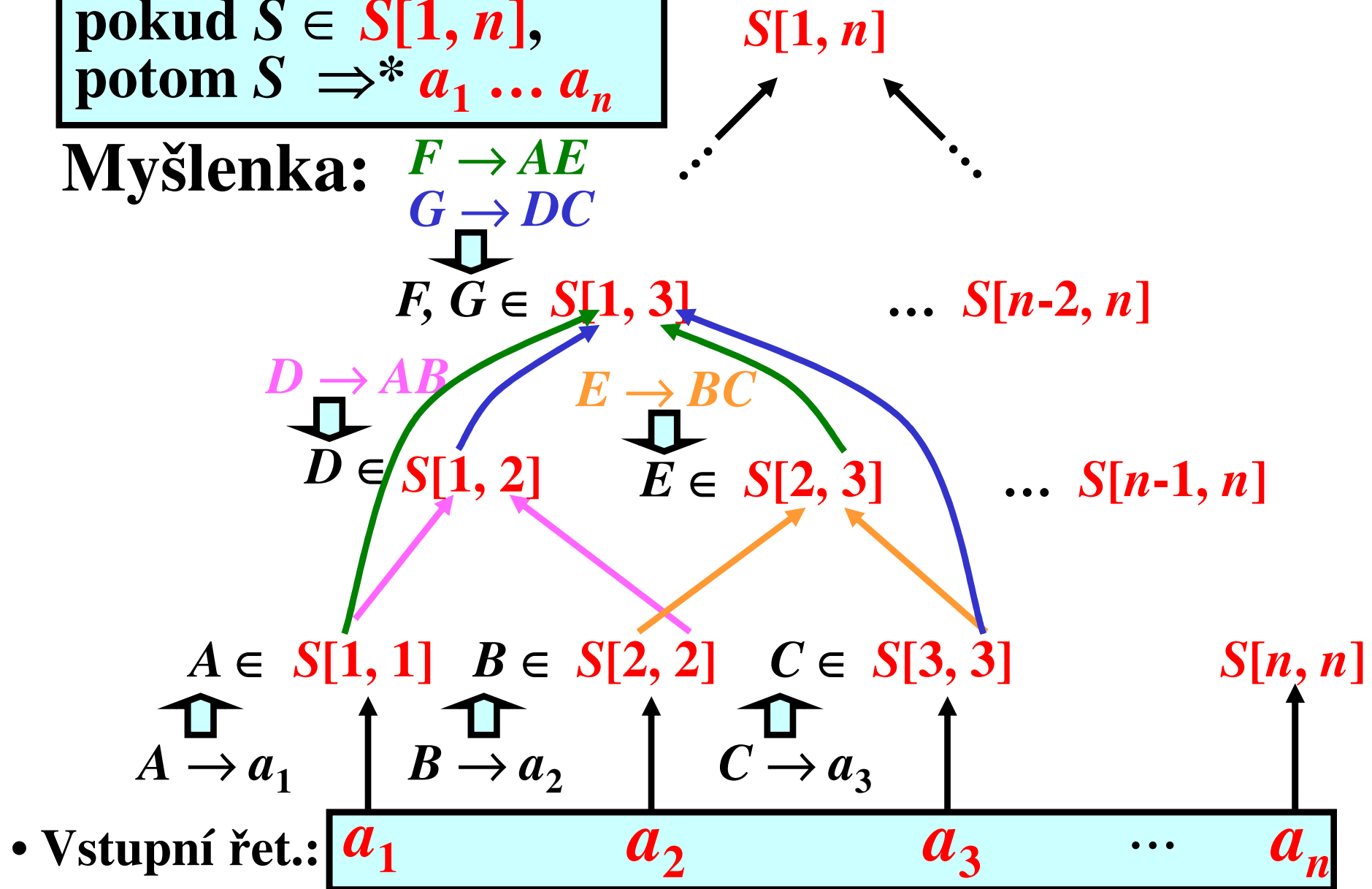


- **Pozn.:** Třída **LR** jazyků =
třída **deterministických jazyků**

Obyčná SA založená na CNF

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

Myšlenka: $F \rightarrow AE$
 $G \rightarrow DC$



Algoritmus: Obecná SA založená na CNF

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

- **Metoda:**

- pro každé a_i , kde $i = 1, \dots, n$:

$$S[i, i] := \{A : A \rightarrow a_i \in P\}$$

- Aplikuj následující pravidlo, dokud žádná z množin $S[i, k]$ nemůže být změněna:

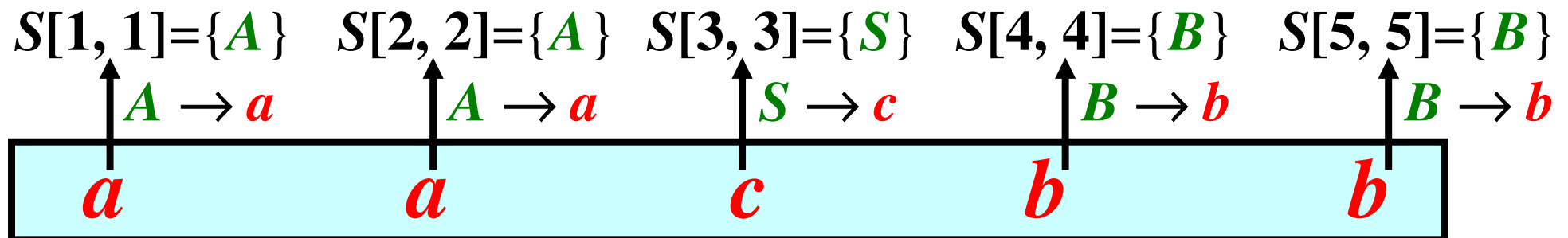
if $A \rightarrow BC \in P$, $B \in S[i, j]$, $C \in S[j+1, k]$,
 kde $1 \leq i \leq j < k \leq n$ then přidej A do $S[i, k]$

- if $S \in S[1, n]$ then napiš('ANO')
else napiš('NE')

Obecná SA založená na CNF: Příklad 1/5

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

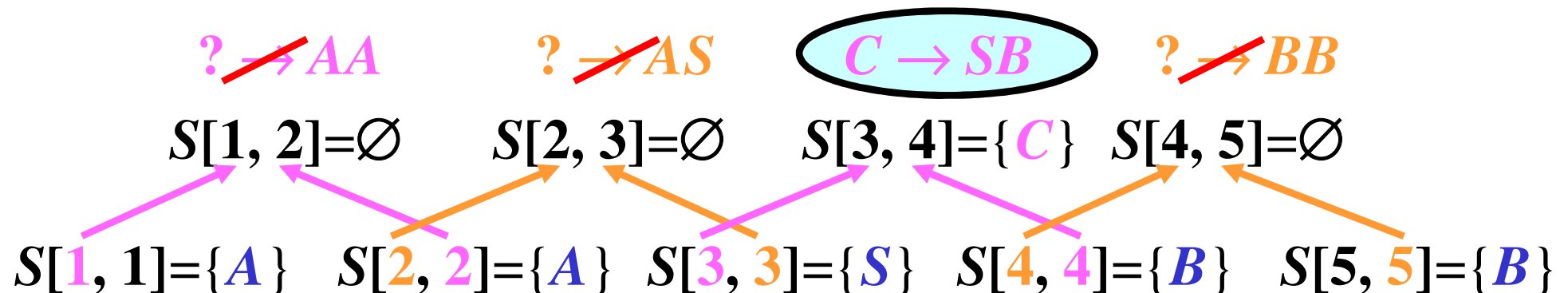
Otázka: $aacbb \in L(G)$?



Obecná SA založená na CNF: Příklad 2/5

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

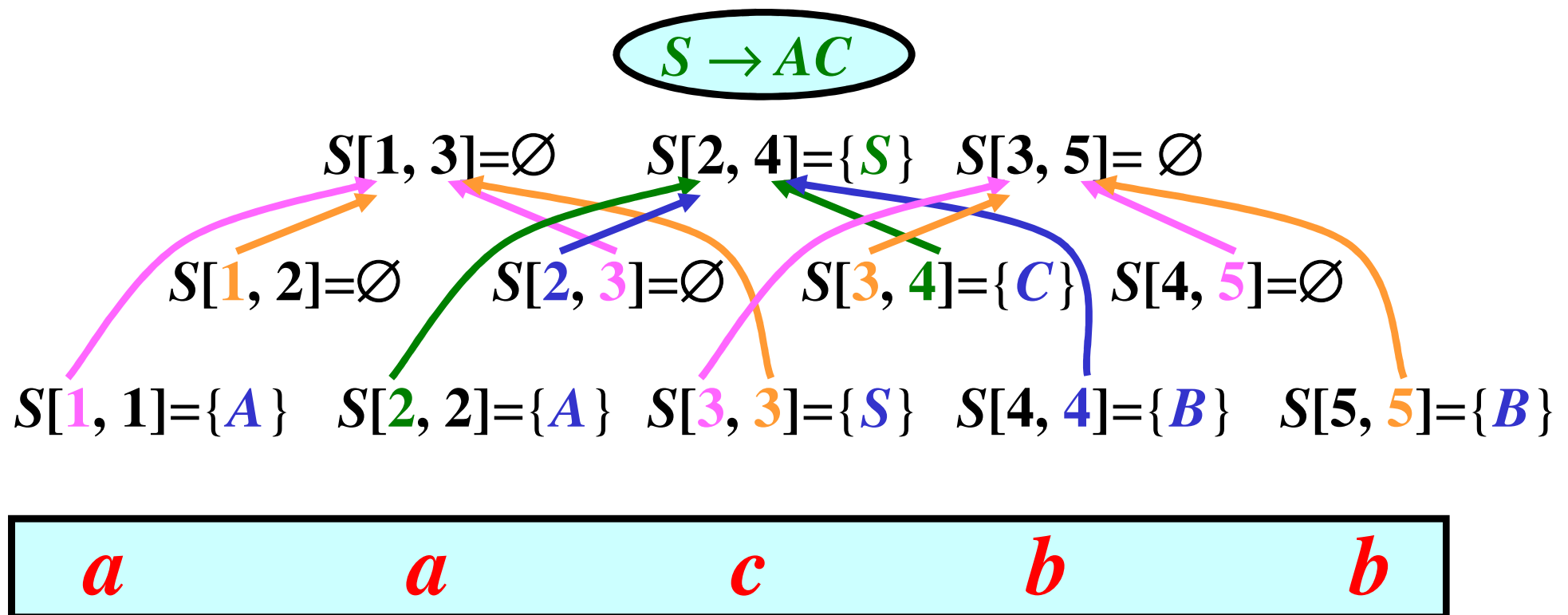
Otázka: $aacbb \in L(G)$?



Obecná SA založená na CNF: Příklad 3/5

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

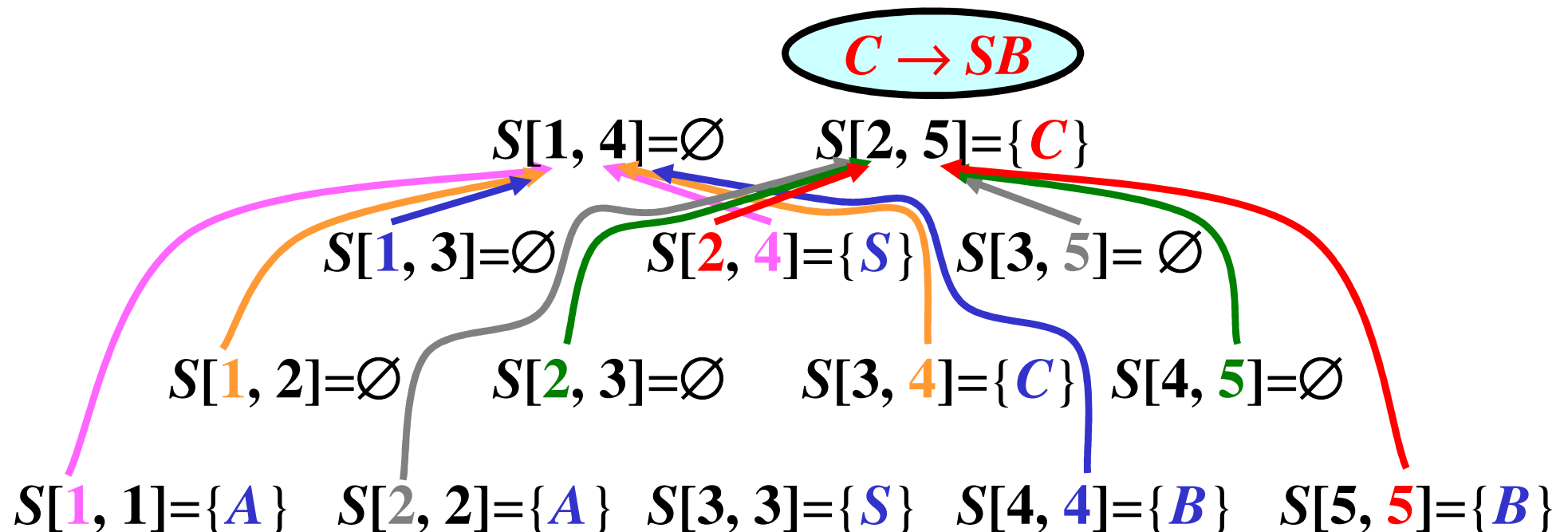
Otázka: $aacbb \in L(G)$?



Obecná SA založená na CNF: Příklad 4/5

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

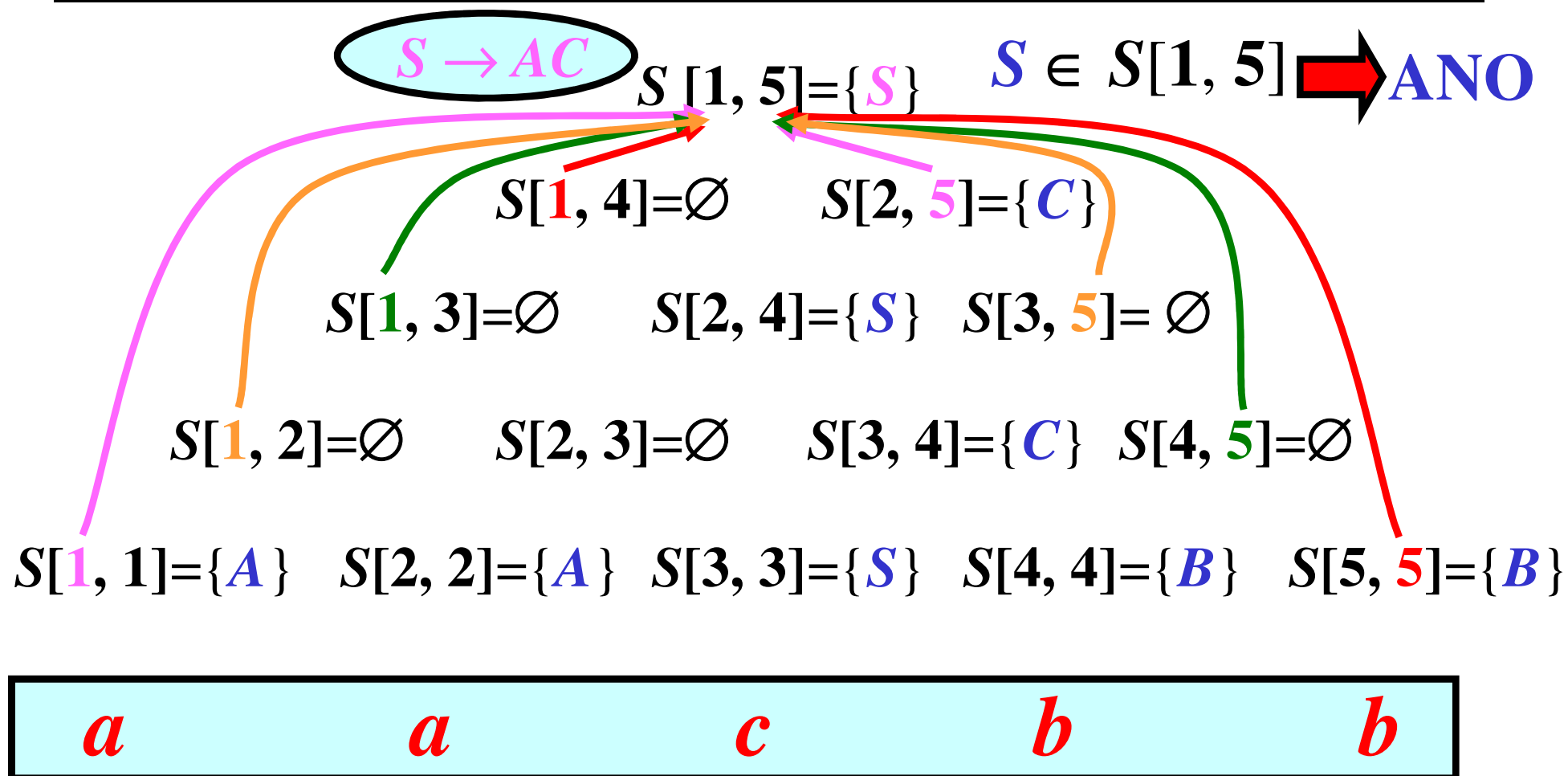
Otázka: $aacbb \in L(G)$?

 a a c b b

Obecná SA založená na CNF: Příklad 5/5

$G = (N, T, P, S)$, 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: $aacbb \in L(G)$?



Pumping lemma pro BKJ

- Necht' 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 $m \geq 0$: $uv^mwx^my \in L$

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 $z = \textcolor{brown}{a}\textcolor{blue}{b}\textcolor{green}{c}\textcolor{blue}{b}\textcolor{brown}{a}$: $z \in L(G)$ a $|z| \geq 5$:

$$\begin{array}{cccccc} \textcolor{brown}{a} & \textcolor{blue}{b} & \textcolor{green}{c} & \textcolor{blue}{b} & \textcolor{brown}{a} \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ u & v & w & x & y \end{array}$$

$$uv^0wx^0y = \textcolor{brown}{a}\textcolor{blue}{b}^0\textcolor{green}{c}\textcolor{blue}{b}^0\textcolor{brown}{a} = \textcolor{brown}{a}\textcolor{green}{c}\textcolor{brown}{a} \in L(G)$$

$$uv^1wx^1y = \textcolor{brown}{a}\textcolor{blue}{b}^1\textcolor{green}{c}\textcolor{blue}{b}^1\textcolor{brown}{a} = \textcolor{brown}{a}\textcolor{blue}{b}\textcolor{green}{c}\textcolor{blue}{b}\textcolor{brown}{a} \in L(G)$$

$$uv^2wx^2y = \textcolor{brown}{a}\textcolor{blue}{b}^2\textcolor{green}{c}\textcolor{blue}{b}^2\textcolor{brown}{a} = \textcolor{brown}{a}\textcolor{blue}{b}\textcolor{blue}{b}\textcolor{green}{c}\textcolor{blue}{b}\textcolor{blue}{b}\textcolor{brown}{a} \in L(G)$$

$$\vdots$$

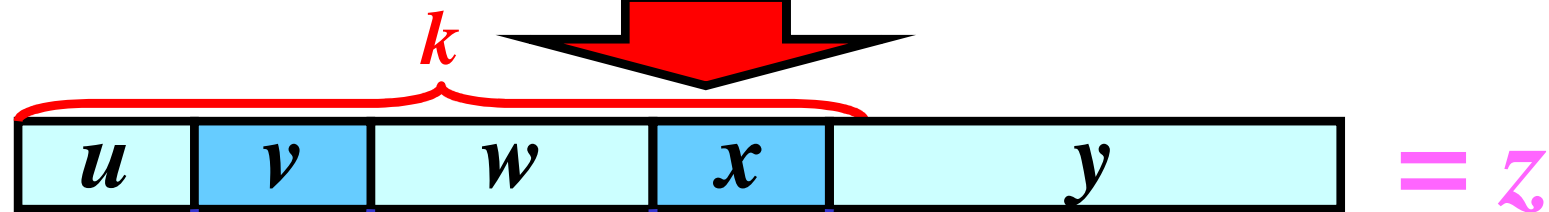
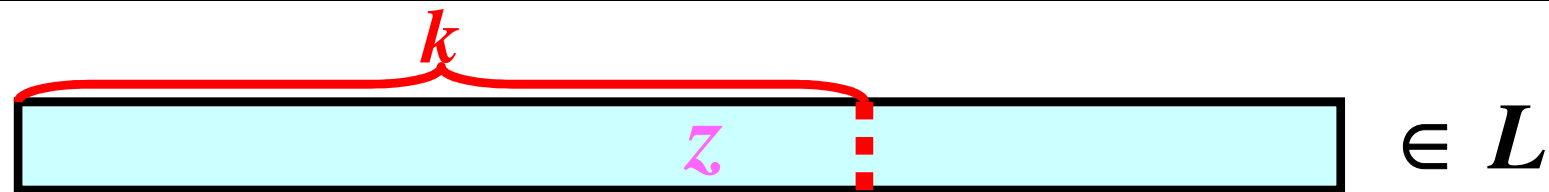
$$vx = \textcolor{blue}{b}\textcolor{blue}{b} \neq \varepsilon$$

$$|vwx| = \textcolor{violet}{3}: 1 \leq \textcolor{violet}{3} \leq \textcolor{red}{5}$$
- pro $z = \textcolor{brown}{a}\textcolor{blue}{b}\textcolor{blue}{b}\textcolor{green}{c}\textcolor{blue}{b}\textcolor{blue}{b}\textcolor{brown}{a}$: $z \in L(G)$ a $|z| \geq 5$:

$$\vdots$$

Pumping lemma: Ilustrace

- L = libovolný bezkontextový jazyk:



- 1) $v \neq \varepsilon$ nebo $x \neq \varepsilon$
- 2) $|vwx| \leq k$



...

Pumping lemma: Aplikace

- Pomocí pumping lemma pro BKJ často provádíme důkaz sporem, že daný jazyk **není** bezkontextový:

Předpokládejme, že L je bezkontextový

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

Pro všechny dekompozice z na $uvwxy$: $vx \neq \varepsilon$, $|vwx| \leq k$, ukážeme existuje $m \geq 0$ pro které $uv^mwx^my \notin L$; } **SPOR**
ale podle PL platí vztah: $uv^mwx^my \in L$

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. Necht' $k \geq 1$ je konstanta z pumping lemma pro daný jazyk L .
- 2) Necht' $z = a^k b^k c^k$: $a^k b^k c^k \in L$, $|z| = |a^k b^k c^k| = 3k \geq k$
- 3) Všechny dekompozice z na $uvwxy$; $vx \neq \varepsilon$, $|vwx| \leq k$:

$\overbrace{aaaaa \dots a}^k \overbrace{bb \dots bb}^k \overbrace{cc \dots cc}^k$
 $aaaaa \dots aabbb \dots bbbcc \dots ccccc$

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

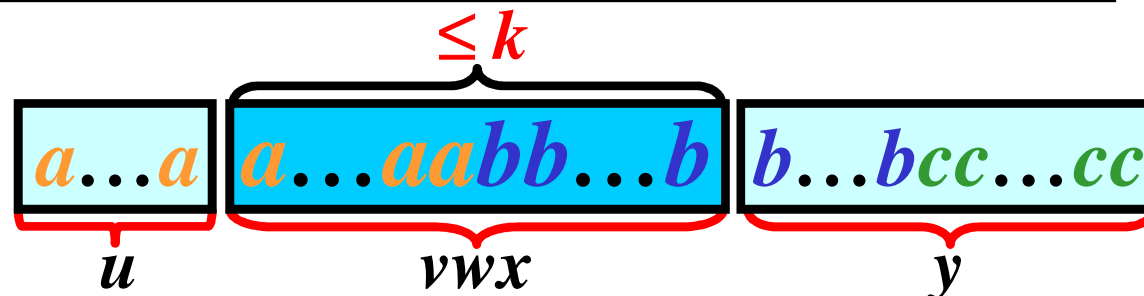
b) $vwx \in \{b\}^* \{c\}^*$,
 $vx \neq \varepsilon$

Pumping lemma: Příklad 2/2

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

• Pumping lemma:

$$uv^0wx^0y \in L$$



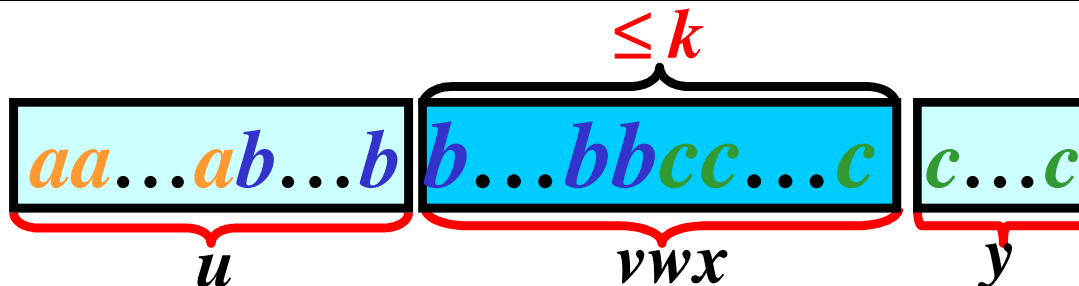
• $uv^0wx^0y = uwy =$ $\notin L$

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

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

• Pumping lemma:

$$uv^0wx^0y \in L$$



• $uv^0wx^0y = uwy =$ $\notin L$

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.

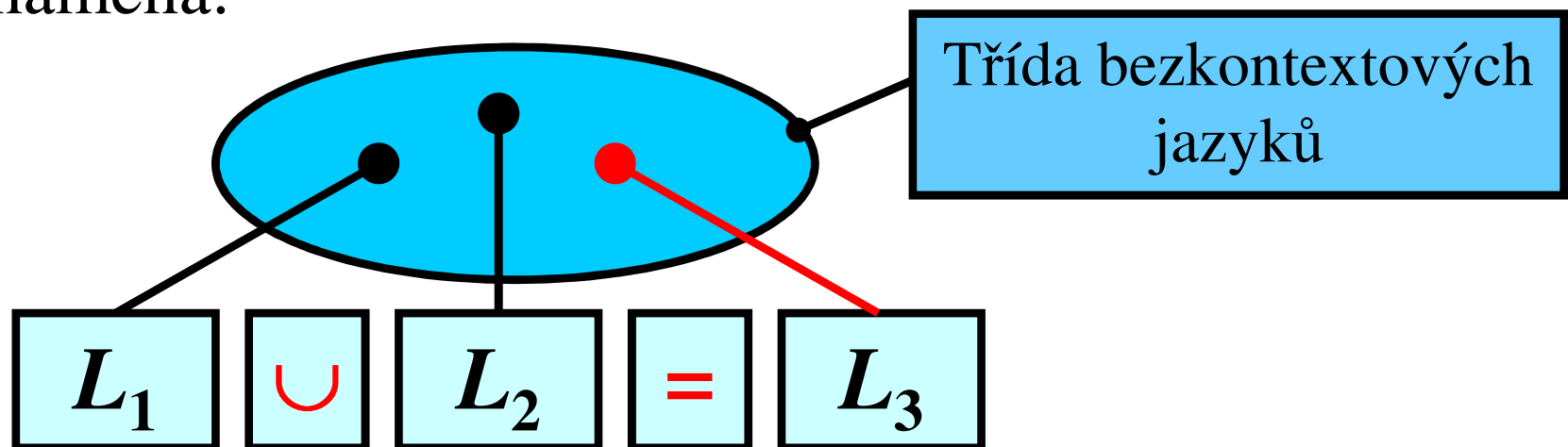
Uzávěrové vlastnosti BKJ

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

Ilustrace:

- Třída bezkontextových jazyků je uzavřená vůči *sjednocení*.

To znamená:



Algoritmus: BKG pro sjednocení

- **Vstup:** $G_1 = (N_1, T_1, P_1, S_1)$ a $G_2 = (N_2, T_2, P_2, S_2)$;
 - **Výstup:** Gramatika $G_u = (N, T, P, S)$ taková, že:
$$L(G_u) = L(G_1) \cup L(G_2)$$
-

- **Metoda:**

- Necht' $S \notin N_1 \cup N_2$, dále necht' $N_1 \cap N_2 = \emptyset$:
 - $T := T_1 \cup T_2$;
 - $N := \{S\} \cup N_1 \cup N_2$;
 - $P := \{S \rightarrow S_1, S \rightarrow S_2\} \cup P_1 \cup P_2$;

Algoritmus: BKG pro konkatenci

- **Vstup:** $G_1 = (N_1, T_1, P_1, S_1)$ a $G_2 = (N_2, T_2, P_2, S_2)$;
 - **Výstup:** $G_c = (N, T, P, S)$ taková, že:
$$L(G_c) = L(G_1) \cdot L(G_2)$$
-

- **Metoda:**

- Necht' $S \notin N_1 \cup N_2$, dále necht' $N_1 \cap N_2 = \emptyset$:
 - $T := T_1 \cup T_2$;
 - $N := \{S\} \cup N_1 \cup N_2$;
 - $P := \{S \rightarrow S_1 S_2\} \cup P_1 \cup P_2$;

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:**
 - Necht' $S \notin N_1$:
 - $N := \{S\} \cup N_1$;
 - $P := \{S \rightarrow S_1 S, S \rightarrow \varepsilon\} \cup P_1$;

Uzávěrové vlastnosti

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

Důkaz:

- Necht' 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: $L(G_c) = L(G_1) \cdot L(G_2)$
 - 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

Doplňěk: 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:

- if $S \Rightarrow^n w$, kde $1 \leq n \leq 2|w| - 1$, then napiš('ANO')
else 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 $S \Rightarrow^* \dots X \dots$, kde S je počáteční neterminál.

Definice: Necht' $G = (N, T, P, S)$ 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: Necht' $G = (N, T, P, S)$ 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:

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

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:

- Necht' $k = 2^{\text{card}(N)}$
 - 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.