3.4 Uzávěrové vlastnosti bezkontextových jazyků

3.4.1 Tvrzení. Bezkontextové jazyky jsou uzavřeny na sjednocení.

To znamená, jsou-li L_1 a L_2 dva bezkontextové jazyky, pak také jazyk $L_1 \cup L_2$ je bezkontextový. □

Zdůvodnění: Jazyky L_1, L_2 jsou bezkontextové, proto existují bezkontextové gramatiky $\mathcal{G}_1 = (N_1, \Sigma, S_1, P_1)$ a $\mathcal{G}_2 = (N_2, \Sigma, S_2, P_2)$ takové, že $\mathcal{G}_1$ generuje L_1 a $\mathcal{G}_2$ generuje L_2. Přejmenujeme neterminály gramatiky $\mathcal{G}_2$ tak, aby gramatiky neměly žádný společný neterminál (tj. $N_1 \cap N_2 = \emptyset$).

Gramatiku $\mathcal{G} = (N, \Sigma, S, P)$ generující jazyk $L_1 \cup L_2$ vytvoříme takto:

a) $N = N_1 \cup N_2 \cup \{S\}$, kde S je nový startovací symbol ($S \notin N_1 \cup N_2$).

b) $P = P_1 \cup P_2 \cup \{S \to S_1, S \to S_2\}$, tj. k pravidlům gramatik $\mathcal{G}_1$ a $\mathcal{G}_2$ přidáme dvě nová pravidla $S \to S_1$ a $S \to S_2$.

Není těžké ukázat, že gramatika $\mathcal{G}$ generuje jazyk $L_1 \cup L_2$.

3.4.2 Tvrzení. Bezkontextové jazyky jsou uzavřeny na zřetězení.

To znamená, jsou-li L_1 a L_2 dva bezkontextové jazyky, pak také jazyk $L_1\, L_2$ je bezkontextový. □

Zdůvodnění: Jazyky L_1, L_2 jsou bezkontextové, proto existují bezkontextové gramatiky $\mathcal{G}_1 = (N_1, \Sigma, S_1, P_1)$ a $\mathcal{G}_2 = (N_2, \Sigma, S_2, P_2)$ takové, že $\mathcal{G}_1$ generuje L_1 a $\mathcal{G}_2$ generuje L_2. Přejmenujeme neterminály gramatiky $\mathcal{G}_2$ tak, aby gramatiky neměly žádný společný neterminál (tj. $N_1 \cap N_2 = \emptyset$).

Gramatiku $\mathcal{G} = (N, \Sigma, S, P)$ generující jazyk $L = L_1 L_2$ vytvoříme takto:

a) $N = N_1 \cup N_2 \cup \{S\}$, kde S je nový startovací symbol ($S \notin N_1 \cup N_2$).

b) $P = P_1 \cup P_2 \cup \{S \to S_1 S_2\}$, tj. k pravidlům gramatik $\mathcal{G}_1$ a $\mathcal{G}_2$ přidáme nové pravidlo $S \to S_1 S_2$.

Není těžké ukázat, že gramatika $\mathcal{G}$ generuje jazyk $L_1 L_2$.

3.4.3 Tvrzení. Bezkontextové jazyky jsou uzavřeny na Kleeneho operaci $\star$. □

To znamená, je-li L bezkontextový jazyk, pak také jazyk $L^\star$ je bezkontextový.

Zdůvodnění: Jazyk L je bezkontextový, proto existuje bezkontextová gramatika $\mathcal{G}_1 = (N_1, \Sigma, S_1, P_1)$ taková, že $L = L(\mathcal{G}_1)$. Gramatiku $\mathcal{G} = (N, \Sigma, S, P)$ generující jazyk $L^\star$ vytvoříme takto:

a) $N = N_1 \cup \{S\}$, kde S je nový startovací symbol ($S \notin N_1$).

b) $P = P_1 \cup \{S \to S_1 S, S \to \varepsilon\}$, tj. k pravidlům gramatik $\mathcal{G}_1$ a $\mathcal{G}_2$ přidáme dvě nová pravidla $S \to S_1 S$, $S \to \varepsilon$.

Není těžké ukázat, že gramatika $\mathcal{G}$ generuje jazyk $L^\star$.

3.4.4 Tvrzení. Bezkontextové jazyky **nejsou** uzavřeny na průnik.

To znamená, jsou-li L_1 a L_2 dva bezkontextové jazyky, pak jazyk $L_1 \cap L_2$ nemusí být bezkontextový. □

Zdůvodnění: Uvažujme jazyk $L_1 = \{0^k 1^n 2^n \mid k, n \geq 1\}$ a jazyk $L_2 = \{0^k 1^k 2^n \mid k, n \geq 1\}$. Oba tyto jazyky jsou bezkontextové, ale jejich průnik je jazyk $L = \{0^n 1^n 2^n \mid n \geq 1\}$, o kterém víme, že není bezkontextový (viz **??**).

3.4.5 Tvrzení. Bezkontextové jazyky **nejsou** uzavřeny na doplněk.

To znamená, je-li L bezkontextový jazyk, pak jeho doplněk $\overline{L}$ nemusí být bezkontextový. □

Zdůvodnění: Kdyby třída bezkontextových jazyků byla uzavřena na doplňky, byla by uzavřena i na průniky, protože

$$L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}},$$

což víme, že neplatí.

3.4.6 Tvrzení. Třída bezkontextových jazyků je uzavřena na průniky s regulárními jazyky.

To znamená, je-li L bezkontextový jazyk a R regulární jazyk, pak jazyk $L \cap R$ je bezkontextový. □

Nástin důkazu: Jazyk L je bezkontextový, tudíž existuje zásobníkový automat, označme ho $A = (Q_1, \Sigma, \Gamma, \delta_1, q_1, Z_1, F_1)$, takový, že A přijímá L koncovým stavem.

Jazyk R je regulární, tudíž existuje deterministický konečný automat $M = (Q_2, \Sigma, \delta_2, q_2, F_2)$, který přijímá jazyk R.

Zkonstruujeme zásobníkový automat B který přijímá jazyk $L \cap R$ koncovým stavem. (Všimněte si, že se v podstatě jedná o součinovou konstrukci, kterou známe z konečných automatů.)

Zásobníkový automat je $B = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$, kde

- $Q = Q_1 \times Q_2$,
- $q_0 = (q_1, q_2)$,
- $Z_0 = Z_1$,
- $F = F_1 \times F_2$,
- přechodová funkce δ je definována takto:

$$\delta((q,p), a, X) = \{((r,s), \gamma) \mid (r, \gamma) \in \delta_1(q, a, X),\ s \in \delta_2^\star(p, a)\},$$

  kde $q \in Q_1, p \in Q_2, a \in \Sigma \cup \{\varepsilon\}, X \in \Gamma$.

Jinými slovy: $((r,s),\gamma) \in \delta((q,p), a, Y)$ právě tehdy, když

a) buď $a \in \Sigma$, $(r,\gamma) \in \delta_1(q, a, Y)$ a $s = \delta_2(p, a)$,
b) nebo $a = \varepsilon$, $(r,\gamma) \in \delta_1(q, \varepsilon, Y)$ a $s = p$.

Uvědomte si, že B se v prvé složce chová jako zásobníkový automat A, ve druhé složce jako deterministický konečný automat M.

Není těžké nahlédnout, že zásobníkový automat B opravdu přijímá jazyk $L \cap R$.

3.4.7 Tvrzení. Bezkontextové jazyky jsou uzavřeny na reverzi.

To znamená, je-li L bezkontextový jazyk, pak také jazyk L^R je bezkontextový. □

Zdůvodnění: Jazyk L je bezkontextový, proto existuje bezkontextová gramatika $\mathcal{G} = (N, \Sigma, S, P)$, která ho generuje. Gramatika $\mathcal{G}_R$, která generuje jazyk L^R je (N, Σ, S, P_R), kde

$$A \to \alpha \in P_R \quad \text{iff} \quad A \to \alpha^R \in P.$$

Není těžké dokázat, že gramatika $\mathcal{G}$ generuje jazyk L^R.

3.4.8 Tvrzení. Třída bezkontextových jazyků je uzavřena na substituce.

Přesněji: Máme dány dvě abecedy Σ a Δ a substituci σ, která každému písmenu $a \in \Sigma$ přiřadí bezkontextový jazyk $\sigma(a) = L_a$ nad abecedou Δ. Je-li L_0 bezkontextový jazyk nad abecedou Σ, pak jeho obraz $\sigma(L_0)$ v substituci σ je také bezkontextový jazyk nad abecedou Δ. (Substituce byla zavedena v 2.7.2.) □

Nástin důkazu: Máme dánu CF gramatiku $\mathcal{G}_0 = (N_0, \Sigma, S_0, P_0)$ takovou, že $L_0 = L(\mathcal{G}_0)$ a dále pro každé $a \in \Sigma$ bezkontextovou gramatiku $\mathcal{G}_a = (N_a, \Delta, S_a, P_a)$ takovou, že $L_a = L(\mathcal{G}_a)$. Předpokládejme, že množiny neterminálů všech gramatik jsou po dvou disjunktní.

Zkonstruujeme CF gramatiku $\mathcal{G} = (N, \Delta, S_0, P)$ takto:

a) Množina neterminálu gramatiky $\mathcal{G}$ se skládá ze všech neterminálů $\mathcal{G}_0$ a $\mathcal{G}_a$, $a \in \Sigma$. Formálně $N = N_0 \cup \bigcup\{N_a \mid a \in \Sigma\}$.

b) Pravidla gramatiky $\mathcal{G}$ se skládají ze všech pravidel gramatik $\mathcal{G}_a$ a pravidla $\mathcal{G}_0$ jsou upravena tak, že v pravých stranách je každý terminál $a \in \Sigma$ je nahrazen neterminálem S_a. Přesněji $P = \bigcup\{P_a \mid a \in \Sigma\} \cup P'$, kde pravidla P' jsou tvořena všemi pravidly z P_0, ve kterých jsme terminál $a \in \Sigma$ nahradili neterminálem S_a.

Není těžké dokázat, že gramatika $\mathcal{G}$ generuje jazyk $\sigma(L_0)$.

3.4.9 Důsledek: Třída bezkontextových jazyků je uzavřena na homomorfismy. □

Protože každý homomorfismus je zvláštním případem substituce a protože každý jazyk obsahující jedno slovo je bezkontextový, vyplývá toto tvrzení z 3.4.8.

Uvedeme ještě jedno tvrzení, ale již bez důkazu.

3.4.10 Tvrzení. Třída bezkontextových jazyků je uzavřena na inversní homomorfní obrazy.

Přesněji: Je-li h homomorfismus abecedy Σ do $\Delta^\star$ a je-li L bezkontextový jazyk nad abecedou Δ, pak $h^{-1}(L)$ je bezkontextový jazyk nad abecedou Σ. □

3.4.11 Dyckovy jazyky. Na závěr si přiblížíme, co „dělá jazyk bezkontextovým“; jinými slovy, co je všem bezkontextovým jazykům společné. Jak uvidíme z následujícího tvrzení, bezkontextové jazyky (které nejsou regulární) v sobě obsahují „správné uzávorkování“ a právě to popisují Dyckovy jazyky.

Definice. Bezkontextový jazyk L se nazývá *Dyckův jazyk*, jestliže je generován gramatikou $\mathcal{G} = (\{S\}, \Sigma, S, P)$, kde $\Sigma = \{a_1, a_1', a_2, a_2', \ldots, a_n, a_n'\}$ a P:

$$S \to \varepsilon \,|\, SS \,|\, a_1 S a_1' \,|\, \ldots \,|\, a_n S a_n'.$$

Jedná se vlastně o jazyk „správně uzávorkovaných výrazů“, kde máme n různých otvíracích závorek a_i a zavíracích závorek a_i'.

3.4.12 Tvrzení. Pro každý bezkontextový jazyk L existuje regulární jazyk R, Dyckův jazyk D a homomorfismus h takový, že $L = h(R \cap D)$. □

3.4.13 Spojení pravidel. Máme CF gramatiku $\mathcal{G}$ a v ní pravidlo $A \to \alpha B \beta$. Jestliže

$$B \to \gamma_1 \,|\, \ldots \,|\, \gamma_k$$

jsou všechna pravidla s levou stranou B, pak nahrazením pravidla $A \to \alpha B \beta$ pravidly

$$A \to \alpha\gamma_1\beta \,|\, \ldots, \,|\, \alpha\gamma_k\beta$$

dostaneme gramatiku, která generuje stejný jazyk jako $\mathcal{G}$.

3.4.14 Odstranění levé rekurze. Předpokládejme, že v CF gramatice $\mathcal{G}$ máme pravidla

$$A \to A\alpha_1,\ A \to A\alpha_2,\ \ldots,\ A \to A\alpha_k$$

a pravidla

$$A \to \beta_1,\ A \to \beta_2,\ \ldots,\ A \to \beta_m$$

jsou všechna ostatní pravidla (tj. β_i nezačíná neterminálem A).

Pak tato pravidla nahradíme pravidly

$$A \to \beta_i,\ A \to \beta_i Z,\ Z \to \alpha_j,\ Z \to \alpha_j Z,\ \ i = 1, \ldots m,\ \ j = 1, \ldots, k;$$

kde Z je nový neterminál.

Tímto nahrazením dostaneme gramatiku generující stejný jazyk, která již nemá pro neterminál A levou rekurzi.