

MASARYKOVA UNIVERZITA
FAKULTA INFORMATIKY



Popis regulárních jazyků pomocí predikátové logiky

DIPLOMOVÁ PRÁCE

Bc. Martin Jonáš

Brno, jaro 2014

Prohlášení

Prohlašuji, že tato diplomová práce je mým původním autorským dílem, které jsem vypracoval samostatně. Všechny zdroje, prameny a literaturu, které jsem při vypracování používal nebo z nich čerpal, v práci řádně cituji s uvedením úplného odkazu na příslušný zdroj.

Vedoucí práce: doc. Mgr. Michal Kunc, Ph.D.

Poděkování

Rád bych poděkoval především vedoucímu své diplomové práce, doc. Mgr. Michalu Kun-
covi, Ph.D., za přínosné rady a připomínky a čas, který věnoval konzultacím. Dále bych rád
poděkoval své rodině za podporu nejen během studia a psaní této diplomové práce.

Shrnutí

Tato diplomová práce se zabývá souvislostí formálních jazyků a predikátové logiky. Je v ní nejprve představen způsob, jakým popisovat formální jazyky pomocí predikátové logiky. Poté je dokázán Büchiho slavný výsledek, podle kterého je libovolný jazyk regulární právě tehdy, když je možné ho popsat monadickou logikou druhého řádu s relacemi „následuje“ a „bezprostředně následuje.“ Dále je dokázán podobný výsledek McNaughtona, podle kterého je libovolný jazyk star-free právě tehdy, když je možné ho popsat logikou prvního řádu s týmiž relacemi. Dále jsou pak definovány Ehrenfeuchtovy-Fraïsseho hry, je ukázána jejich souvislost s logikou a dokázán Thomasův výsledek, podle kterého je libovolný jazyk lokálně prahově testovatelný právě tehdy, když je možné ho popsat logikou prvního řádu jen s relací „bezprostředně následuje.“

Abstract

The thesis deals with connection between formal languages and predicate logic. First, the way to define a formal language using predicate logic is described. Then the famous result of Büchi, which states that the class of regular languages is equal to the class of languages definable by monadic second order logic with “greater than” and “successor” relations, is proven. Then we prove similar theorem of McNaughton, according to which the class of star-free languages is equal to the class of languages definable by first order logic with the same relations. Then Ehrenfeucht-Fraïsse games are introduced and we show their relation to first order definability of languages. We use this relation to prove the theorem of Thomas, according to which the class of locally threshold testable languages is equal to the class of languages definable by first order logic using only “successor” predicate.

Klíčová slova

formální jazyky, regulární jazyky, predikátová logika, monadická logika druhého řádu, monadic second order logic, star-free, Ehrenfeucht-Fraïsse, lokálně prahově testovatelné jazyky, locally treshold testable languages

Obsah

Úvod	1
1 Formální jazyky a predikátová logika	2
1.1 <i>Syntaxe</i>	2
1.2 <i>Sémantika</i>	4
2 Regulární jazyky a logika $\text{MSO}[+1, <]$	7
3 Star-free jazyky a logika $\text{FO}[<]$	17
3.1 <i>Star-free jazyky</i>	17
3.2 <i>Popisná síla logiky</i> $\text{FO}[<]$	22
4 Ehrenfeuchtovy-Fraïssého hry	29
4.1 <i>Definice</i>	29
4.2 <i>Souvislost s logikou prvního řádu</i>	32
4.3 <i>Použití na logiku</i> $\text{FO}[<]$	36
5 Lokálně prahově testovatelné jazyky a logika $\text{FO}[+1]$	38
5.1 <i>Lokálně prahově testovatelné jazyky</i>	38
5.2 <i>Popisná síla logiky</i> $\text{FO}[+1]$	40
Závěr	46
Literatura	47

Úvod

Na popis regulárních jazyků je známých několik různých formalismů – operační přístup pomocí konečných automatů, denotační přístup pomocí regulárních výrazů, algebraický přístup pomocí konečných pologrup nebo generování jazyků pomocí gramatik. V této práci zavedeme další formalismus, a to popis formálních jazyků pomocí predikátové logiky. Kombinace myšlenek logiky a formálních jazyků není ničím novým, již v roce 1936 popsal Alan Turing ve svém článku [Tur36] schopnost predikátové logiky prvního řádu popsat výpočet Turingova stroje, a tím předvedl nerozhodnutelnost této logiky. Avšak základem této práce bude článek J. Richarda Büchiho [Bü60], který ukázal, že každý regulární jazyk se dá popsat formulí monadické logiky druhého řádu. Na rozdíl od zmíněných nástrojů na popis regulárních jazyků je tento přístup *intenzionální* – určuje přímo vlastnosti, které musí slova splňovat, aby mohla patřit do daného jazyka.

V první kapitole práce je tato logika představena formálně a je definován způsob, jak pomocí ní popisovat formální jazyky. Zároveň je jako příklad uvedeno několik jazyků a způsob, jakým je lze popsat pomocí monadické logiky druhého řádu s predikáty „následuje“ a „bezprostředně následuje“. V následující kapitole je uvedeno několik základních poznatků o této logice a zbytek této kapitoly je věnován důkazu Büchiho věty, podle které je libovolný jazyk regulární právě tehdy, když ho lze popsat touto logikou. Ve třetí kapitole jsou nejprve definovány star-free jazyky a je o nich dokázáno několik tvrzení, které jsou potřebné v následujícím textu. Zbytek této kapitoly je opět věnován souvislosti této třídy jazyků s logikou – totiž McNaughtonovy větě, podle které je libovolný jazyk star-free právě tehdy, když ho lze popsat logikou prvního řádu s predikáty „následuje“ a „bezprostředně následuje“.

Ve čtvrté kapitole jsou zavedeny Ehrenfeuchtovy-Fraïsseho hry jako způsob, kterým lze dokazovat nevyjádřitelnost některých vlastností v logice prvního řádu. Tato souvislost her s logikou prvního řádu je v této kapitole dokázána a ilustrována několika příklady. Na konci této kapitoly jsou Ehrenfeuchtovy-Fraïsseho hry použity k důkazu tvrzení, podle kterého lze logikou druhého řádu s predikáty „následuje“ a „bezprostředně následuje“ popsat více jazyků než logikou prvního řádu se stejnými predikáty. V poslední, páté, kapitole je nejprve definována třída lokálně prahově testovatelných jazyků a poté je tento koncept ilustrován na několika příkladech jazyků, o kterých je ukázáno, jestli jsou, nebo nejsou lokálně prahově testovatelné. Na konci této kapitoly je pak uveden důkaz Thomasovy věty, podle které je libovolný jazyk lokálně prahově testovatelný právě tehdy, když je popsatečný logikou prvního řádu, která používá jen predikát „bezprostředně následuje“.

Od čtenáře práce se předpokládají základní znalosti teorie formálních jazyků v rozsahu prvních dvou kapitol textu [ČKK06]. Jiné znalosti nejsou pro čtení této práce potřebné. V práci je kladen důraz na co největší samostatnost – všechna používaná tvrzení jsou v textu také dokázána. Jedním z cílů práce je dokázat všechny potřebné výsledky bez použití algebraického popisu regulárních jazyků pomocí konečných pologrup.

Kapitola 1

Formální jazyky a predikátová logika

Syntakticky se logika zavedená v této práci nijak neliší od standardní definice monadické logiky druhého řádu. Jedná se o klasickou predikátovou logiku rozšířenou o možnost kvantifikace přes množiny objektů. Jediným rozdílem je, že přímo v definici máme speciální predikáty popisující přítomnost znaků abecedy na daných pozicích slova. Rozdíl v sémantice této logiky je patrnější. Neuvažujeme libovolné modely standardní monadické logiky druhého řádu, ale jen konečné a lineárně uspořádané. To přesně odpovídá motivaci, se kterou tuto logiku zavádíme, chceme totiž popisovat konečná slova. Pokud vynecháme požadavek konečnosti modelu, je možné touto logikou popisovat jazyky nekonečných slov.

V Büchiho článku je definovaná logika s přesně danými predikáty. My zvolíme obecnější definici, která určuje základní strukturu této logiky a umožňuje k ní přidat libovolné vhodné predikáty. Zejména se budeme věnovat monadické logice druhého řádu s predikáty „následuje“ a „bezprostředně následuje“ a ukážeme o ní, že popisuje právě všechny regulární jazyky. Formalismus představený v této kapitole je kombinací myšlenek z [Str94] a [Esp12]. Z [Str94] používáme právě možnost definovat logiku s libovolnými predikáty a z [Esp12] používáme definici sémantiky pomocí interpretací jako dvojic obsahujících slovo a zobrazení, které určuje hodnoty volných proměnných.

1.1 Syntaxe

Definice 1.1. Necht $\mathcal{X}_1 = \{x, y, z, \dots\}$ je spočetná množina proměnných prvního řádu s přirozeným uspořádáním, $\mathcal{X}_2 = \{X, Y, Z, \dots\}$ spočetná množina proměnných druhého řádu s přirozeným uspořádáním a $\Sigma = \{a, b, c, \dots\}$ konečná abeceda. Buď dále $R = \{R_1, R_2, \dots, R_k\}$ konečná množina predikátových symbolů pro vhodné $k \in \mathbb{N}$. Označme j_k aritu predikátového symbolu R_k . Definujeme induktivně množinu $\text{MSO}[R_1, R_2, \dots, R_k](\Sigma)$ formulí monadické predikátové logiky druhého řádu nad abecedou Σ s relačními symboly $R_1, R_2, \dots, R_k$ následujícím způsobem:

1. $Q_a(x)$ je formule $\text{MSO}[R_1, R_2, \dots, R_k](\Sigma)$ pro všechna $a \in \Sigma, x \in \mathcal{X}_1$.
2. $R_i(x_1, x_2, \dots, x_{j_i})$ je formule $\text{MSO}[R_1, R_2, \dots, R_k](\Sigma)$ pro všechna $1 \leq i \leq k$ a $x_1, x_2, \dots, x_{j_i} \in \mathcal{X}_1$.
3. $x \in X$ je formule $\text{MSO}[R_1, R_2, \dots, R_k](\Sigma)$ pro všechna $x \in \mathcal{X}_1, X \in \mathcal{X}_2$.

Pokud φ_1 a φ_2 jsou formule $\text{MSO}[R_1, R_2, \dots, R_k](\Sigma)$, pak také

4. $\neg\varphi_1$ je formule $\text{MSO}[R_1, R_2, \dots, R_k](\Sigma)$,
5. $\varphi_1 \wedge \varphi_2$ je formule $\text{MSO}[R_1, R_2, \dots, R_k](\Sigma)$,

6. $\exists x \varphi_1$ je formule $\text{MSO}[R_1, R_2, \dots, R_k](\Sigma)$ pro všechna $x \in \mathcal{X}_1$,

7. $\exists X \varphi_1$ je formule $\text{MSO}[R_1, R_2, \dots, R_k](\Sigma)$ pro všechna $X \in \mathcal{X}_2$.

Nic jiného není formule $\text{MSO}[R_1, R_2, \dots, R_k](\Sigma)$.

Definice 1.2. Pokud z předchozí induktivní definice vypustíme body 3 a 7, mluvíme o $\text{FO}[R_1, R_2, \dots, R_k](\Sigma)$ – množině formulí predikátové logiky prvního řádu nad abecedou Σ s relačními symboly $R_1, R_2, \dots, R_k$.

Definice 1.3. Formulí φ , která vznikla jen pomocí bodu 1, 2 nebo 3, nazveme *atomickou formulí*. Pokud je formule φ atomickou formulí nebo negací atomické formule, nazveme ji *literálem*.

Pro snadnější vyjadřování budeme v následujícím textu používat následující syntaktické zkratky:

- $\varphi_1 \vee \varphi_2$ jako zkratku pro $\neg(\neg\varphi_1 \wedge \neg\varphi_2)$
- $\varphi_1 \rightarrow \varphi_2$ jako zkratku pro $\neg\varphi_1 \vee \varphi_2$
- $\varphi_1 \leftrightarrow \varphi_2$ jako zkratku pro $(\varphi_1 \rightarrow \varphi_2) \wedge (\varphi_2 \rightarrow \varphi_1)$
- $\forall x \varphi$ jako zkratku pro $\neg \exists x \neg \varphi$

Definice 1.4. Nechť φ je formule $\text{MSO}[R_1, R_2, \dots, R_k](\Sigma)$. Induktivně vzhledem ke struktuře formule φ definujeme množinu $\text{free}_1(\varphi)$ *volných proměnných prvního řádu* formule φ a množinu $\text{free}_2(\varphi)$ *volných proměnných druhého řádu* formule φ .

- Je-li $\varphi \equiv Q_a(x)$, pak $\text{free}_1(\varphi) = \{x\}$ a $\text{free}_2(\varphi) = \emptyset$.
- Je-li $\varphi \equiv R_i(x_1, x_2, \dots, x_{j_i})$, pak $\text{free}_1(\varphi) = \{x_1, x_2, \dots, x_{j_i}\}$ a $\text{free}_2(\varphi) = \emptyset$.
- Je-li $\varphi \equiv x \in X$, pak $\text{free}_1(\varphi) = \{x\}$ a $\text{free}_2(\varphi) = \{X\}$.
- Je-li $\varphi \equiv \neg\varphi_1$, pak $\text{free}_1(\varphi) = \text{free}_1(\varphi_1)$ a $\text{free}_2(\varphi) = \text{free}_2(\varphi_1)$.
- Je-li $\varphi \equiv \varphi_1 \wedge \varphi_2$, pak $\text{free}_1(\varphi) = \text{free}_1(\varphi_1) \cup \text{free}_1(\varphi_2)$ a $\text{free}_2(\varphi) = \text{free}_2(\varphi_1) \cup \text{free}_2(\varphi_2)$.
- Je-li $\varphi \equiv \exists x \varphi_1$, pak $\text{free}_1(\varphi) = \text{free}_1(\varphi_1) \setminus \{x\}$ a $\text{free}_2(\varphi) = \text{free}_2(\varphi_1)$.
- Je-li $\varphi \equiv \exists X \varphi_1$, pak $\text{free}_1(\varphi) = \text{free}_1(\varphi_1)$ a $\text{free}_2(\varphi) = \text{free}_2(\varphi_1) \setminus \{X\}$.

Pokud $x \in \text{free}_1(\varphi)$, říkáme, že proměnná prvního řádu x se ve formuli φ vyskytuje *volně*. Pokud $X \in \text{free}_2(\varphi)$, říkáme, že proměnná druhého řádu X se ve formuli φ vyskytuje *volně*. Jako $\text{free}(\varphi)$ označíme $\text{free}_1(\varphi) \cup \text{free}_2(\varphi)$. V případě, že $\text{free}(\varphi) = \emptyset$, nazveme formuli φ *uzavřenou*.

1.2 Sémantika

Nyní popíšeme sémantiku formulí této logiky. Naším cílem je definovat relaci $\models$ takovou, že $w \models \varphi$ intuitivně znamená, že formule φ říká něco pravdivého o slově $w \in \Sigma^*$. Protože budeme tuto relaci definovat induktivně vzhledem ke struktuře formule φ , není možné se omezit jen na uzavřené formule. Z tohoto důvodu nebudeme pravdivost formule φ určovat jen pro slova $w \in \Sigma^*$, ale také pro interpretace $(w, \mathcal{J})$, které kromě samotného slova w obsahují i hodnoty volných proměnných formule φ .

Definice 1.5. Interpretací formule $\varphi \in \text{MSO}[R_1, R_2, \dots, R_k](\Sigma)$ nazveme dvojici $(w, \mathcal{J})$, kde $w \in \Sigma^*$ a $\mathcal{J}$ je zobrazení, které každé volné proměnné prvního řádu x formule φ přiřazuje pozici $\mathcal{J}(x) \in \{1, \dots, |w|\}$ a každé volné proměnné druhého řádu X formule φ přiřazuje podmnožinu pozic $\mathcal{J}(X) \subseteq \{1, \dots, |w|\}$.

Definice 1.6. Realizací logiky $\text{MSO}[R_1, R_2, \dots, R_k](\Sigma)$ nazveme zobrazení $\mathcal{R}$, které každému j -árnímu predikátovému symbolu R_i přiřadí relaci $\mathcal{R}(R_i) \subseteq (\mathbb{Z}^+)^j$.

Definice 1.7. Nechť $f: A \rightarrow B$ je libovolné zobrazení. Pak zápisem $f[b/a]: A \cup \{a\} \rightarrow B \cup \{b\}$ rozumíme zobrazení určené pro libovolné $x \in A \cup \{a\}$ předpisem

$$f[b/a](x) = \begin{cases} b, & \text{jestliže } x = a, \\ f(x), & \text{jinak.} \end{cases}$$

Definice 1.8. Nechť Σ je abeceda a $w \in \Sigma^*$ slovo. Pak pro každé $i \in \{1, \dots, |w|\}$ jako $w[i] \in \Sigma$ označíme písmeno slova w na pozici i .

Definice 1.9. Pro formuli $\varphi \in \text{MSO}[R_1, R_2, \dots, R_k](\Sigma)$, její interpretaci $(w, \mathcal{J})$ a realizaci $\mathcal{R}$ definujeme relaci $\models^{\mathcal{R}}$ induktivně vzhledem ke struktuře formule φ :

- $(w, \mathcal{J}) \models^{\mathcal{R}} Q_a(x) \stackrel{\text{def}}{\iff} w[\mathcal{J}(x)] = a$
- $(w, \mathcal{J}) \models^{\mathcal{R}} R_i(x_1, x_2, \dots, x_{j_i}) \stackrel{\text{def}}{\iff} (\mathcal{J}(x_1), \mathcal{J}(x_2), \dots, \mathcal{J}(x_{j_i})) \in \mathcal{R}(R_i)$
- $(w, \mathcal{J}) \models^{\mathcal{R}} x \in X \stackrel{\text{def}}{\iff} \mathcal{J}(x) \in \mathcal{J}(X)$
- $(w, \mathcal{J}) \models^{\mathcal{R}} \neg \varphi_1 \stackrel{\text{def}}{\iff} (w, \mathcal{J}) \not\models^{\mathcal{R}} \varphi_1$
- $(w, \mathcal{J}) \models^{\mathcal{R}} \varphi_1 \wedge \varphi_2 \stackrel{\text{def}}{\iff} (w, \mathcal{J}|_{\text{free}(\varphi_1)}) \models^{\mathcal{R}} \varphi_1$ a zároveň $(w, \mathcal{J}|_{\text{free}(\varphi_2)}) \models^{\mathcal{R}} \varphi_2$
- $(w, \mathcal{J}) \models^{\mathcal{R}} \exists x \varphi_1 \stackrel{\text{def}}{\iff} x \notin \text{free}(\varphi_1), |w| \geq 1$ a $(w, \mathcal{J}) \models^{\mathcal{R}} \varphi_1$, nebo
 $x \in \text{free}(\varphi_1), |w| \geq 1$ a pro nějaké $i \in \{1, \dots, |w|\}$ platí
 $(w, \mathcal{J}[i/x]) \models^{\mathcal{R}} \varphi_1$
- $(w, \mathcal{J}) \models^{\mathcal{R}} \exists X \varphi_1 \stackrel{\text{def}}{\iff} X \notin \text{free}(\varphi_1)$ a $(w, \mathcal{J}) \models^{\mathcal{R}} \varphi_1$, nebo
 $X \in \text{free}(\varphi_1)$ a pro nějaké $S \subseteq \{1, \dots, |w|\}$ platí
 $(w, \mathcal{J}[S/X]) \models^{\mathcal{R}} \varphi_1$

Pokud $(w, \mathcal{J}) \models^{\mathcal{R}} \varphi$, říkáme, že interpretace $(w, \mathcal{J})$ je *modelem* formule φ . Pro uzavřenou formuli φ navíc klademe $w \models^{\mathcal{R}} \varphi \stackrel{\text{def}}{\iff} (w, \emptyset) \models^{\mathcal{R}} \varphi$.

Poznámka 1.10. V předchozí definici bylo nutné pro formuli ve tvaru $\exists x \varphi_1$, kde x není volná proměnná formule φ_1 , definovat sémantiku odlišným způsobem. V takovém případě totiž dvojice $(w, \mathcal{J}[i/x])$ není interpretací formule φ_1 . Pro formuli $\exists X \varphi_1$ je situace podobná.

Definice 1.11. Nechť $\varphi \in \text{MSO}[R_1, R_2, \dots, R_k](\Sigma)$ je uzavřená formule a $\mathcal{R}$ realizace logiky $\text{MSO}[R_1, R_2, \dots, R_k](\Sigma)$. Pak jazyk $L(\varphi) = \{w \in \Sigma^* \mid w \models^{\mathcal{R}} \varphi\}$ nazveme *jazyk definovaný formulí φ* . Jazyk $L \subseteq \Sigma^*$ nazveme $\text{MSO}[R_1, R_2, \dots, R_k]$ -*definovatelný*, pokud existuje formule $\varphi \in \text{MSO}[R_1, R_2, \dots, R_k](\Sigma)$, která ho definuje.

Definice 1.12. Nechť $\varphi \in \text{MSO}[R_1, R_2, \dots, R_k](\Sigma)$ a $\psi \in \text{MSO}[\hat{R}_1, \hat{R}_2, \dots, \hat{R}_l](\Sigma)$ jsou formule. Buď dále $\mathcal{R}$ realizace logiky $\text{MSO}[R_1, R_2, \dots, R_k](\Sigma)$ a $\hat{\mathcal{R}}$ realizace logiky $\text{MSO}[\hat{R}_1, \hat{R}_2, \dots, \hat{R}_l](\Sigma)$. Pak říkáme, že formule φ a ψ jsou *ekvivalentní*, když mají stejné volné proměnné a pro všechny jejich interpretace $(w, \mathcal{J})$ platí

$$(w, \mathcal{J}) \models^{\mathcal{R}} \varphi \Leftrightarrow (w, \mathcal{J}) \models^{\hat{\mathcal{R}}} \psi.$$

V dalším textu budeme jako predikátové symboly používat převážně relace $+1$ a $<$. Realizací logiky $\text{MSO}[+1, <](\Sigma)$ je zobrazení $\mathcal{R}$ takové, že

$$\begin{aligned} \mathcal{R}(+1) &= \{(m, n) \in \mathbb{Z}^+ \times \mathbb{Z}^+ \mid m = n + 1\}, \\ \mathcal{R}(<) &= \{(m, n) \in \mathbb{Z}^+ \times \mathbb{Z}^+ \mid m < n\}. \end{aligned}$$

Případně použijeme vhodná zúžení tohoto zobrazení pro logiky obsahující jen jeden z těchto predikátových symbolů. Místo $<(x, y)$ budeme psát $x < y$ a místo $+1(x, y)$ budeme psát $x = y + 1$. Pokud bude realizace logiky zřejmá z kontextu, budeme místo $\models^{\mathcal{R}}$ psát jen $\models$.

Dále budeme používat následující syntaktické zkratky:

- $x = y$ jako zkratku pro $\neg(x < y) \wedge \neg(y < x)$
- $\text{first}(x)$ jako zkratku pro $\neg \exists y (x = y + 1)$
- $\text{last}(x)$ jako zkratku pro $\neg \exists y (y = x + 1)$

Příklad 1.13. Ukážeme příklady $\text{MSO}[+1, <]$ -definovatelných jazyků.

- $L_1 = \{a\}\{a, b\}^* \subseteq \{a, b\}^*$. Tedy L_1 je jazyk všech slov, která začínají písmenem a . Jazyk L_1 můžeme popsat formulí

$$\varphi_1 \equiv \exists x (Q_a(x) \wedge \text{first}(x)).$$

Tato formule říká, že v každém slově existuje pozice x taková, že na ní je písmeno a a x je první pozice ve slově.

- $L_2 = \{a\}\{b\}^* \subseteq \{a, b\}^*$. Tedy L_2 je jazyk všech slov, která začínají písmenem a a za ním mohou následovat jen písmena b . Jazyk L_2 můžeme popsat formulí

$$\varphi_2 \equiv \exists x (Q_a(x) \wedge \text{first}(x) \wedge \forall y (x < y \rightarrow Q_b(y))).$$

Tato formule říká, že v každém slově opět existuje pozice x taková, že na ní je písmeno a , x je první pozice ve slově a na všech pozicích odlišných od x je písmeno b .

- $L_3 = \{a, b\}^* \{ab\} \{a, b\}^* \subseteq \{a, b\}^*$. Tedy L_3 je jazyk všech slov, která obsahují podslovo ab . Jazyk L_3 můžeme popsat formulí

$$\varphi_3 \equiv \exists x \exists y (Q_a(x) \wedge Q_b(y) \wedge y = x + 1).$$

Tato formule říká, že v každém slově existují pozice x a y takové, že na pozici x je písmeno a , na pozici y je písmeno b a pozice y následuje bezprostředně za pozicí x .

- $L_4 = \{a, b\}^* \{c\} \{a, b\}^* \{c\} \{a, b\}^* \subseteq \{a, b, c\}^*$. Tedy L_4 je jazyk všech slov, která obsahují právě dva znaky c . Jazyk L_4 můžeme popsat formulí

$$\varphi_4 \equiv \exists x \exists y (\neg(x = y) \wedge Q_c(x) \wedge Q_c(y) \wedge \forall z (Q_c(z) \rightarrow (z = x \vee z = y))).$$

Tato formule říká, že v každém slově existují dvě rozdílné pozice x a y takové, že na obou pozicích je písmeno c a zároveň každé písmeno c je na jedné z těchto pozic.

- $L_5 = \{\varepsilon\} \subseteq \{a, b\}^*$. Jazyk L_5 můžeme popsat formulí

$$\varphi_5 \equiv \forall x \neg(x = x).$$

Tato formule říká, že pro všechny pozice x platí $\neg(x = x)$. Ale $x = x$ platí pro každou pozici, takže tuto formuli splňuje jenom slovo, ve kterém žádné pozice nejsou – prázdné slovo.

- $L_6 = \emptyset \subseteq \{a, b\}^*$. Jazyk L_6 můžeme popsat formulí

$$\varphi_6 \equiv \exists x \neg(x = x).$$

Význam této formule je podobný formulí φ_5 , ale s tím rozdílem, že požadujeme existenci pozice x . Tím vyloučíme i prázdné slovo, protože v tom žádná taková pozice x neexistuje.

- $L_7 = (\{a, b\}^2)^+ \subseteq \{a, b\}^*$. Tedy L_7 je jazyk všech slov nad abecedou $\{a, b\}$ kladné sudé délky.

$$\begin{aligned} \varphi_7 \equiv & \exists X (\exists x (x \in X \wedge \text{first}(x)) \wedge \\ & \wedge \exists y (\neg(y \in X) \wedge \text{last}(y)) \wedge \\ & \wedge \forall x \forall y (y = x + 1 \rightarrow (x \in X \leftrightarrow \neg(y \in X)))). \end{aligned}$$

Tato formule říká, že v každém slově existuje podmnožina pozic X taková, že první pozice ve slově do X patří, poslední pozice do X nepatří a pro každé dvě následující pozice do X patří právě jedna z nich. Tedy do X patří právě každá lichá pozice ve slově.

Kapitola 2

Regulární jazyky a logika $\text{MSO}[+1, <]$

Přirozenou otázkou je, jakou popisnou sílu má právě definovaná logika. Odpověď na tuto otázku samozřejmě závisí na volbě predikátů, které dovolíme používat. Například pomocí logiky $\text{MSO}[+]$ lze popsat některé jazyky, které jsou kontextové a nejsou bezkontextové, ale nelze pomocí ní popsat některé regulární jazyky (viz [Tho06]). Nebo, jak ukázal Gödel v [Gö34], pomocí logiky $\text{MSO}[1, =, +, \times]$ ¹ lze popsat všechny rekurzivně spočetné jazyky. Jak už jsme naznačili v úvodu práce, logikou $\text{MSO}[+1, <]$ lze popsat právě všechny regulární jazyky. Tento výsledek nyní dokážeme spolu s několika dalšími tvrzeními o MSO logice s různými kombinacemi predikátů $=$, $+1$ a $<$.

Jak ukazuje následující věta, nepotřebujeme k definování syntaktické zkratky $x = y$ v logice druhého řádu ve skutečnosti žádný další predikát. Avšak syntaktickou zkratku $x = y$ budeme používat i v logice prvního řádu, a proto jsme ji zavedli pomocí predikátu $<$.

Věta 2.1. *Nechť Σ je libovolná abeceda. Jazyk $L \subseteq \Sigma^*$ je $\text{MSO}[\]$ -definovatelný právě tehdy, když je $\text{MSO}[=]$ -definovatelný.*

Důkaz. Implikace $\Rightarrow$ je zřejmá. Pro důkaz implikace $\Leftarrow$ stačí ukázat, že formuli $x = y$ logiky $\text{MSO}[=]$ lze vyjádřit jako ekvivalentní formuli $\varphi(x, y) \in \text{MSO}[\]$. Takovou formulí je

$$\varphi(x, y) = \forall X (\forall z \forall w (z \in X \Leftrightarrow w \in X)).$$

Pokud $\mathcal{J}(x)$ a $\mathcal{J}(y)$ jsou opravdu stejné pozice, formule $\varphi(x, y)$ zřejmě platí. Naopak, pokud jsou $\mathcal{J}(x)$ a $\mathcal{J}(y)$ různé pozice, formule $z \in X \Leftrightarrow w \in X$ neplatí pro volbu $\mathcal{J}(X) = \{\mathcal{J}(x)\}$, $\mathcal{J}(z) = \mathcal{J}(x)$ a $\mathcal{J}(w) = \mathcal{J}(y)$. $\square$

Věta 2.2. *Nechť Σ je libovolná abeceda. Jazyk $L \subseteq \Sigma^*$ je $\text{MSO}[<]$ -definovatelný právě tehdy, když je $\text{MSO}[+1]$ -definovatelný.*

Důkaz. „ $\Rightarrow$ “: Stačí ukázat, že predikát $<$ lze vyjádřit v logice $\text{MSO}[+1](\Sigma)$. Ukážeme tedy, že umíme najít formuli $\psi(x, y) \in \text{MSO}[+1](\Sigma)$ s volnými proměnnými x a y ekvivalentní formuli $\varphi \equiv x < y$.

Pro přehlednost výsledné formule zavedeme čtyři pomocné formule, které využijeme při konstrukci formule ψ . Formule $\text{continuous}(X, x, y)$ má volné proměnné X , x a y a je pravdivá v interpretaci $(w, \mathcal{J})$, právě když $\mathcal{J}(X)$ tvoří souvislý blok pozic, který na levé straně může skončit pouze začátkem slova nebo pozicí $\mathcal{J}(x)$ a na pravé straně koncem slova nebo pozicí $\mathcal{J}(y)$. Tedy pokud je nějaká pozice v v $\mathcal{J}(X)$ a je odlišná od $\mathcal{J}(x)$, tak její levý soused musí také patřit do $\mathcal{J}(X)$. Analogicky musí platit, že pokud je nějaká pozice z v $\mathcal{J}(X)$

1. Realizace predikátových symbolů této logiky je zobrazení takové, že $\mathcal{R}(1) = \{1\}$, $\mathcal{R}(=) = \text{id}_{\mathbb{Z}^+}$, $\mathcal{R}(+) = \{(m, n, k) \in (\mathbb{Z}^+)^3 \mid m + n = k\}$ a $\mathcal{R}(\times) = \{(m, n, k) \in (\mathbb{Z}^+)^3 \mid mn = k\}$.

a je odlišná od y , pak její pravý soused musí také patřit do $\mathcal{J}(X)$.

$$\begin{aligned} \text{continuous}(X, x, y) \equiv \forall z \forall v (v = z + 1 \rightarrow ((v \in X \wedge \neg(x = v)) \rightarrow z \in X) \wedge \\ \wedge ((z \in X \wedge \neg(y = z)) \rightarrow v \in X))) \end{aligned}$$

V této předchozí formuli nemůžeme použít definovanou syntaktickou zkratku pro $x = v$, protože v logice $\text{MSO}[+1]$ nemáme predikát $<$. Místo toho je potřeba použít formuli definovanou v důkazu Věty 2.1.

Další dvě formule jsou $\text{leftBoundary}(X, x)$ a $\text{rightBoundary}(X, x)$. Formule $\text{leftBoundary}(X, x)$ má volné proměnné X a x a $(w, \mathcal{J})$ je jejím modelem, právě když pozice $\mathcal{J}(x)$ patří do $\mathcal{J}(X)$ a levý soused $\mathcal{J}(x)$ (pokud existuje) v $\mathcal{J}(X)$ neleží. Analogicky $\text{rightBoundary}(X, x)$ je pravdivá, když pozice $\mathcal{J}(x)$ patří do množiny $\mathcal{J}(X)$ a její pravý soused (pokud existuje) do ní nepatří.

$$\begin{aligned} \text{leftBoundary}(X, x) &\equiv x \in X \wedge \neg \exists z (z \in X \wedge x = z + 1) \\ \text{rightBoundary}(X, x) &\equiv x \in X \wedge \neg \exists z (z \in X \wedge z = x + 1) \end{aligned}$$

Poslední formule je $\text{twoConsecutive}(X)$, která má jednu volnou proměnnou X a $(w, \mathcal{J})$ je jejím modelem, právě když $\mathcal{J}(X)$ obsahuje nějaké dvě sousední pozice.

$$\text{twoConsecutive}(X) \equiv \exists z \exists v (z \in X \wedge v \in X \wedge v = z + 1)$$

Pak hledaná formule ψ je:

$$\begin{aligned} \psi(x, y) \equiv \exists X (\text{leftBoundary}(X, x) \wedge \text{rightBoundary}(X, y) \wedge \\ \wedge \text{continuous}(X, x, y) \wedge \text{twoConsecutive}(X)) \end{aligned}$$

Formule ψ je skutečně ekvivalentní s formulí $\varphi \equiv x < y$. Předpokládejme $\mathcal{J}(x) < \mathcal{J}(y)$, pak množina $S = \{\mathcal{J}(x), \dots, \mathcal{J}(y)\}$ tvoří souvislý blok pozic, $\mathcal{J}(x)$ je její levý okraj, $\mathcal{J}(y)$ je její pravý okraj a množina S má alespoň dva sousední prvky. Předpokládejme naopak, že množina $\mathcal{J}(X)$ tvoří souvislý blok pozic, $\mathcal{J}(x)$ je její levý okraj, $\mathcal{J}(y)$ je její pravý okraj a $\mathcal{J}(X)$ má alespoň dva po sobě následující prvky. Pak jistě $\mathcal{J}(x) < \mathcal{J}(y)$.

„ $\Leftarrow$ “: Naopak ukážeme, že predikát $+1$ lze vyjádřit v logice $\text{MSO}[<](\Sigma)$. Ukážeme tedy, že umíme najít formuli $\psi(x, y) \in \text{MSO}[<](\Sigma)$ s volnými proměnnými x a y ekvivalentní formuli $\varphi \equiv y = x + 1$. Taková formule je

$$\psi(x, y) \equiv x < y \wedge \neg \exists z ((x < z) \wedge (z < y)). \quad \square$$

Důsledek 2.3. *Nechť Σ je libovolná abeceda. Jazyk $L \subseteq \Sigma^*$ je $\text{MSO}[<, +1]$ -definovatelný právě tehdy, když je $\text{MSO}[+1]$ -definovatelný.*

Důkaz. Jedna inkluze je zřejmá, protože z definice platí $\text{MSO}[+1] \subseteq \text{MSO}[<, +1]$. Druhá inkluze vyplývá z předchozího důkazu, ve kterém jsme ukázali, že libovolná $\text{MSO}[<]$ formule se dá vyjádřit jako ekvivalentní formule logiky $\text{MSO}[+1]$. Toto tvrzení stačí použít v indukci vzhledem ke struktuře formule $\text{MSO}[<, +1]$. $\square$

Lemma 2.4. *Nechť Σ je libovolná abeceda. Každý regulární jazyk $L \subseteq \Sigma^*$ je MSO[+1]-definovatelný.*

Důkaz. Buď Σ abeceda a $L \subseteq \Sigma^*$ regulární jazyk. Tedy existuje konečný deterministický automat $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ takový, že $L(\mathcal{A}) = L$. Stavy automatu označíme $Q = \{q_0, q_1, \dots, q_m\}$. Dále buď $w = w_1 w_2 \dots w_n \in \Sigma^*$ libovolné slovo. Ukážeme, že vlastnost „výpočet automatu $\mathcal{A}$ nad slovem w skončí v akceptujícím stavu“ lze vyjádřit formulí $\psi_{\mathcal{A}}$ logiky MSO[+1].

Budeme předpokládat, že $\varepsilon \notin L$. V opačném případě použijeme dále popsanou konstrukci, abychom získali formuli ψ' takovou, že $L(\psi') = L \setminus \{\varepsilon\}$, a výslednou formuli $\psi_{\mathcal{A}}$ pak vyjádříme jako $\psi_{\mathcal{A}} \equiv \psi' \vee \varphi_5$ (kde $\varphi_5 \equiv \forall x \neg(x = x)$ je formule z příkladu 1.13, která definuje jazyk $\{\varepsilon\}$).

Ukážeme, že slovo w patří do jazyka L , právě když existují množiny

$$P_{q_0}, P_{q_1}, \dots, P_{q_m} \subseteq \{1, 2, \dots, n\},$$

které splňují následující podmínky:

1. Množiny $P_{q_0}, P_{q_1}, \dots, P_{q_m}$ jsou po dvou disjunktní a jejich sjednocení je množina $\{1, 2, \dots, n\}$.
2. $1 \in P_{\delta(q_0, w_1)}$.
3. Pokud $j \in P_{q_i}$ a $j < n$, pak $j + 1 \in P_{\delta(q_i, w_{j+1})}$.
4. $n \in \bigcup_{q \in F} P_q$.

Buď $w \in L$. Pak automat $\mathcal{A}$ akceptuje slovo w . Definujeme pro libovolné $q \in Q$ množinu P_q předpisem

$$P_q = \{j \in \{1, 2, \dots, n\} \mid \hat{\delta}(q_0, w_1 w_2 \dots w_j) = q\}.$$

Tedy pozice j patří do množiny P_q , právě když automat $\mathcal{A}$ po přečtení prvních j znaků slova w skončí ve stavu q . Pak množiny $P_{q_0}, P_{q_1}, \dots, P_{q_m}$ jsou po dvou disjunktní a jejich sjednocení je množina $\{1, 2, \dots, n\}$, protože po každém kroku je automat právě v jednom stavu. Po přečtení prvního písmene je automat ve stavu $\delta(q_0, w_1)$, a tedy $1 \in P_{\delta(q_0, w_1)}$. Pokud je automat ve stavu q_i a nepřečetl celý vstup, bude po přečtení následujícího písmene ve stavu $\delta(q_i, w_{j+1})$. A protože $w \in L$, platí $\hat{\delta}(q_0, w_1 w_2 \dots w_n) \in F$, a tedy také $n \in \bigcup_{q \in F} P_q$.

Buďte naopak $P_{q_1}, \dots, P_{q_m}$ množiny splňující tyto 4 podmínky. Protože tyto množiny jsou po dvou disjunktní a jejich sjednocení je množina $\{1, 2, \dots, n\}$, můžeme pro každé $i \in \{1, 2, \dots, n\}$ označit symbolem $[i]$ stav $q \in Q$ takový, že $i \in P_q$. Pak ukážeme, že pro každý neprázdný prefix w' slova w platí $\hat{\delta}(q_0, w') = [[w']]$. Důkaz povedeme indukcí vzhledem k délce slova w' :

- Pro $|w'| = 1$ platí $w' = w_1$. Z druhé podmínky dostáváme $1 \in P_{\delta(q_0, w_1)}$, a tedy $[1] = \delta(q_0, w_1) = \hat{\delta}(q_0, w')$.
- Pro $|w'| = k > 1$ platí $w' = w_1 w_2 \dots w_k$. Pak z indukčního předpokladu $\hat{\delta}(q_0, w_1 w_2 \dots w_{k-1}) = [k-1]$. Podle třetí podmínky $k \in P_{\delta([k-1], w_k)}$. Celkově tedy $\hat{\delta}(q_0, w') = \delta(\hat{\delta}(q_0, w_1 w_2 \dots w_{k-1}), w_k) = \delta([k-1], w_k) = [k] = [[w']]$.

Protože w je prefix w , tak také $\hat{\delta}(q_0, w) = [|w|] = [n]$. Tedy podle čtvrté podmínky platí $\hat{\delta}(q_0, w) \in F$ a proto $w \in L$.

Označme pro každé $a \in \Sigma$ symbolem $i_a \in \{0, \dots, m\}$ takový index, pro který platí $\delta(q_0, a) = q_{i_a}$. K dokončení důkazu tvrzení tedy stačí najít formuli logiky MSO[<, +1], která popisuje existenci takových množin $P_{q_0}, P_{q_1}, \dots, P_{q_m}$, které splňují uvedené podmínky.

Uvedeme zvlášť formuli pro každou podmínku:

$$\begin{aligned}\psi_1 &\equiv \forall x \left(\bigvee_{0 \leq i \leq m} x \in X_i \wedge \bigwedge_{0 \leq i < j \leq m} (x \in X_i \rightarrow x \notin X_j) \right) \\ \psi_2 &\equiv \exists x \left(\text{first}(x) \wedge \left(\bigvee_{a \in \Sigma} (Q_a(x) \wedge x \in X_{i_a}) \right) \right) \\ \psi_3 &\equiv \forall x \forall y \left(y = x + 1 \rightarrow \bigvee_{\substack{a \in \Sigma \\ i, j \in \{0, \dots, m\} \\ \delta(q_i, a) = q_j}} (x \in X_i \wedge Q_a(y) \wedge y \in X_j) \right) \\ \psi_4 &\equiv \exists x \left(\text{last}(x) \wedge \left(\bigvee_{\substack{i \in \{0, \dots, m\} \\ q_i \in F}} (x \in X_i) \right) \right)\end{aligned}$$

Pak hledaná formule je

$$\psi_A \equiv \exists X_0 \exists X_1 \dots \exists X_m (\psi_1 \wedge \psi_2 \wedge \psi_3 \wedge \psi_4).$$

□

Poznámka 2.5. Ve formuli ψ_1 jsme použili zápis $\bigvee_{0 \leq i \leq m} (x \in X_i)$, který není součástí definice syntaxe naší logiky. Avšak $\bigvee_{0 \leq i \leq m} (x \in X_i)$ chápeme jako syntaktickou zkratku pro disjunkci $m + 1$ formulí, kde i -tá formule vznikne z $x \in X_i$ dosazením hodnoty i . Podobně používáme také syntaktické zkratky $\bigwedge_{0 \leq i < j \leq m} (\varphi)$ a $\bigvee_{a \in \Sigma} (\varphi)$, jejichž zamýšlený význam je zřejmý. V případě formule $\bigvee_{a \in \Sigma} (\varphi)$ je však nutné, aby množina Σ byla konečná.

Naším dalším cílem bude ukázat, že implikace z předchozí věty platí i druhým směrem, tedy že každý MSO[+1]-definovatelný jazyk je regulární. Protože budeme z formule definující jazyk L induktivně konstruovat automat, který tento jazyk akceptuje, musíme se opět vypořádat s volnými proměnnými. Potřebujeme tedy interpretace $(w, \mathcal{J})$ vyjádřit jako slova nad nějakou vhodnou abecedou. Pro formuli s k volnými proměnnými bude touto abecedou množina $\Sigma \times \{0, 1\}^k$. Intuitivně – ke každé pozici přidáme navíc informaci, jestli je v dané interpretaci hodnotou proměnné prvního řádu, respektive jestli náleží do množiny pozic určených proměnnou druhého řádu.

Definice 2.6. Nechť Σ je abeceda a φ je formule logiky MSO[$R_1, \dots, R_m$](Σ) s volnými proměnnými prvního řádu $x_1, x_2, \dots, x_k$ a volnými proměnnými druhého řádu $X_1, X_2, \dots, X_l$. Dále buď $(w, \mathcal{J})$ interpretace formule φ pro $w = w_1 w_2 \dots w_n \in \Sigma^*$. Pak zápisem $\text{enc}(w, \mathcal{J})$ rozumíme slovo $v_1 v_2 \dots v_n$ nad abecedou $\Sigma \times \{0, 1\}^{k+l}$ takové, že

$$v_i = (w_i, \overline{x_1}, \overline{x_2}, \dots, \overline{x_k}, \overline{X_1}, \overline{X_2}, \dots, \overline{X_l})$$

pro všechna $1 \leq i \leq n$, kde

$$\overline{x_j} = \begin{cases} 1, & \text{jestliže } i = \mathcal{J}(x_j), \\ 0, & \text{jinak.} \end{cases}$$

$$\overline{X_j} = \begin{cases} 1, & \text{jestliže } i \in \mathcal{J}(X_j), \\ 0, & \text{jinak.} \end{cases}$$

Tedy slovo $\text{enc}(w, \mathcal{J})$ nezávisí sémantice konkrétní formule, ale jen na jejích volných proměnných. Pro lepší vizuální reprezentaci budeme znaky abecedy $\Sigma \times \{0, 1\}^{k+l}$ místo do řádku zapisovat do sloupce.

Příklad 2.7. Ukážeme příklad slova $\text{enc}(w, \mathcal{J})$ pro nějaké w a $\mathcal{J}$. Nechť φ je formule s volnými proměnnými prvního řádu x, y, z a volnými proměnnými druhého řádu X, Y . Buď dále $w = abaab$ a $\mathcal{J}$ zobrazení takové, že

$$\begin{aligned} \mathcal{J}(x) &= 4, \\ \mathcal{J}(y) &= 1, \\ \mathcal{J}(z) &= 1, \\ \mathcal{J}(X) &= \{1, 3, 4\}, \\ \mathcal{J}(Y) &= \{3, 5\}. \end{aligned}$$

Potom

$$\text{enc}(w, \mathcal{J}) = \begin{bmatrix} a \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} b \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} a \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} a \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} b \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Definice 2.8. Nechť Σ je libovolná abeceda, φ formule logiky $\text{MSO}[R_1, \dots, R_m](\Sigma)$ s k volnými proměnnými prvního řádu a l volnými proměnnými druhého řádu. Pak $\text{Enc}(\varphi) \subseteq (\Sigma \times \{0, 1\}^{k+l})^*$ značíme jazyk $\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \text{ je interpretace formule } \varphi\}$.

Definice 2.9. Nechť Σ je abeceda, φ formule logiky $\text{MSO}[R_1, \dots, R_m](\Sigma)$ a x její volná proměnná prvního řádu. Buď dále $\text{free}_1(\varphi) = \{x_1, \dots, x_k\}$, $\text{free}_2(\varphi) = \{X_1, \dots, X_l\}$ a $1 \leq i \leq k$ takový index, pro který $x_i = x$. Označme $\Sigma_\varphi^{x=1}$ a $\Sigma_\varphi^{x=0}$ následující množiny.

$$\begin{aligned} \Sigma_\varphi^{x=1} &= \{(a, I) \in \Sigma \times \{0, 1\}^{k+l} \mid p_i^{k+l}(I) = 1\} \\ \Sigma_\varphi^{x=0} &= \{(a, I) \in \Sigma \times \{0, 1\}^{k+l} \mid p_i^{k+l}(I) = 0\} \end{aligned}$$

Lemma 2.10. Nechť Σ je libovolná abeceda a φ formule logiky $\text{MSO}[+1](\Sigma)$. Pak jazyk $\text{Enc}(\varphi)$ je regulární.

Důkaz. Buď Σ libovolná abeceda a φ formule logiky $\text{MSO}[+1](\Sigma)$ s k volnými proměnnými prvního řádu a l volnými proměnnými druhého řádu. Jediné omezení, které musí slovo

$w \in \Sigma \times (\{0, 1\}^{k+l})^*$ splňovat, aby patřilo do jazyka $\text{Enc}(\varphi)$, je, že hodnota každé volné proměnné prvního řádu se musí v tomto slově vyskytnout právě jednou. A proto

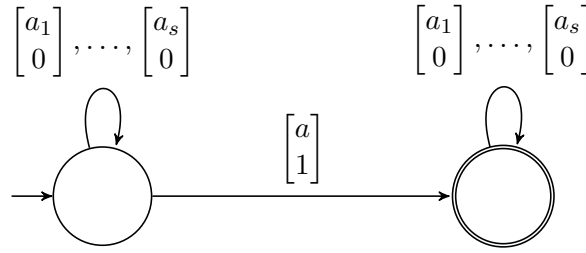
$$\text{Enc}(\varphi) = \bigcap_{x \in \text{free}_1(\varphi)} (\Sigma_{\varphi}^{x=0})^* \Sigma_{\varphi}^{x=1} (\Sigma_{\varphi}^{x=0})^*.$$

Tento jazyk je regulární, protože se jedná o konečný průnik regulárních jazyků. $\square$

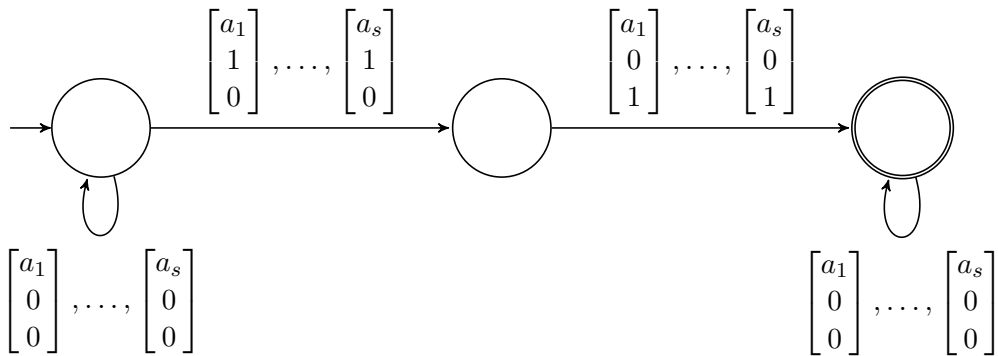
Lemma 2.11. *Nechť Σ je libovolná abeceda a φ formule logiky $\text{MSO}[+1](\Sigma)$. Pak jazyk $\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi\}$ je regulární.*

Důkaz. Buď $\Sigma = \{a_1, a_2, \dots, a_s\}$ libovolná abeceda. Tvzení dokážeme indukcí vzhledem ke struktuře formule φ .

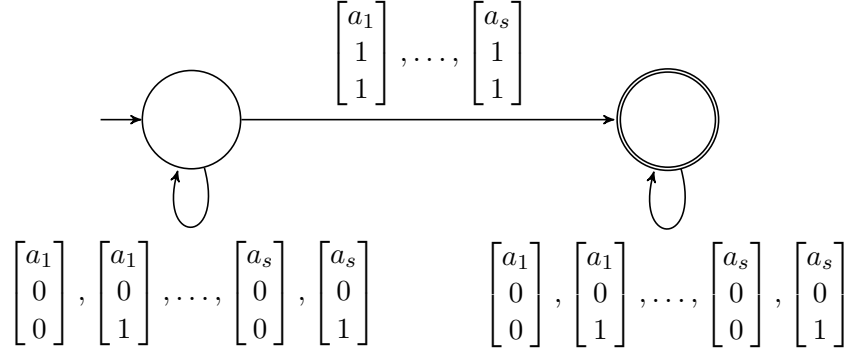
- Je-li $\varphi \equiv Q_a(x)$: Pak $\text{free}(\varphi) = \{x\}$. Jazyk $\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi\}$ je regulární, protože je rozpoznávaný následujícím konečným automatem.



- Je-li $\varphi \equiv y = x + 1$ pro $y \neq x$: Pak $\text{free}(\varphi) = \{x, y\}$. Vztah $(w, \mathcal{J}) \models \varphi$ platí právě tehdy, když $\mathcal{J}(y) = \mathcal{J}(x) + 1$. Jazyk $\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi\}$ je regulární, protože je rozpoznávaný následujícím konečným automatem.



- Je-li $\varphi \equiv x = x + 1$: Protože $(w, \mathcal{J}) \models x = x + 1$ není pravda pro žádnou interpretaci $(w, \mathcal{J})$, platí $\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models x = x + 1\} = \emptyset$.
- Je-li $\varphi \equiv x \in X$: Pak $\text{free}(\varphi) = \{x, X\}$. Vztah $(w, \mathcal{J}) \models \varphi$ platí právě tehdy, když $\mathcal{J}(x) \in \mathcal{J}(X)$. Jazyk $\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi\}$ je regulární, protože je rozpoznávaný následujícím konečným automatem.



- Je-li $\varphi \equiv \neg\varphi_1$: Pak z indukčního předpokladu je jazyk $\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi_1\}$ regulární. Navíc, protože $\text{free}(\varphi) = \text{free}(\varphi_1)$, platí

$$\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi\} = \text{Enc}(\varphi) \setminus \{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi_1\}.$$

Takže z Lemmatu 2.10 a uzavřenosti regulárních jazyků na rozdíl plyne, že jazyk $\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi\}$ je regulární.

- Je-li $\varphi \equiv \varphi_1 \wedge \varphi_2$: Pak podle indukčního předpokladu jsou jazyky $L_1 = \{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi_1\}$ a $L_2 = \{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi_2\}$ regulární. Ale protože formule φ_1 a φ_2 nemusí mít stejné volné proměnné, není vždy pravda, že jazyk $\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi\}$ je průnikem těchto dvou jazyků. Nejprve musíme interpretace obou formulí φ_1 a φ_2 rozšířit tak, aby obsahovaly hodnoty všech proměnných z $\text{free}(\varphi) = \text{free}(\varphi_1) \cup \text{free}(\varphi_2)$. K tomuto využijeme uzavřenost regulárních jazyků na úplný vzor v libovolném homomorfismu. Nechť

$$\begin{aligned} \text{free}_1(\varphi) &= \{x_1, x_2, \dots, x_k\}, \\ \text{free}_2(\varphi) &= \{X_1, X_2, \dots, X_l\}, \\ \text{free}_1(\varphi_1) &= \{x'_1, x'_2, \dots, x'_m\}, \\ \text{free}_2(\varphi_1) &= \{X'_1, X'_2, \dots, X'_n\}. \end{aligned}$$

Pro libovolné $1 \leq i \leq m$ označme $\text{num}(x'_i)$ takové $1 \leq j \leq k$, pro které $x'_i = x_j$. Analogicky pro libovolné $1 \leq i \leq n$ označme $\text{num}(X'_i)$ takové $1 \leq j \leq l$, pro které $X'_i = X_j$.

Definujeme tedy funkci $f_1: \Sigma \times \{0, 1\}^{k+l} \rightarrow \Sigma \times \{0, 1\}^{m+n}$ pro každé $a \in \Sigma$ a $I \in \{0, 1\}^{k+l}$ předpisem

$$f_1((a, I)) = \left(a, \left(p_{\text{num}(x'_1)}^{k+l}(I), \dots, p_{\text{num}(x'_m)}^{k+l}(I), p_{\text{num}(X'_1)+k}^{k+l}(I), \dots, p_{\text{num}(X'_n)+k}^{k+l}(I) \right) \right)$$

Funkci f_1 rozšíříme na homomorfismus $\hat{f}_1: (\Sigma \times \{0, 1\}^{k+l})^* \rightarrow (\Sigma \times \{0, 1\}^{m+n})^*$:

$$\begin{aligned} \hat{f}_1(\varepsilon) &= \varepsilon \\ \hat{f}_1(bw) &= f_1(b) \cdot \hat{f}_1(w) \text{ pro každé } b \in \Sigma \times \{0, 1\}^{k+l} \text{ a } w \in (\Sigma \times \{0, 1\}^{k+l})^* \end{aligned}$$

Analogicky definujeme homomorfismus $\hat{f}_2: (\Sigma \times \{0, 1\}^{k+l})^* \rightarrow (\Sigma \times \{0, 1\}^{|\text{free}(\varphi_2)|})^*$. Pak jazyk $\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi\}$ je regulární, protože

$$\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi\} = \hat{f}_1^{-1}(L_1) \cap \hat{f}_2^{-1}(L_2),$$

regulární jazyky jsou uzavřené na průnik a úplný vzor v libovolném homomorfismu.

- Je-li $\varphi \equiv \exists x \varphi_1$ a $x \notin \text{free}(\varphi_1)$: Pak z indukčního předpokladu je jazyk $L_1 = \{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi_1\}$ regulární. Z definice sémantiky platí $\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi\} = L_1 \setminus \{\varepsilon\}$, a tedy jazyk $\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi\}$ je také regulární z uzavřenosti regulárních jazyků na rozdíl.
- Je-li $\varphi \equiv \exists x \varphi_1$ a $x \in \text{free}(\varphi_1)$: Pak z indukčního předpokladu je jazyk $\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi_1\}$ regulární a $\text{free}(\varphi) = \text{free}(\varphi_1) \setminus \{x\}$. Tedy interpretace formule φ odpovídají interpretacím formule φ_1 bez hodnoty volné proměnné x . Definujeme homomorfismus $\hat{f}: (\Sigma \times \{0, 1\}^{|\text{free}(\varphi_1)|})^* \rightarrow (\Sigma \times \{0, 1\}^{|\text{free}(\varphi)|})^*$, který z interpretací odstraní hodnotu proměnné x . Nechť

$$\begin{aligned} \text{free}_1(\varphi_1) &= \{x_1, x_2, \dots, x_k\}, \\ \text{free}_2(\varphi_1) &= \{X_1, X_2, \dots, X_l\} \end{aligned}$$

a nechť $1 \leq i \leq k$ je takové, že $x_i = x$. Pak definujeme funkci $f: \Sigma \times \{0, 1\}^{k+l} \rightarrow \Sigma \times \{0, 1\}^{k+l-1}$ pro každé $a \in \Sigma$ a $I \in \{0, 1\}^{k+l}$ předpisem

$$f((a, I)) = (a, (p_1^{k+l}(I), \dots, p_{i-1}^{k+l}(I), p_{i+1}^{k+l}(I), \dots, p_{k+l}^{k+l}(I)))$$

a tuto funkci opět rozšíříme na hledaný homomorfismus $\hat{f}$ předpisem

$$\begin{aligned} \hat{f}(\varepsilon) &= \varepsilon, \\ \hat{f}(bw) &= f(b) \cdot \hat{f}(w) \text{ pro každé } b \in \Sigma \times \{0, 1\}^{k+l} \text{ a } w \in (\Sigma \times \{0, 1\}^{k+l})^*. \end{aligned}$$

Pak $\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi\}$ je regulární, protože

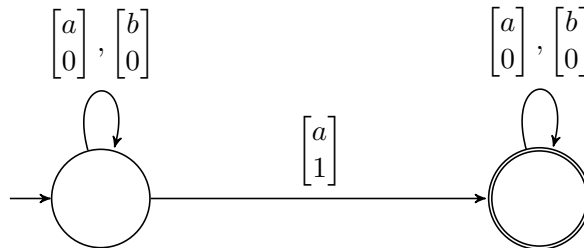
$$\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi\} = \hat{f}(\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi_1\})$$

a regulární jazyky jsou uzavřené na obraz v libovolném homomorfismu.

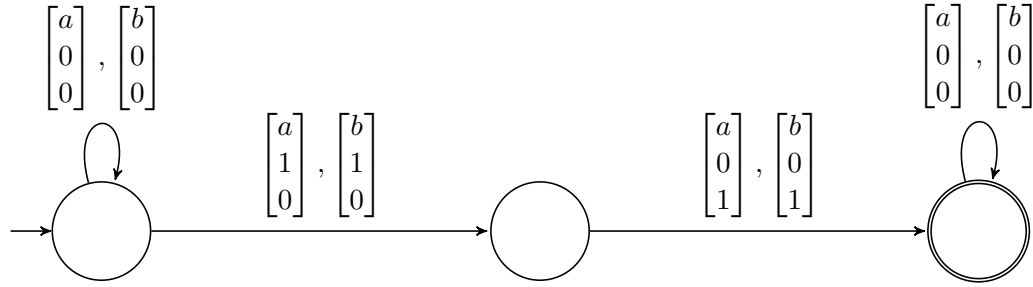
- Je-li $\varphi \equiv \exists X \varphi_1$ a $X \notin \text{free}(\varphi_1)$: Pak z indukčního předpokladu je jazyk $L = \{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi_1\}$ regulární. Z definice sémantiky platí $L = \{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi\}$, a proto je tento jazyk také regulární.
- Je-li $\varphi \equiv \exists X \varphi_1$ a $X \in \text{free}(\varphi_1)$: Důkaz je analogický s případem pro $\varphi \equiv \exists x \varphi_1$. $\square$

Příklad 2.12. Nechť $\Sigma = \{a, b\}$ je abeceda a $\varphi = \exists x \exists y (Q_a(x) \wedge y = x + 1)$ je formule logiky MSO[+1](Σ). Ukážeme, jak předchozím postupem zkonstruovat automat akceptující jazyk $\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi\} \subseteq (\Sigma \times \{0, 1\}^0)^*$. Postupně budeme konstruovat automaty od nejmenších podformulí.

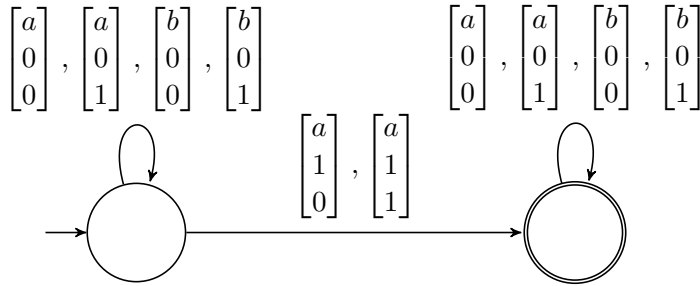
1. Automat rozpoznávající jazyk $\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models Q_a(x)\}$.



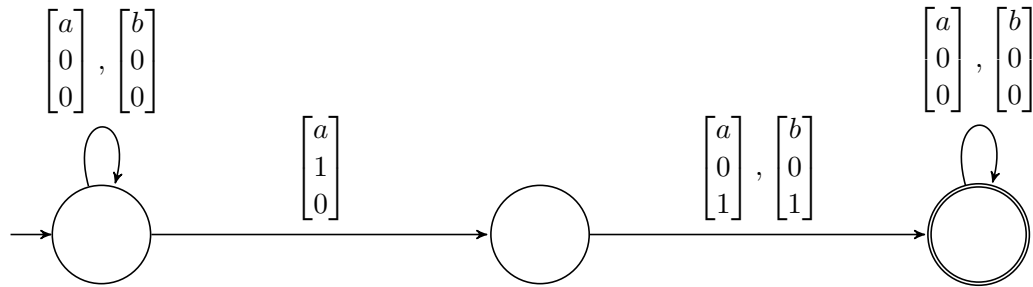
2. Automat rozpoznávající jazyk $\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models y = x + 1\}$.



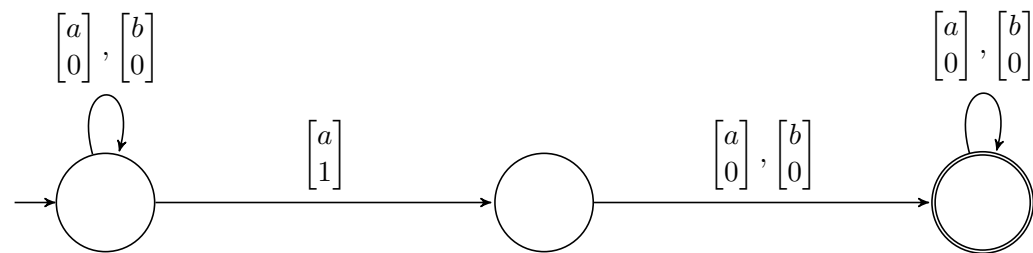
3. Abychom mohli udělat průnik těchto dvou jazyků, musíme do prvního automatu přidat volnou proměnnou y , aby měly obě interpretace stejné volné proměnné. Tím dostaneme automat



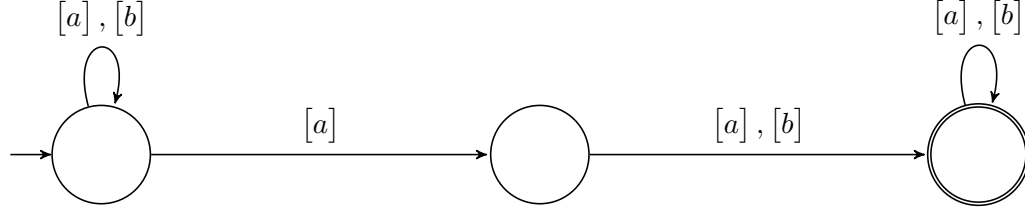
4. Automat rozpoznávající jazyk $\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models Q_a(x) \wedge y = x + 1\}$ zkonstruujeme paralelní synchronní kompozicí předchozích dvou automatů.



5. Automat rozpoznávající jazyk $\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \exists y (Q_a(x) \wedge y = x + 1)\}$ získáme odstraněním informace o hodnotě proměnné y z interpretací akceptovaných předchozím automatem.



6. A konečně automat rozpoznávající jazyk $\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \exists x \exists y (Q_a(x) \wedge y = x + 1)\}$ získáme odstraněním informace o hodnotě proměnné x z interpretací akceptovaných předchozím automatem.



Důsledek 2.13. *Nechť Σ je libovolná abeceda a $L \subseteq \Sigma^*$ je $\text{MSO}[+1]$ -definovatelný jazyk. Pak jazyk L je regulární.*

Důkaz. Nechť $\varphi \in \text{MSO}[+1](\Sigma)$ je formule definující jazyk L . Podle předchozí Věty 2.11 je jazyk $\{\text{enc}(w, \emptyset) \mid (w, \emptyset) \models \varphi\} \subseteq (\Sigma \times \{0, 1\}^0)^*$ regulární. Uvažme funkci $f: \Sigma \times \{0, 1\}^0 \rightarrow \Sigma^*$ určenou pro každé $a \in \Sigma$ předpisem

$$f((a, \emptyset)) = a$$

a její rozšíření na homomorfismus $\hat{f}: (\Sigma \times \{0, 1\}^0)^* \rightarrow \Sigma^*$. Pak $\hat{f}(\{\text{enc}(w, \emptyset) \mid (w, \emptyset) \models \varphi\}) = L$, a tudíž jazyk L je regulární. $\square$

Důsledek 2.14 (Büchi). *Nechť Σ je libovolná abeceda. Pak jazyk $L \subseteq \Sigma^*$ je regulární právě tehdy, když je $\text{MSO}[+1]$ -definovatelný.*

Důkaz. Přímý důsledek Lemmatu 2.4 a předchozího důsledku. $\square$

Kapitola 3

Star-free jazyky a logika $\text{FO}[\leq]$

Tato kapitola je zaměřena na prvořadový fragment dříve definované logiky $\text{MSO}[\leq, +1]$ – logiku $\text{FO}[\leq]$. Ukážeme, že třídu jazyků definovatelných touto logikou lze charakterizovat také pomocí varianty regulárních výrazů, ve kterých nahradíme Kleeneho hvězdičku operací doplňku. Takto popsatelné jazyky se nazývají *star-free* jazyky. Tudiž hlavním cílem této kapitoly je důkaz věty, podle které je jazyk $\text{FO}[\leq]$ -definovatelný, právě když je *star-free*. V této kapitole však nedokážeme, že pomocí logiky $\text{MSO}[\leq, +1]$ lze popsat více jazyků než pomocí logiky $\text{FO}[\leq]$. Příkladem $\text{MSO}[\leq, +1]$ -definovatelného jazyka, který není $\text{FO}[\leq, +1]$ je jazyk $\{aa\}^*$. Tento fakt ukážeme v následující kapitole pomocí Ehrenfeuchtových-Fraïsseho her – nástroje teorie modelů, jímž lze dokázat nevyjádřitelnost některých vlastností v logice prvního řádu.

3.1 Star-free jazyky

Jak bylo uvedeno výše, *star-free* jazyky jsou právě takové jazyky, které lze popsat regulárním výrazem, ve kterém nahradíme Kleeneho hvězdičku operací doplňku. Dá se ukázat (viz [Sch65]), že regulární jazyk $L \subseteq \Sigma^*$ je *star-free* právě tehdy, když existuje takové číslo $k \geq 0$, že pro všechna slova $x, y, z \in \Sigma^*$ platí $xy^kz \in L \Leftrightarrow xy^{k+1}z \in L$. Proto se někdy *star-free* jazyky označují jako *noncounting* jazyky. Neformálně se jedná o vlastnost, že příslušnost slov do *star-free* jazyka se nemůže cyklicky řídit počtem v nich obsažených znaků počítáno ve zbytkových třídách modulo nějaké číslo. Jak je poznamenáno v [Reg09], toto omezení není ve většině přirozených ani umělých jazyků nijak svazující – stěží najdeme například programovací jazyk, který by v názvech proměnných vynucoval počet některých znaků dělitelný pevně daným číslem.

Definice 3.1. Nechť Σ je libovolná abeceda. Definujeme induktivně množinu *star-free regulárních výrazů nad abecedou Σ* následujícím předpisem.

1. Pro každé $a \in \Sigma$ je a *star-free* regulární výraz nad abecedou Σ .
2. $\emptyset$ je *star-free* regulární výraz nad abecedou Σ .

Pokud r a s jsou *star-free* regulární výrazy nad abecedou Σ , pak také

3. $r \cup s$ je *star-free* regulární výraz nad abecedou Σ ,
4. $r \cdot s$ je *star-free* regulární výraz nad abecedou Σ ,
5. $\text{co-}r$ je *star-free* regulární výraz nad abecedou Σ .

Nic jiného není *star-free* regulární výraz nad abecedou Σ .

Definice 3.2. Ke každému star-free regulárnímu výrazu r nad abecedou Σ definujeme induktivně jazyk $L(r) \subseteq \Sigma^*$, který popisuje.

1. $L(a) = \{a\}$
2. $L(\emptyset) = \emptyset$
3. $L(r \cup s) = L(r) \cup L(s)$
4. $L(r \cdot s) = L(r) \cdot L(s)$
5. $L(\text{co-}r) = \Sigma^* \setminus L(r)$

Definice 3.3. Jazyk $L \subseteq \Sigma^*$ nazveme *star-free*, pokud existuje takový star-free regulární výraz r nad abecedou Σ , který splňuje $L(r) = L$. Množinu všech star-free jazyků nad abecedou Σ značíme $\text{SF}(\Sigma)$.

Protože z De Morganových pravidel vyplývá uzavřenost třídy všech star-free jazyků na průnik a rozdíl, zavedeme syntaktické zkratky $r \cap s = \text{co-}(\text{co-}r \cup \text{co-}s)$ a $r \setminus s = r \cap \text{co-}s$. Zároveň budeme v zápisu star-free regulárního výrazu vynechávat operační symbol zřetězení, pokud tím bude zachována jednoznačnost zápisu. Pro libovolnou abecedu Σ lze jazyk Σ^* popsat star-free regulárním výrazem $\text{co-}\emptyset$. Proto zavedeme také syntaktickou zkratku $\Sigma^* = \text{co-}\emptyset$.

Příklad 3.4. Uvedeme příklady několika star-free jazyků.

- Každý konečný jazyk L_1 je star-free, protože lze vyjádřit jako sjednocení jednoprvkových jazyků, odpovídajících slovům jazyka L_1 .
- Jazyk $L_2 = \{a, b\}^* \{ab\} \{a, b\}^*$ nad abecedou $\{a, b\}$ je star-free, protože lze popsat star-free výrazem $\Sigma^* ab \Sigma^*$.
- Jazyk $L_3 = \{a\}^* \{b\}^*$ nad abecedou $\{a, b\}$ je star-free, protože lze popsat star-free výrazem $\text{co-}(\Sigma^* ba \Sigma^*)$. Jazyk L_3 obsahuje totiž právě taková slova, která neobsahují podslovo ba .
- Jazyk $L_4 = \{ab\}^+$ nad abecedou $\{a, b\}$ je star-free, protože lze popsat star-free výrazem $(ab \Sigma^*) \cap (\Sigma^* ab) \cap (\text{co-}(\Sigma^* aa \Sigma^*)) \cap (\text{co-}(\Sigma^* bb \Sigma^*))$. Jazyk L_4 obsahuje totiž právě taková slova, která začínají a končí na ab a zároveň neobsahují dva znaky a , ani dva znaky b za sebou.
- Pro libovolnou abecedu Σ a podmnožinu $A \subseteq \Sigma$ je jazyk A^* nad abecedou Σ star-free, protože lze popsat star-free regulárním výrazem $\text{co-}\bigcup_{a \notin A} (\Sigma^* a \Sigma^*)$.
- Zejména tedy jazyk $L_5 = \{\varepsilon\} = \emptyset^*$ je star-free.

Věta 3.5. Nechť Σ a Φ jsou libovolné abecedy splňující $\Sigma \subseteq \Phi$. Pak platí $\text{SF}(\Sigma) = \text{SF}(\Phi) \cap 2^{\Sigma^*}$.

Důkaz. Nejprve ukážeme, že platí inkluze $\text{SF}(\Sigma) \subseteq \text{SF}(\Phi) \cap 2^{\Sigma^*}$. Nechť $L \in \text{SF}(\Sigma)$ je libovolný star-free jazyk nad abecedou Σ . Tedy existuje star-free regulární výraz r nad abecedou Σ splňující $L(r) = L$. Není však možné pouze použít výraz r jako star-free regulární výraz nad abecedou Φ , protože pak by kvůli odlišné sémantice doplňku mohl popisovat jiný jazyk. Místo toho nicméně stačí induktivně nahradit v r každý výskyt výrazu $\text{co-}s$ výrazem

$\Sigma^* \cap \text{co-}s$. V příkladě 3.4 jsme ukázali, že jazyk Σ^* je star-free jazyk nad abecedou Φ , protože platí $\Sigma \subseteq \Phi$. Tedy platí $\text{SF}(\Sigma) \subseteq \text{SF}(\Phi)$. Inkluze $\text{SF}(\Sigma) \subseteq 2^{\Sigma^*}$ je zřejmá.

Dále ukážeme, že platí i opačná inkluze $\text{SF}(\Sigma) \supseteq \text{SF}(\Phi) \cap 2^{\Sigma^*}$. Indukcí vzhledem ke struktuře star-free regulárního výrazu ukážeme, že pro libovolný star-free regulární výraz r nad abecedou Φ existuje star-free regulární výraz r' nad abecedou Σ takový, že $L(r') = L(r) \cap \Sigma^*$.

- Je-li $r = a$, kde $a \in \Sigma$, pak $r' = a$.
- Je-li $r = a$, kde $a \notin \Sigma$, pak $r' = \emptyset$.
- Je-li $r = \emptyset$, pak $r' = \emptyset$.
- Je-li $r = s \cup t$, pak z indukčního předpokladu existují star-free regulární výrazy s', t' nad abecedou Σ takové, že $L(s') = L(s) \cap \Sigma^*$ a $L(t') = L(t) \cap \Sigma^*$. Tedy stačí zvolit $r' = s' \cup t'$, protože pak bude platit

$$L(r') = L(s') \cup L(t') = (L(s) \cap \Sigma^*) \cup (L(t) \cap \Sigma^*) = (L(s) \cup L(t)) \cap \Sigma^* = L(r) \cap \Sigma^*.$$

- Je-li $r = s \cdot t$, pak z indukčního předpokladu existují star-free regulární výrazy s', t' nad abecedou Σ takové, že $L(s') = L(s) \cap \Sigma^*$ a $L(t') = L(t) \cap \Sigma^*$. Tedy stačí zvolit $r' = s' \cdot t'$, protože pak bude platit

$$L(r') = L(s') \cdot L(t') = (L(s) \cap \Sigma^*) \cdot (L(t) \cap \Sigma^*) = (L(s) \cdot L(t)) \cap \Sigma^* = L(r) \cap \Sigma^*.$$

- Je-li $r = \text{co-}s$, pak z indukčního předpokladu existuje star-free regulární výraz s' nad abecedou Σ takový, že $L(s') = L(s) \cap \Sigma^*$. Tedy stačí zvolit $r' = \text{co-}s'$, protože pak bude platit

$$L(r') = \Sigma^* \setminus L(s') = \Sigma^* \setminus (L(s) \cap \Sigma^*) = (\Sigma^* \setminus L(s)) \cap \Sigma^* = (\Phi^* \setminus L(s)) \cap \Sigma^* = L(r) \cap \Sigma^*.$$

Bud' $L \in \text{SF}(\Phi)$ libovolný jazyk, který splňuje $L \subseteq \Sigma^*$. Pak existuje star-free regulární výraz r nad abecedou Φ takový, že $L(r) = L$. Použitím právě dokázaného tvrzení dostaneme, že existuje takový star-free regulární výraz r' nad abecedou Σ , pro který platí $L(r') = L(r) \cap \Sigma^*$. Ale $L(r) \subseteq \Sigma^*$, takže platí $L(r') = L(r) = L$, a tedy $L \in \text{SF}(\Sigma)$. Tím je dokázána i opačná inkluze. $\square$

Následující lemma ukazuje, že každý star-free jazyk lze pomocí libovolných dvou disjunktních podmnožin jeho abecedy rozložit do určitého tvaru, který využijeme při důkazu souvislosti star-free jazyků s definovatelností pomocí logiky FO[<]. Toto tvrzení a jeho důkaz pochází z [DG08].

Lemma 3.6 (splitting lemma). *Bud' Σ libovolná abeceda, $A, B \subseteq \Sigma$ disjunktní podmnožiny a $L \subseteq \Sigma^*$ star-free jazyk. Pak jazyk $L \cap B^*AB^*$ lze psát ve tvaru*

$$L \cap B^*AB^* = \bigcup_{1 \leq i \leq n} K_i a_i L_i,$$

kde $n \in \mathbb{N}$ a $K_i, L_i \in \text{SF}(B)$, $a_i \in A$ pro každé $1 \leq i \leq n$.

Důkaz. Tvrzení stačí dokázat pro případ, kdy $A = \{a\}$. V případě, kdy $|A| > 1$, lze totiž jazyk $L \cap B^*AB^*$ vyjádřit jako $L \cap B^*AB^* = \bigcup_{a' \in A} (L \cap B^*\{a'\}B^*)$ a použít tvrzení pro případ, kdy množina A je jednoprvková. Naopak pokud $A = \emptyset$, platí $L \cap B^*AB^* = \emptyset$, a tedy stačí zvolit $n = 0$.

Tento případ dokážeme indukcí vzhledem ke struktuře star-free regulárního výrazu, který popisuje jazyk L .

- Je-li $L = L(a)$ pro $a \in A$, pak $L \cap B^*AB^* = \{\varepsilon\}A\{\varepsilon\}$.
- Je-li $L = L(b)$ pro $b \notin A$, pak $L \cap B^*AB^* = \emptyset = \emptyset A \emptyset$.
- Je-li $L = L(\emptyset)$, pak $L \cap B^*AB^* = \emptyset = \emptyset A \emptyset$.
- Je-li $L = L(r \cup s)$, pak je platnost tvrzení zřejmá.
- Je-li $L = L(r \cdot s)$, pak z indukčního předpokladu lze jazyky $L(r) \cap B^*AB^*$ a $L(s) \cap B^*AB^*$ psát jako

$$L(r) \cap B^*AB^* = \bigcup_{1 \leq i \leq m} K_i a L_i,$$

$$L(s) \cap B^*AB^* = \bigcup_{1 \leq i \leq n} M_i a N_i$$

pro vhodná $m, n \in \mathbb{N}$, $K_i, L_i \in \text{SF}(B)$ pro každé $1 \leq i \leq m$ a $M_i, N_i \in \text{SF}(B)$ pro každé $1 \leq i \leq n$. Navíc platí

$$\begin{aligned} (L(r) \cdot L(s)) \cap B^*AB^* &= ((L(r) \cap B^*AB^*) \cdot (L(s) \cap B^*AB^*)) \\ &\quad \cup ((L(r) \cap B^*) \cdot (L(s) \cap B^*AB^*)) \\ &= \bigcup_{1 \leq i \leq m} (K_i a L_i \cdot (L(s) \cap B^*)) \cup \bigcup_{1 \leq i \leq n} ((L(r) \cap B^*) \cdot M_i a N_i). \end{aligned}$$

Tvrzení pak vyplývá z Věty 3.5, protože $L_i \cdot (L(s) \cap B^*) \in \text{SF}(B)$ pro každé $1 \leq i \leq m$ a $(L(r) \cap B^*) \cdot M_i \in \text{SF}(B)$ pro každé $1 \leq i \leq n$.

- Je-li $L = L(\text{co-}r)$, pak z indukčního předpokladu lze jazyk $L(r) \cap B^*AB^*$ psát jako

$$L(r) \cap B^*AB^* = \bigcup_{1 \leq i \leq n} K_i a L_i$$

pro vhodné $n \in \mathbb{N}$ a $K_i, L_i \in \text{SF}(B)$ pro každé $1 \leq i \leq n$.

Bez újmy na obecnosti můžeme předpokládat, že množiny K_i jsou po dvou disjunktní. V opačném případě, kdy $K_i \cap K_j \neq \emptyset$ pro nějaká $1 \leq i < j \leq n$, můžeme sjednocení $K_i a L_i \cup K_j a L_j$ přepsat do tvaru

$$K_i a L_i \cup K_j a L_j = (K_i \setminus K_j) a L_i \cup (K_j \setminus K_i) a L_j \cup (K_i \cap K_j) a (L_i \cup L_j).$$

Také můžeme bez újmy na obecnosti předpokládat, že $\bigcup_{1 \leq i \leq n} K_i = B^*$. V opačném případě můžeme sjednocení $\bigcup_{1 \leq i \leq n} K_i a L_i$ přepsat do tvaru

$$\bigcup_{1 \leq i \leq n} K_i a L_i = \left(\bigcup_{1 \leq i \leq n} K_i a L_i \right) \cup \left(\left(B^* \setminus \bigcup_{1 \leq i \leq n} K_i \right) a \emptyset \right).$$

Tedy množina $\{K_i\}_{1 \leq i \leq n}$ tvoří rozklad množiny B^* . Ukážeme, že pak platí

$$(L(\text{co-}r)) \cap B^*AB^* = \bigcup_{1 \leq i \leq n} K_iA(B^* \setminus L_i).$$

Bud' $w \in (L(\text{co-}r)) \cap B^*AB^*$ libovolné slovo. Tudíž $w = uav$ pro nějaká slova $u, v \in B^*$. Protože množina $\{K_i\}_{1 \leq i \leq n}$ tvoří rozklad množiny B^* , musí existovat takové $1 \leq i \leq n$, pro které $u \in K_i$. Ale $w \notin L(r)$, takže $v \notin L_i$, a proto $w \in K_iA(B^* \setminus L_i)$.

Bud' naopak $w \in \bigcup_{1 \leq i \leq n} K_iA(B^* \setminus L_i)$. Tudíž existuje takové $1 \leq i \leq n$ a slova $u \in K_i, v \in (B^* \setminus L_i)$, pro která $w = uav$. Jistě platí $w \in B^*AB^*$. Pro důkaz sporem předpokládejme, že platí také $w \in L(r)$. Tedy existuje takové $1 \leq j \leq n$, pro které $w \in K_jAL_j$. Protože množina $\{K_i\}_{1 \leq i \leq n}$ tvoří rozklad množiny B^* , platí $K_i = K_j$, a tedy také $L_i = L_j$. Tudíž $v \in L_i$, což je spor. $\square$

Třída star-free jazyků není na rozdíl od třídy regulárních jazyků uzavřená na úplné obrazy v homomorfismech, takže v důkazu souvislosti star-free jazyků s definovatelností v logice FO[<] nebude možné použít tuto uzávěrovou vlastnost podobně, jako jsme ji použili v důkazu Lemmatu 2.11. Jako protipříklad uvažme abecedu $\Sigma = \{a\}$, jazyk $L = \{a\}^*$ nad touto abecedou a homomorfismus $f: \Sigma^* \rightarrow \Sigma^*$ jednoznačně určený podmínkou $f(a) = aa$. Pak jazyk $f(L) = \{aa\}^*$ není star-free, což ukážeme v následující kapitole.

Na úplné vzory v libovolném homomorfismu třída star-free jazyků uzavřená je [CMMP12]. V dalším textu však bude stačit tvrzení, podle kterého je třída star-free jazyků uzavřená na úplný vzor v každém homomorfismu, který navíc splňuje určitou podmínku. Následující definice a věta pochází z [PST01]. Tvrzení dokázané v [PST01] je silnější – podle něj úplný vzor jazyka v abecedním homomorfismu nemůže zvýšit jeho hvězdnou výšku. Věta 3.8 je speciálním případem, kdy uvažujeme star-free jazyk, tedy jazyk hvězdné výšky 0. Slabší formulace tohoto tvrzení nám umožní podat důkaz bez použití teorie konečných pologrup.

Definice 3.7. Necht Σ a Φ jsou libovolné abecedy. O homomorfismu $f: \Sigma^* \rightarrow \Phi^*$ řekneme, že je *abecední*, pokud $f(a) \in \Phi \cup \{\varepsilon\}$ pro každé $a \in \Sigma$.

Věta 3.8. Necht Σ a Φ jsou libovolné abecedy, $f: \Sigma^* \rightarrow \Phi^*$ libovolný abecední homomorfismus a L libovolný star-free jazyk nad abecedou Φ . Pak $f^{-1}(L) \in \text{SF}(\Sigma)$.

Důkaz. Bud' $f: \Sigma^* \rightarrow \Phi^*$ libovolný abecední homomorfismus. Ukážeme indukcí vzhledem ke struktuře, že pro každý star-free regulární výraz r nad abecedou Φ existuje takový star-free regulární výraz r' nad abecedou Σ , který splňuje $L(r') = f^{-1}(L(r))$.

- Je-li $r = a$, pak $f^{-1}(L(r)) = \{w \in \Sigma^* \mid f(w) = a\}$. Označme $\Sigma_\varepsilon \subseteq \Sigma$ množinu $\Sigma_\varepsilon = \{b \in \Sigma \mid f(b) = \varepsilon\}$. Pak

$$\{w \in \Sigma^* \mid f(w) = a\} = \bigcup_{\substack{b \in \Sigma \\ f(b)=a}} \Sigma_\varepsilon^* b \Sigma_\varepsilon^*,$$

a tedy jazyk $f^{-1}(L(r))$ je star-free, protože je konečným sjednocením star-free jazyků.

- Je-li $r = \emptyset$, pak $f^{-1}(L(\emptyset)) = f^{-1}(\emptyset) = \emptyset = L(\emptyset)$.

- Je-li $r = s \cup t$, pak $f^{-1}(L(s \cup t)) = f^{-1}(L(s) \cup L(t)) = f^{-1}(L(s)) \cup f^{-1}(L(t))$. Z indukčního předpokladu existují takové star-free regulární výrazy s', t' nad abecedou Σ , pro které $L(s') = f^{-1}(L(s))$ a $L(t') = f^{-1}(L(t))$. A tudíž $f^{-1}(L(s \cup t)) = L(s') \cup L(t') = L(s' \cup t')$.
- Je-li $r = s \cdot t$, pak $f^{-1}(L(s \cdot t)) = f^{-1}(L(s) \cdot L(t))$. Z indukčního předpokladu existují takové star-free regulární výrazy s', t' nad abecedou Σ , pro které $L(s') = f^{-1}(L(s))$ a $L(t') = f^{-1}(L(t))$. Ukážeme-li, že platí $f^{-1}(L(s) \cdot L(t)) = f^{-1}(L(s)) \cdot f^{-1}(L(t))$, potom $f^{-1}(L(s) \cdot L(t)) = L(s') \cdot L(t') = L(s' \cdot t')$ a tento případ bude dokázán.

Nejdříve ukážeme, že platí inkluze $f^{-1}(L(s) \cdot L(t)) \subseteq f^{-1}(L(s)) \cdot f^{-1}(L(t))$. Buď $w \in f^{-1}(L(s) \cdot L(t))$ libovolné slovo, tedy $f(w) \in L(s) \cdot L(t)$. Takže existují taková slova $u \in L(s), v \in L(t)$, pro která $f(w) = uv$. Rozlišíme tři případy.

- Pokud $u = f(w)$ a $v = \varepsilon$, pak $f(w) \in L(s)$ a $f(\varepsilon) = \varepsilon \in L(t)$. Tudíž $w \in f^{-1}(L(s)) \cdot f^{-1}(L(t))$.
- Pokud $u = \varepsilon$ a $v = f(w)$, tvrzení plyne analogicky.
- Jinak označme $w = w_1 \dots w_n$, kde $w_i \in \Sigma$ pro každé $1 \leq i \leq n$. Protože f je homomorfismus, platí $f(w) = f(w_1) \cdot f(w_2) \dots f(w_n)$. Protože f je dokonce abecední homomorfismus, platí $f(w_i) \in \Phi \cup \{\varepsilon\}$ pro každé $1 \leq i \leq n$. Tudíž musí existovat takové $1 \leq k \leq n$, pro které $u = f(w_1) \dots f(w_k)$ a $v = f(w_{k+1}) \dots f(w_n)$. Tedy $f(w_1) \dots f(w_k) = f(w_1 \dots w_k) \in L(s)$ a $f(w_{k+1}) \dots f(w_n) = f(w_{k+1} \dots w_n) \in L(t)$. Tím pádem platí $w_1 \dots w_k \in f^{-1}(L(s))$ a $w_{k+1} \dots w_n \in f^{-1}(L(t))$, a proto $w \in f^{-1}(L(s)) \cdot f^{-1}(L(t))$.

Nyní ukážeme, že také platí inkluze $f^{-1}(L(s) \cdot L(t)) \supseteq f^{-1}(L(s)) \cdot f^{-1}(L(t))$. Buď tedy $w \in f^{-1}(L(s)) \cdot f^{-1}(L(t))$ libovolné slovo. Pak existují slova $u \in f^{-1}(L(s)), v \in f^{-1}(L(t))$ taková, že $w = uv$. Tudíž platí $f(u) \in L(s), f(v) \in L(t)$, a tedy $f(uv) = f(u) \cdot f(v) \in L(s) \cdot L(t)$. Tím dostáváme $w = uv \in f^{-1}(L(s) \cdot L(t))$.

- Je-li $r = \text{co-}s$, pak $f^{-1}(L(\text{co-}s)) = f^{-1}(\Phi^* \setminus L(s)) = \Sigma^* \setminus f^{-1}(L(s))$, protože pro libovolné zobrazení úplný vzor a doplněk komutují. Z indukčního předpokladu existuje takový star-free regulární výraz s' nad abecedou Σ , pro který $L(s') = f^{-1}(L(s))$. Tedy $f^{-1}(L(\text{co-}s)) = L(\text{co-}s')$. $\square$

3.2 Popisná síla logiky FO[<]

Jak bylo naznačeno v úvodu této kapitoly, dokážeme nyní tvrzení, podle kterého je libovolný jazyk FO[<]-definovatelný právě tehdy, když je star-free. K tomu formulujeme a dokážeme na začátku této části Větu o relativizaci, kterou v důkazu využijeme. Nejprve však ukážeme, že stejně jako v případě logiky druhého řádu lze predikát +1 vyjádřit pomocí formule používající predikát <. Poznamenejme, že na rozdíl od logiky druhého řádu neplatí opačné tvrzení a logiky FO[<] a FO[+1] nejsou stejně silné. Avšak prostředky potřebné k důkazu tohoto tvrzení budou vybudovány až v následující kapitole.

Věta 3.9. *Nechť Σ je abeceda. Jazyk $L \subseteq \Sigma^*$ je FO[+1, <]-definovatelný právě tehdy, když je FO[<]-definovatelný.*

Důkaz. Důkaz je analogický s důkazem Věty 2.2. □

Definice 3.10. Nechť Σ je abeceda, φ formule logiky FO[<] taková, že $x \in \text{free}(\varphi)$. Buď dále $(w, \mathcal{J})$ interpretace formule φ . Nechť $w = w_1 \dots w_n \in \Sigma^*$. Definujeme slova $w_{<x}^{\mathcal{J}}, w_{>x}^{\mathcal{J}}, w_{\leq x}^{\mathcal{J}}, w_{\geq x}^{\mathcal{J}} \in \Sigma^*$ jako

- $w_{<x}^{\mathcal{J}} \stackrel{\text{def}}{=} w_1 \dots w_{\mathcal{J}(x)-1},$
- $w_{>x}^{\mathcal{J}} \stackrel{\text{def}}{=} w_{\mathcal{J}(x)+1} \dots w_n,$
- $w_{\leq x}^{\mathcal{J}} \stackrel{\text{def}}{=} w_1 \dots w_{\mathcal{J}(x)},$
- $w_{\geq x}^{\mathcal{J}} \stackrel{\text{def}}{=} w_{\mathcal{J}(x)} \dots w_n.$

Tedy například $w_{<x}^{\mathcal{J}}$ je takový prefix slova w , za kterým následuje pozice určená hodnotou x v interpretaci $(w, \mathcal{J})$. Je zřejmé, že pro délku takto definovaného slova platí vztah $|w_{<x}^{\mathcal{J}}| = \mathcal{J}(x) - 1$. Jak ukazuje následující věta, která pochází z [Str94], lze ke každé formuli φ logiky FO[<] najít formuli $\varphi_{<x}$ s volnou proměnnou x , ve které všechny proměnné formule φ mohou označovat jen pozice před pozicí určenou hodnotou x . Toto tvrzení následně využijeme k důkazu, že každý star-free jazyk je FO[<]-definovatelný.

Věta 3.11 (o relativizaci). *Nechť Σ je abeceda a φ formule logiky FO[<](Σ). Pro každou proměnnou x prvního řádu splňující $x \notin \text{free}(\varphi)$ existuje formule $\varphi_{<x} \in \text{FO[<]}(\Sigma)$ s volnými proměnnými $\text{free}(\varphi_{<x}) = \text{free}(\varphi) \cup \{x\}$ taková, že pro každou interpretaci $(w, \mathcal{J})$ formule $\varphi_{<x}$ platí*

$$(w, \mathcal{J}) \models \varphi_{<x} \Leftrightarrow (w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\varphi)}) \text{ je interpretace formule } \varphi \text{ a } (w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\varphi)}) \models \varphi. \quad (3.1)$$

Důkaz. Nechť Σ je abeceda, ukážeme indukcí vzhledem ke struktuře, že ke každé formuli $\varphi \in \text{FO[<]}(\Sigma)$ a proměnné x prvního řádu splňující $x \notin \text{free}(\varphi)$ existuje formule $\varphi_{<x} \in \text{FO[<]}(\Sigma)$ s volnými proměnnými $\text{free}(\varphi_{<x}) = \text{free}(\varphi) \cup \{x\}$ splňující podmínku (3.1).

- Je-li $\varphi \equiv Q_a(y)$, pak $\text{free}(\varphi) = \{y\}$. Ukážeme, že hledaná formule je

$$\varphi_{<x} \equiv Q_a(y) \wedge y < x.$$

Nechť $(w, \mathcal{J}) \models Q_a(y) \wedge y < x$, pak $w[\mathcal{J}(y)] = a$ a zároveň $\mathcal{J}(y) < \mathcal{J}(x)$. Tudíž $\mathcal{J}(y) \leq |w_{<x}^{\mathcal{J}}|$, a proto je $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\{y\}})$ interpretace formule φ . Navíc $w_{<x}^{\mathcal{J}}[\mathcal{J}(y)] = a$, a tudíž platí také $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\{y\}}) \models Q_a(y)$.

Nechť naopak $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\{y\}})$ je interpretace formule φ a $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\{y\}}) \models Q_a(y)$. Tedy jistě $\mathcal{J}(y) \leq |w_{<x}^{\mathcal{J}}| = \mathcal{J}(x) - 1 < \mathcal{J}(x)$, a proto $(w, \mathcal{J}) \models y < x$. Navíc $w[\mathcal{J}(y)] = w_{<x}^{\mathcal{J}}[\mathcal{J}(y)] = a$, a tedy $(w, \mathcal{J}) \models Q_a(y)$. Celkově tedy $(w, \mathcal{J}) \models Q_a(y) \wedge y < x$.

- Je-li $\varphi \equiv y < z$, pak $\text{free}(\varphi) = \{y, z\}$. Ukážeme, že hledaná formule je

$$\varphi_{<x} \equiv y < z \wedge z < x.$$

Nechť $(w, \mathcal{J}) \models \varphi_{<x} = y < z \wedge z < x$, pak $\mathcal{J}(y) < \mathcal{J}(z) < \mathcal{J}(x)$. Tudíž $\mathcal{J}(y) < \mathcal{J}(z) \leq |w_{<x}^{\mathcal{J}}|$, a proto je $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\{y,z\}})$ interpretace formule φ . Navíc platí také $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\{y,z\}}) \models y < z$.

Nechť naopak $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\{y,z\}})$ je interpretace formule φ a $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\{y,z\}}) \models y < z$. Tedy $\mathcal{J}(y) < \mathcal{J}(z)$. Stejně jako v předchozím případě se ukáže $\mathcal{J}(z) < \mathcal{J}(x)$, a tudíž platí $(w, \mathcal{J}) \models y < z \wedge z < x$.

- Je-li $\varphi \equiv \psi \wedge \sigma$, pak $\text{free}(\varphi) = \text{free}(\psi) \cup \text{free}(\sigma)$. Ukážeme, že hledaná formule je

$$\varphi_{<x} \equiv \psi_{<x} \wedge \sigma_{<x}.$$

Nechť $(w, \mathcal{J}) \models \psi_{<x} \wedge \sigma_{<x}$, pak z definice sémantiky platí $(w, \mathcal{J}|_{\text{free}(\psi_{<x})}) \models \psi_{<x}$ a $(w, \mathcal{J}|_{\text{free}(\sigma_{<x})}) \models \sigma_{<x}$. Z indukčního předpokladu je dvojice $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\psi)})$ interpretace formule ψ , dvojice $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\sigma)})$ je interpretace formule σ a platí $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\psi)}) \models \psi$ a $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\sigma)}) \models \sigma$. Protože $\text{free}(\varphi) = \text{free}(\psi) \cup \text{free}(\sigma)$, je $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\psi \wedge \sigma)})$ interpretace formule $\varphi \equiv \psi \wedge \sigma$. Z definice sémantiky platí také $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\psi \wedge \sigma)}) \models \psi \wedge \sigma$.

Nechť naopak $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\varphi)})$ je interpretace formule $\varphi \equiv \psi \wedge \sigma$ a platí $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\varphi)}) \models \psi \wedge \sigma$. Pak, protože $\text{free}(\varphi) = \text{free}(\psi) \cup \text{free}(\sigma)$, je dvojice $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\psi)})$ interpretace formule ψ a dvojice $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\sigma)})$ je interpretace formule σ . Navíc z definice sémantiky platí $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\psi)}) \models \psi$ a $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\sigma)}) \models \sigma$. Z indukčního předpokladu pak dostaneme $(w, \mathcal{J}|_{\text{free}(\psi_{<x})}) \models \psi_{<x}$ a $(w, \mathcal{J}|_{\text{free}(\sigma_{<x})}) \models \sigma_{<x}$. Celkově tedy z definice sémantiky platí $(w, \mathcal{J}) \models \psi_{<x} \wedge \sigma_{<x}$.

- Je-li $\varphi \equiv \neg\psi$, pak $\text{free}(\varphi) = \text{free}(\psi)$. Ukážeme, že hledaná formule je

$$\varphi_{<x} \equiv \neg(\psi_{<x}) \wedge \bigwedge_{y \in \text{free}(\psi)} (y < x).$$

Nechť $(w, \mathcal{J}) \models \varphi_{<x}$. Pak z definice sémantiky platí $(w, \mathcal{J}) \not\models \psi_{<x}$ a pro každé $y \in \text{free}(\psi)$ platí $\mathcal{J}(y) < \mathcal{J}(x)$. Tedy $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\psi)})$ je interpretace formule ψ . Z indukčního předpokladu dostaneme, že $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\psi)})$ není interpretace formule ψ nebo $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\psi)}) \not\models \psi$. Protože už víme, že $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\psi)})$ je interpretace formule ψ , platí $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\psi)}) \not\models \psi$. Tedy z definice sémantiky také platí $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\psi)}) \models \neg\psi$. A protože $\text{free}(\varphi) = \text{free}(\psi)$, je toto tvrzení ekvivalentní s $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\varphi)}) \models \neg\psi$.

Nechť naopak $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\varphi)})$ je interpretace formule $\varphi \equiv \neg\psi$ a platí $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\varphi)}) \models \neg\psi$. Dvojice $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\varphi)})$ je jistě také interpretací formule ψ . Z definice sémantiky platí $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\varphi)}) \not\models \psi$. A tedy, protože $\text{free}(\varphi) = \text{free}(\psi)$, platí $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\psi)}) \not\models \psi$. Z indukčního předpokladu pak dostaneme $(w, \mathcal{J}) \not\models \psi_{<x}$. Z definice sémantiky tudíž

platí $(w, \mathcal{J}) \models \neg \psi_{<x}$. Buď dále $y \in \text{free}(\psi)$ libovolné. Protože $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\psi)})$ je interpretací formule ψ , musí platit $\mathcal{J}(y) \leq |w_{<x}^{\mathcal{J}}| < \mathcal{J}(x)$. Celkově tedy dostáváme $(w, \mathcal{J}) \models \varphi_{<x}$.

- Je-li $\varphi \equiv \exists y \psi$, pak $\text{free}(\varphi) = \text{free}(\psi) \setminus \{y\}$. Ukážeme, že hledaná formule je

$$\varphi_{<x} \equiv \exists y (\psi_{<x} \wedge y < x).$$

Nechť $(w, \mathcal{J}) \models \varphi_{<x}$. Tedy $|w| \geq 1$ a existuje takové $i \in \{1, \dots, |w|\}$, které splňuje $(w, \mathcal{J}[i/y]) \models \psi_{<x}$ a $(w, \mathcal{J}[i/y]) \models y < x$. Tudíž platí $i = \mathcal{J}[i/y](y) < \mathcal{J}[i/y](x) = \mathcal{J}(x)$. Tedy z indukčního předpokladu dostaneme, že

$$(w_{<x}^{\mathcal{J}[i/y]}, \mathcal{J}[i/y]|_{\text{free}(\psi)})$$

je interpretace formule ψ a platí

$$(w_{<x}^{\mathcal{J}[i/y]}, \mathcal{J}[i/y]|_{\text{free}(\psi)}) \models \psi.$$

Rozlišíme dva případy.

- Pokud $y \notin \text{free}(\psi)$, pak platí $\text{free}(\varphi) = \text{free}(\psi)$ a $\mathcal{J}[i/y]|_{\text{free}(\psi)} = \mathcal{J}|_{\text{free}(\psi)}$. Tím pádem platí $(w_{<x}^{\mathcal{J}[i/y]}, \mathcal{J}|_{\text{free}(\psi)}) \models \psi$. Tudíž z definice sémantiky platí také $(w_{<x}^{\mathcal{J}[i/y]}, \mathcal{J}|_{\text{free}(\psi)}) \models \exists y \psi$. A z $\text{free}(\varphi) = \text{free}(\psi)$ a $w_{<x}^{\mathcal{J}[i/y]} = w_{<x}^{\mathcal{J}}$ dostaneme, že $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\varphi)})$ je interpretace formule φ a platí $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\varphi)}) \models \exists y \psi$.
- Pokud naopak $y \in \text{free}(\psi)$ a $\text{free}(\varphi) = \text{free}(\psi) \setminus \{y\}$, platí rovnost $\mathcal{J}[i/y]|_{\text{free}(\psi)} = (\mathcal{J}|_{\text{free}(\psi)})[i/y] = (\mathcal{J}|_{\text{free}(\varphi)})[i/y]$. Tím pádem také platí

$$(w_{<x}^{\mathcal{J}[i/y]}, (\mathcal{J}|_{\text{free}(\varphi)})[i/y]) \models \psi.$$

A protože $w_{<x}^{\mathcal{J}[i/y]} = w_{<x}^{\mathcal{J}}$ a $i \in \{1, \dots, |w_{<x}^{\mathcal{J}}|\}$, dostaneme, že $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\varphi)})$ je interpretace formule φ a platí $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\varphi)}) \models \exists y \psi$.

Nyní ukážeme opačnou implikaci. Tedy nechť naopak $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\varphi)})$ je interpretace formule φ a $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\varphi)}) \models \varphi$. Opět rozlišíme dva případy.

- Pokud $y \notin \text{free}(\psi)$, platí z definice sémantiky $|w_{<x}^{\mathcal{J}}| \geq 1$ a $(w_{<x}^{\mathcal{J}}, \mathcal{J}|_{\text{free}(\varphi)}) \models \psi$. Pak, protože platí $\text{free}(\varphi) = \text{free}(\psi)$, dostaneme z indukčního předpokladu $(w, \mathcal{J}) \models \psi_{<x}$. Tedy pokud zvolíme $\mathcal{J}(y) = 1$, platí $\mathcal{J}(y) < \mathcal{J}(x)$, a proto platí také $(w, \mathcal{J}) \models \exists y (\psi_{<x} \wedge y < x)$.
- Pokud naopak $y \in \text{free}(\psi)$, pak $|w_{<x}^{\mathcal{J}}| \geq 1$ a existuje takové $i \in \{1, \dots, |w_{<x}^{\mathcal{J}}|\}$, které splňuje

$$(w_{<x}^{\mathcal{J}}, (\mathcal{J}|_{\text{free}(\varphi)})[i/y]) \models \psi.$$

Opět použijeme rovnosti $\mathcal{J}[i/y]|_{\text{free}(\psi)} = (\mathcal{J}|_{\text{free}(\psi)})[i/y] = (\mathcal{J}|_{\text{free}(\varphi)})[i/y]$ a $w_{<x}^{\mathcal{J}[i/y]} = w_{<x}^{\mathcal{J}}$ a dostaneme

$$(w_{<x}^{\mathcal{J}[i/y]}, \mathcal{J}[i/y]|_{\text{free}(\psi)}) \models \psi.$$

Tento vztah dává smysl, protože pokud $i \in \{1, \dots, |w_{<x}^{\mathcal{J}}|\}$, pak je také $(w_{<x}^{\mathcal{J}[i/y]}, \mathcal{J}[i/y]|_{\text{free}(\psi)})$ interpretací formule ψ . Tudíž podle indukčního předpokladu platí $(w, \mathcal{J}[i/y]) \models \psi_{<x}$. Navíc platí

$$\mathcal{J}[i/y](y) = i < \mathcal{J}(x) = \mathcal{J}[i/y](x).$$

Celkově tedy dostaneme požadovaný vztah $(w, \mathcal{J}) \models \exists y (\psi_{<x} \wedge y < x)$. $\square$

Definice 3.12. Formulí $\varphi_{<x}$ z předchozí věty nazýváme *relativizací* formule φ . Analogicky zavedeme relativizace $\varphi_{>x}$, $\varphi_{\leq x}$ a $\varphi_{\geq x}$.

Lemma 3.13. *Nechť Σ je libovolná abeceda. Každý star-free jazyk $L \subseteq \Sigma^*$ je FO[<]-definovatelný.*

Důkaz. Ukážeme indukcí vzhledem ke struktuře star-free regulárního výrazu popisujícího jazyk L , že existuje uzavřená formule $\varphi \in \text{FO}[<](\Sigma)$, která definuje jazyk L .

- Je-li $L = L(a)$, pak $L = \{a\}$ a tento jazyk definuje uzavřená formule $\varphi \equiv \exists x (Q_a(x) \wedge \forall y (x = y))$.
- Je-li $L = \emptyset$, pak tento jazyk definuje uzavřená formule $\varphi \equiv \exists x \neg(x = x)$.
- Je-li $L = L(r \cup s)$, pak $L = L(r) \cup L(s)$. Z indukčního předpokladu existují takové uzavřené formule $\psi, \sigma \in \text{FO}[<](\Sigma)$, že $L(\psi) = L(r)$ a $L(\sigma) = L(s)$. Tedy jazyk $L(r \cup s)$ se dá definovat uzavřenou formulí $\varphi \equiv \psi \vee \sigma$.
- Je-li $L = L(r \cdot s)$, pak $L = L(r) \cdot L(s)$. Z indukčního předpokladu existují takové uzavřené formule $\psi, \sigma \in \text{FO}[<](\Sigma)$, že $L(\psi) = L(r)$ a $L(\sigma) = L(s)$. Uvažme dva případy.
 - Pokud $\varepsilon \notin L(s)$, zvolíme $\varphi \equiv \exists x (\psi_{<x} \wedge \sigma_{\geq x})$. Ukážeme, že pak $L(\varphi) = L(r) \cdot L(s)$. Nejprve ukážeme inkluzi $L(\varphi) \subseteq L(r) \cdot L(s)$. Buď tedy $w \in L(\varphi)$ libovolné slovo, tedy platí $|w| \geq 1$ a pro nějaké $i \in \{1, \dots, |w|\}$ platí $(w, \emptyset[i/x]) \models \psi_{<x}$ a $(w, \emptyset[i/x]) \models \sigma_{\geq x}$. Z Věty o relativizaci (3.11) pak vyplývá $(w_{<x}^{\emptyset[i/x]}, \emptyset[i/x]|_{\text{free}(\psi)}) \models \psi$ a $(w_{\geq x}^{\emptyset[i/x]}, \emptyset[i/x]|_{\text{free}(\sigma)}) \models \sigma$. Protože $\text{free}(\psi) = \text{free}(\sigma) = \emptyset$, tak také $(w_{<x}^{\emptyset[i/x]}, \emptyset) \models \psi$ a $(w_{\geq x}^{\emptyset[i/x]}, \emptyset) \models \sigma$. Tedy $w_{<x}^{\emptyset[i/x]} \in L(r)$ a $w_{\geq x}^{\emptyset[i/x]} \in L(s)$, a proto $w = w_{<x}^{\emptyset[i/x]} \cdot w_{\geq x}^{\emptyset[i/x]} \in L(r) \cdot L(s)$.
 - Nyní ukážeme inkluzi $L(\varphi) \supseteq L(r) \cdot L(s)$. Buď tedy naopak $w \in L(r) \cdot L(s)$ libovolné slovo, tudíž $w = uv$ pro vhodná slova $u \in L(r)$ a $v \in L(s)$. Víme, že platí $(u, \emptyset) \models \psi$ a $(v, \emptyset) \models \sigma$. Zvolme $i = |u| + 1$. Pak $i \leq |w|$, protože $v \neq \varepsilon$. Uvažme interpretaci $(w, \emptyset[i/x])$ formulí $\psi_{<x}$ a $\sigma_{\geq x}$. Pro tuto interpretaci platí $u = w_{<x}^{\emptyset[i/x]}$ a $v = w_{\geq x}^{\emptyset[i/x]}$. Tím pádem také $(w_{<x}^{\emptyset[i/x]}, \emptyset) \models \psi$ a $(w_{\geq x}^{\emptyset[i/x]}, \emptyset) \models \sigma$. Protože $\text{free}(\psi) = \text{free}(\sigma) = \emptyset$, tak jistě platí $(w_{<x}^{\emptyset[i/x]}, \emptyset[i/x]|_{\text{free}(\psi)}) \models \psi$ a $(w_{\geq x}^{\emptyset[i/x]}, \emptyset[i/x]|_{\text{free}(\sigma)}) \models \sigma$. Tedy opět z Věty o relativizaci vyplývá $(w, \emptyset[i/x]) \models \psi_{<x}$ a $(w, \emptyset[i/x]) \models \sigma_{\geq x}$. Celkově tedy platí $(w, \emptyset) \models \exists x (\psi_{<x} \wedge \sigma_{\geq x})$, a proto $w \in L(\varphi)$.
 - Pokud $\varepsilon \in L(s)$, lze jazyk L psát jako $L = L(r) \cup (L(r) \cdot (L(s) \setminus \{\varepsilon\}))$. Protože jsme už ukázali platnost tvrzení pro sjednocení a pro zřetězení, ve kterém druhý jazyk neobsahuje prázdné slovo, stačí najít formuli definující jazyk $L(s) \setminus \{\varepsilon\}$. Buď x taková proměnná, pro kterou platí $x \notin \text{free}(\sigma)$. Potom jazyk $L(s) \setminus \{\varepsilon\}$ je definovatelný formulí $\exists x \sigma$.

- Je-li $L = L(\text{co-}r)$, pak $L = \Sigma^* \setminus L(r)$. Z indukčního předpokladu existuje taková uzavřená formule $\psi \in \text{FO}[<](\Sigma)$, že $L(\psi) = L(r)$. Tedy jazyk $\Sigma^* \setminus L(r)$ se dá definovat uzavřenou formulí $\varphi \equiv \neg\psi$. $\square$

Stejně jako v případě logiky druhého řádu dokážeme i druhý směr této implikace, podle kterého je každý FO[<]-definovatelný jazyk star-free. Postup bude analogický jako v případě logiky druhého řádu – nejprve ukážeme, že pro libovolnou formuli $\varphi \in \text{FO}[<](\Sigma)$ je jazyk $\text{Enc}(\varphi)$ star-free, a poté, že také jazyk $\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi\}$ je star-free.

Lemma 3.14. *Nechť Σ je libovolná abeceda a φ formule logiky FO[<](Σ). Pak jazyk $\text{Enc}(\varphi)$ je star-free.*

Důkaz. V důkazu Lemmatu 2.10 jsme zdůvodnili rovnost

$$\text{Enc}(\varphi) = \bigcap_{x \in \text{free}(\varphi)} (\Sigma_{\varphi}^{x=0})^* \Sigma_{\varphi}^{x=1} (\Sigma_{\varphi}^{x=0})^*.$$

A tedy jazyk $\text{Enc}(\varphi)$ je star-free, protože se jedná o konečný průnik star-free jazyků. $\square$

Lemma 3.15. *Nechť Σ je libovolná abeceda a φ formule logiky FO[<](Σ). Pak jazyk $\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi\}$ je star-free.*

Důkaz. Tvrzení dokážeme indukcí vzhledem ke struktuře formule φ .

- Je-li $\varphi \equiv Q_a(x)$, pak

$$\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi\} = (\Sigma_{\varphi}^{x=0})^* \cdot (a, (1)) \cdot (\Sigma_{\varphi}^{x=0})^*.$$

- Je-li $\varphi \equiv x < y$, pak

$$\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi\} = (\Sigma_{\varphi}^{x=0} \cap \Sigma_{\varphi}^{y=0})^* \cdot \Sigma_{\varphi}^{x=1} \cdot (\Sigma_{\varphi}^{x=0} \cap \Sigma_{\varphi}^{y=0})^* \cdot \Sigma_{\varphi}^{y=1} \cdot (\Sigma_{\varphi}^{x=0} \cap \Sigma_{\varphi}^{y=0})^*.$$

- Je-li $\varphi \equiv \neg\psi$, pak z indukčního předpokladu je jazyk $\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \psi\}$ star-free. Jazyk $\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi\}$ je tedy také star-free, protože platí

$$\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi\} = \text{Enc}(\varphi) \setminus \{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \psi\}$$

a jazyk $\text{Enc}(\varphi)$ je star-free podle Lemmatu 3.14.

- Je-li $\varphi \equiv \psi_1 \wedge \psi_2$, označme $\text{free}(\varphi) = \{x_1, \dots, x_n\}$. Pak z indukčního předpokladu jsou jazyky $L_1 = \{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \psi_1\}$ a $L_2 = \{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \psi_2\}$ star-free. Označme $k = |\text{free}(\psi_1)|$ a $l = |\text{free}(\psi_2)|$. Stejně jako v případě logiky druhého řádu nemusí být hledaným jazykem průnik těchto jazyků, protože formule ψ_1 a ψ_2 nemusí mít stejné volné proměnné. Uvažme však homomorfismy $\hat{f}_1: (\Sigma \times \{0, 1\}^n)^* \rightarrow (\Sigma \times \{0, 1\}^k)^*$ a $\hat{f}_2: (\Sigma \times \{0, 1\}^n)^* \rightarrow (\Sigma \times \{0, 1\}^l)^*$, které odeberou hodnoty volných proměnných, které se v příslušných formulích nevyskytují, definované v důkazu Lemmatu 2.11. Protože homomorfismy $\hat{f}_1$ a $\hat{f}_2$ jsou abecední, z Věty 3.8 dostaneme, že také jazyk

$$\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi\} = \hat{f}_1^{-1}(L_1) \cap \hat{f}_2^{-1}(L_2)$$

je star-free.

- Je-li $\varphi \equiv \exists x \psi$ a $x \notin \text{free}(\varphi)$, pak z indukčního předpokladu je jazyk $L_1 = \{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \psi\}$ star-free. Z definice sémantiky platí $\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi\} = L_1 \setminus \{\varepsilon\}$, a tedy jazyk $\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi\}$ je také star-free, protože jazyk $\{\varepsilon\}$ je star-free (viz Příklad 3.4) a třída star-free jazyků je uzavřená na rozdíl.
- Je-li $\varphi \equiv \exists x \psi$ a $x \in \text{free}(\varphi)$, označme $\text{free}(\varphi) = \{x_1, \dots, x_n\}$. Pak z indukčního předpokladu je jazyk $L_1 = \{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \psi\}$ star-free. Proto můžeme použít Větu 3.6 a volbou $A = \Sigma_\psi^{x=1}$ a $B = \Sigma_\psi^{x=0}$ dostaneme, že jazyk $L_1 \cap (\Sigma_\psi^{x=0})^* \Sigma_\psi^{x=1} (\Sigma_\psi^{x=0})^*$ lze psát ve tvaru

$$L_1 \cap (\Sigma_\psi^{x=0})^* \Sigma_\psi^{x=1} (\Sigma_\psi^{x=0})^* = \bigcup_{1 \leq i \leq m} K_i c_i L_i,$$

kde $m \in \mathbb{N}$ a $K_i, L_i \in \text{SF}(B)$, $c_i \in A$ pro každé $1 \leq i \leq m$. Protože navíc platí inkluze $L_1 \subseteq B^* A B^*$, platí rovnost $L_1 = L_1 \cap B^* A B^*$, a tedy

$$L_1 = \bigcup_{1 \leq i \leq m} K_i a_i L_i.$$

Označme $1 \leq i \leq n$ takový index, pro který $x_i = x$. Připomeňme znovu funkci $f: \Sigma \times \{0, 1\}^n \rightarrow \Sigma \times \{0, 1\}^{n-1}$, kterou jsme použili v důkazu Lemmatu 2.11, definovanou pro každé $a \in \Sigma$ a $I \in \{0, 1\}^n$ předpisem

$$f((a, I)) = (a, (p_1^n(I), \dots, p_{i-1}^n(I), p_{i+1}^n(I), \dots, p_n^n(I))).$$

Tentokrát tuto funkci rozšíříme na homomorfismus $\hat{f}: (\Sigma_\psi^{x=0})^* \rightarrow (\Sigma \times \{0, 1\}^{n-1})^*$ předpisem

$$\begin{aligned} \hat{f}(\varepsilon) &= \varepsilon, \\ \hat{f}(bw) &= f(b) \cdot \hat{f}(w) \text{ pro každé } b \in \Sigma_\psi^{x=0} \text{ a } w \in (\Sigma_\psi^{x=0})^*. \end{aligned}$$

Protože tento homomorfismus je navíc izomorfismus, jsou jazyky $\hat{f}(K_i)$ a $\hat{f}(L_i)$ star-free jazyky nad abecedou $\Sigma \times \{0, 1\}^{n-1}$ pro každé $1 \leq i \leq m$. A tedy jazyk

$$\{\text{enc}(w, \mathcal{J}) \mid (w, \mathcal{J}) \models \varphi\} = \bigcup_{1 \leq i \leq m} \hat{f}(K_i) \cdot f(a_i) \cdot \hat{f}(L_i)$$

je star-free. □

Důsledek 3.16. *Nechť Σ je libovolná abeceda a $L \subseteq \Sigma^*$ je FO[<]-definovatelný jazyk. Pak jazyk L je star-free.*

Důkaz. K důkazu lze použít stejný argument jako v Důsledku 2.13, protože definovaný homomorfismus $\hat{f}$ je jistě izomorfismus. □

Důsledek 3.17 (McNaughton). *Nechť Σ je libovolná abeceda. Pak jazyk $L \subseteq \Sigma^*$ je star-free právě tehdy, když je FO[<]-definovatelný.*

Důkaz. Přímý důsledek Lemmatu 3.13 a předchozího důsledku. □

Kapitola 4

Ehrenfeuchtovy-Fraïssého hry

V předchozí kapitole jsme popsali jazyky definovatelné logikou $\text{FO}[+1]$. Přirozeně se nabízí otázka, jestli je tato třída, obdobně jako v případě logiky druhého řádu, shodná se třídou jazyků popsatečných logikou $\text{FO}[<]$. Abychom mohli odpovědět na tuto otázku, představíme Ehrenfeuchtovy-Fraïssého hry – nástroj teorie modelů, kterým lze dokázat nevyjádřitelnost daných vlastností v logice prvního řádu s danými predikáty.

Ehrenfeuchtova-Fraïssého hra je hrou dvou hráčů. Dvě struktury považujeme za stejné, pokud splňují právě stejné formule dané logiky. První hráč se snaží poukázat na rozdíl dvou struktur. Naopak druhý hráč se snaží napodobovat jeho tahy tak, aby rozdíly mezi oběma strukturami zakryl, a tím ukázal, že obě struktury jsou vzhledem k dané logice stejné.

Motivací může být například dokázat, že daná struktura splňuje formuli $\forall x \varphi(x)$. Tento úkol se dá chápat v kontextu teorie her – protivník vybere libovolný prvek a této struktury a naším úkolem je ukázat, že platí formule $\varphi(a)$. Pokud formule není splnitelná, protivník má vítěznou strategii. Podobně v případě formule $\exists x \varphi(x)$ jsme na tahu první a naším úkolem je najít takový prvek a , pro který platí $\varphi(a)$. V případě složitějších formulí se tahy obou hráčů v tomto smyslu střídají.

Alternativní charakterizaci podmínky, kterou dvě struktury musí splňovat, aby byly ekvivalentní vzhledem k dané logice, popsal Roland Fraïssé [Fra54]. Vyjádření této podmínky v podobě hry dvou hráčů pak pochází od Andrzeje Ehrenfeuchta [Ehr61]. Tento nástroj zde nebudeme formulovat v plné obecnosti, ale s ohledem na logiku definovanou v této práci. Tedy budeme pracovat jen s konečnými a lineárně uspořádanými strukturami – konečnými slovy. Následující definice, mírně upravená pro naši logiku nad slovy, pochází z [Imm99]. Myšlenky některých důkazů pak pochází z [Str94].

4.1 Definice

Definice 4.1. Hra $\mathcal{G}_m$ nad logikou $\text{FO}[R_1, R_2, \dots, R_k](\Sigma)$ s realizací $\mathcal{R}$ je hrou dvou hráčů, kterou hrají hráči Spoiler a Duplicator s dvojicí slov u a v nad danou abecedou Σ . Hra $\mathcal{G}_m$ se hraje m tahů s m páry *hratelných* kamenů. V každém tahu položí Spoiler jeden z hratelných kamenů na vybranou pozici v libovolném ze slov u a v . Duplicator pak položí odpovídající kámen na vybranou pozici v druhém z těchto slov. Na začátku hry můžou být navíc na obou slovech položeny *fixní kameny*. Vždy však z každého páru fixních kamenů není umístěn ani jeden, nebo je na každém slově jeden.

Označme τ množinu fixních kamenů. Pozici po r -tém tahu je možné vyjádřit jako dvě parciální funkce $\alpha_r: \tau \cup \{x_1, \dots, x_m\} \rightarrow \{1, \dots, |u|\}$ a $\beta_r: \tau \cup \{x_1, \dots, x_m\} \rightarrow \{1, \dots, |v|\}$. Takovou dvojici funkcí (α_r, β_r) nazveme *m -konfigurací*, pokud mají shodné definiční obory, tedy $\text{dom}(\alpha_r) = \text{dom}(\beta_r)$.

Pokud se hra nachází v m -konfiguraci (α_r, β_r) a platí $\alpha_r(x_i) = l, \beta_r(x_i) = l'$, znamená to,

že po r -tém tahu se jeden z kamenů x_i nachází na l -té pozici slova u a druhý z kamenů x_i se nachází na l' -té pozici slova v . Pokud funkce α_r a β_r nejsou pro argument x_i definovány, znamená to, že kameny x_i ještě nebyly umístěny na žádné ze slov.

Zavedeme notaci $\mathcal{G}_m(u, \alpha_0, v, \beta_0)$ pro iniciální konfiguraci hry s m kameny a slovy u, v . V této konfiguraci platí $\text{dom}(\alpha_0) = \text{dom}(\beta_0) = \tau$, a tedy žádný z hratelných kamenů ještě není umístěn na žádném slově. V r -tém tahu vezme Spoiler zatím neumístěný hratelný kámen x_r a umístí ho na vybranou pozici l v libovolném ze slov. Duplicator pak vezme odpovídající druhý kámen x_r a umístí ho na vybranou pozici l' ve druhém ze slov. Pokud Spoiler umístil kámen na první slovo, m -konfigurace po r -tém tahu bude $(\alpha_r, \beta_r) = (\alpha_{r-1}[l/x_r], \beta_{r-1}[l'/x_r])$. V opačném případě, pokud Spoiler umístil kámen na druhé slovo, bude m -konfigurace po r -tém tahu $(\alpha_r, \beta_r) = (\alpha_{r-1}[l'/x_r], \beta_{r-1}[l/x_r])$.

Po každém tahu určuje m -konfigurace relaci $\beta_r \circ \alpha_r^{-1} \subseteq \{1, \dots, |u|\} \times \{1, \dots, |v|\}$, ve které jsou pozice ve slovech u a v , na kterých jsou umístěny odpovídající si kameny. Řekneme, že Duplicator vyhrál r -tý tah hry, právě tehdy, když pro každý kámen $y \in \tau \cup \{x_1, \dots, x_r\}$ platí

$$u[\alpha_r(y)] = v[\beta_r(y)] \quad (4.1)$$

a pro každý j -ární relační symbol R_i a každou j -tici kamenů $(y_1, \dots, y_j) \in (\tau \cup \{x_1, \dots, x_r\})^j$ platí

$$(\alpha_r(y_1), \dots, \alpha_r(y_j)) \in \mathcal{R}(R_i) \Leftrightarrow (\beta_r(y_1), \dots, \beta_r(y_j)) \in \mathcal{R}(R_i). \quad (4.2)$$

Intuitivně – Duplicator vyhrál r -tý tah hry, pokud se mu na každý z r kamenů položených Spoilerem podařilo najít ve druhém slově takovou pozici, aby pozice určené kameny v jednom slově splňovaly právě ty stejné atomické formule, jako splňují pozice určené odpovídajícími kameny ve slově druhém.

Konečně řekneme, že Duplicator vyhrál celou hru, pokud vyhrál každý tah této hry.

Definice 4.2. Pro libovolnou abecedu Σ a logiku $\text{FO}[R_1, R_2, \dots, R_k](\Sigma)$ definujeme pro každé $m \in \mathbb{N}$ binární relaci $\sim_m$.

Pro slova $u, v \in \Sigma^*$, množinu fixních kamenů τ a funkce $\alpha_0: \tau \rightarrow \{1, \dots, |u|\}, \beta_0: \tau \rightarrow \{1, \dots, |v|\}$ klademe $(u, \alpha_0) \sim_m (v, \beta_0)$ právě tehdy, když má Duplicator vítěznou strategii pro m -tahovou hru nad odpovídající logikou začínající v konfiguraci $\mathcal{G}_m(u, \alpha_0, v, \beta_0)$.

Příklad 4.3. Ukážeme příklady Ehrenfeuchtových-Fraïssého her pro různé logiky. Zvolme abecedu $\Sigma = \{a, b\}$.

1. Uvažme logiku $\text{FO}[(\Sigma)]$. Pak platí $(ab, \emptyset) \not\sim_1 (aa, \emptyset)$.

Vítězná strategie pro Spoilera je položit v prvním tahu hry kámen na pozici písmena b ve slově ab . Pak Duplicator zřejmě nemá odpovídající tah, protože ve slově aa neexistuje žádná pozice i , pro kterou by platilo $aa[i] = b$.

2. Uvažme opět logiku $\text{FO}[(\Sigma)]$. Pak platí $(babb, \emptyset) \sim_m (baaa, \emptyset)$ pro každé $m \in \mathbb{N}$.

Buď $m \in \mathbb{N}$ libovolné přirozené číslo. Pak vítězná strategie v m -tahové hře pro Duplicatora je po každém tahu Spoilera na pozici i v libovolném slově $u \in \{babb, baaa\}$ odpovědět tahem na takovou pozici j ve druhém slově, na které je písmeno $u[i]$. Pak je podmínka (4.1) po každém Duplicatorově tahu zřejmě splněna, a proto po m tazích Duplicator zvítězí.

3. Nyní uvažme stejná slova jako v předchozím případě, ale tentokrát budeme hrát hru nad logikou $\text{FO}[=(\Sigma)]$. Pak platí $(babb, \emptyset) \sim_1 (baaa, \emptyset)$, ale $(babb, \emptyset) \not\sim_2 (baaa, \emptyset)$.

Vítězná strategie v jednotahové hře je pro Duplicatora stejná jako v předchozím případě, tedy položit kámen x_1 na pozici se stejným písmenem, na které položil svůj kámen x_1 Spoiler. Pak je jistě splněna podmínka (4.1). Podmínka (4.2) je splněna také, protože jediná možná dvojice kamenů je (x_1, x_1) a pro tu jistě platí $\alpha_1(x_1) = \alpha_1(x_1) \Leftrightarrow \beta_1(x_1) = \beta_1(x_1)$.

Vítězná strategie v dvoutahové hře pro Spoilera je v prvním tahu položit kámen x_1 na jedno z písmen a ve slově $baaa$. A nezávisle na Duplicatorově odpovědi ve druhém tahu položit kámen x_2 na odlišné písmeno a ve slově $baaa$. Pak Duplicator nemá odpověď, pro kterou by platilo $babb[\alpha_2(x_1)] = baaa[\beta_2(x_1)] = a$, $babb[\alpha_2(x_2)] = baaa[\beta_2(x_2)] = a$ a $\beta_2(x_1) \neq \beta_2(x_2)$, protože ve slově $babb$ nejsou dvě různé pozice, na kterých je písmeno a .

4. Uvažme opět logiku $\text{FO}[=(\Sigma)]$. Pak $(baab, \emptyset) \sim_m (abab, \emptyset)$ pro libovolné $m \in \mathbb{N}$.

Bud' $m \in \mathbb{N}$ libovolné přirozené číslo. Pak vítězná strategie v m -tahové hře pro Duplicatora je po každém tahu Spoilera na pozici i v libovolném slově $u \in \{baab, abab\}$ odpovědět tahem na takovou pozici j ve druhém slově, na které je písmeno $u[i]$. Tím zaručí platnost podmínky (4.1). Navíc musí Duplicator tuto pozici zvolit tak, že pokud již na pozici i je nějaký kámen, vybere ve druhém slově takovou pozici, na které se v něm nachází odpovídající kámen. A naopak pokud na pozici i zatím žádný kámen není, musí Duplicator zvolit takovou pozici v druhém slově, na které také zatím žádný kámen není. To je vždy možné, protože slova $baab$ a $abab$ obsahují právě stejný počet všech písmen. Tím zaručí také platnost podmínky (4.2).

5. Nyní uvažme opět slova $baab$ a $abab$, ale tentokrát budeme hrát hru nad logikou $\text{FO}[\prec](\Sigma)$. Pak platí $(baab, \emptyset) \sim_1 (abab, \emptyset)$, ale $(baab, \emptyset) \not\sim_2 (abab, \emptyset)$.

Vítězná strategie v jednotahové hře je pro Duplicatora stejná jako v předchozím případě. Po prvním tahu je jistě splněna podmínka (4.1), protože Duplicator položil kámen na stejné písmeno. Podmínka (4.2) je splněna také, protože jediná možná dvojice kamenů je (x_1, x_1) a pro tu jistě platí $\alpha_1(x_1) < \alpha_1(x_1) \Leftrightarrow \beta_1(x_1) < \beta_1(x_1)$.

Vítězná strategie v dvoutahové hře pro Spoilera je v prvním tahu položit kámen x_1 na první písmeno ve slově $baab$. Duplicator musí jako odpověď položit kámen na nějaké písmeno b ve slově $abab$. V druhém tahu Spoiler položí kámen x_2 na první písmeno ve slově $abab$. Pak Duplicator nemá odpověď. Ve slově $baab$ totiž neexistuje žádná pozice, kam by Duplicator položil kámen x_2 a poté by platilo $1 = \alpha_2(x_1) > \alpha_2(x_2)$. Tato podmínka vyplývá z toho, že musí platit $\beta_2(x_1) > 1$, protože na první pozici slova $abab$ není písmeno b , a $\beta_2(x_2) = 1$, a tedy $\beta_2(x_1) > \beta_2(x_2)$.

Názorně vypadá situace před druhým tahem Duplicatora takto. Podle volby Duplicatora v prvním tahu nastane buď

$$\begin{array}{cc} b & a & a & b \\ x_1 & & & \end{array} \quad \begin{array}{cc} a & b & a & b \\ x_2 & x_1 & & \end{array},$$

nebo

$$\begin{array}{cc} b & a & a & b \\ x_1 & & & \end{array} \quad \begin{array}{cc} a & b & a & b \\ x_2 & & x_1 & \end{array}.$$

V ani jednom případě nemá Duplicator odpověď, která by zaručila stejné pořadí kamenů x_1 a x_2 v obou slovech.

Věta 4.4. Pro libovolnou abecedu Σ , logiku $\text{FO}[R_1, R_2, \dots, R_k](\Sigma)$ a $m \in \mathbb{N}$ je relace $\sim_m$ ekvivalence.

Důkaz. Relace $\sim_m$ je zřejmě reflexivní. Vítěznou strategií Duplicatora je pokládat kameny na stejné pozice jako Spoiler, ale ve druhém slově. Relace $\sim_m$ je také zřejmě symetrická. Pokud má Duplicator vítěznou strategii pro hru začínající v konfiguraci $\mathcal{G}_m(u, \alpha_0, v, \beta_0)$, může vyhrát úplně stejnými tahy hru $\mathcal{G}_m(v, \beta_0, u, \alpha_0)$.

Zbývá ukázat tranzitivitu. Buďte $u, v, w \in \Sigma^*$ libovolná slova nad abecedou Σ , τ libovolná množina fixních kamenů a $\alpha_0: \tau \rightarrow \{1, \dots, |u|\}$, $\beta_0: \tau \rightarrow \{1, \dots, |v|\}$ a $\gamma_0: \tau \rightarrow \{1, \dots, |w|\}$ libovolná zobrazení. Ukážeme, že pokud platí $(u, \alpha_0) \sim_m (v, \beta_0)$ a zároveň $(v, \beta_0) \sim_m (w, \gamma_0)$, pak také musí platit $(u, \alpha_0) \sim_m (w, \gamma_0)$.

Předpokládejme tedy, že Duplicator má vítěznou strategii pro m -tahové hry začínající v konfiguracích $\mathcal{G}_m(u, \alpha_0, v, \beta_0)$ a $\mathcal{G}_m(v, \beta_0, w, \gamma_0)$, a definujeme Duplicatorovu strategii pro m -tahovou hru začínající v konfiguraci $\mathcal{G}_m(u, \alpha_0, w, \gamma_0)$. Intuitivně, Duplicator bude zároveň hrát obě hry $\mathcal{G}_m(u, \alpha_0, v, \beta_0)$ a $\mathcal{G}_m(v, \beta_0, w, \gamma_0)$ tak, aby v obou těchto hrách byly na slově v položené stejné kameny a na stejných pozicích. Tahy Spoilera v jedné z těchto her bude Duplicator simulovat pomocí své vítězné strategie ve druhé hře.

Uvažme tedy pro libovolné $0 \leq r < m$ situaci po r -tém kole m -tahové hry začínající v konfiguraci $\mathcal{G}_m(u, \alpha_0, w, \gamma_0)$. Tedy je odehráno také r kol ve hře začínající v konfiguraci $\mathcal{G}_m(u, \alpha_0, v, \beta_0)$ a r kol ve hře začínající v konfiguraci $\mathcal{G}_m(v, \beta_0, w, \gamma_0)$, a navíc ve slově v jsou v obou těchto hrách položeny kameny na stejných pozicích. Předpokládejme, že Spoiler v tahu $r + 1$ položil kámen x_{r+1} na pozici l ve slově u . Položil-li Spoiler kámen x_{r+1} na slovo w , je argumentace analogická.

Duplicator v tahu $r + 1$ použije svoji strategii ve hře začínající v konfiguraci $\mathcal{G}_m(u, \alpha_0, v, \beta_0)$, aby získal odpovídající pozici l' ve slově v . Poté ve hře začínající v konfiguraci $\mathcal{G}_m(v, \beta_0, w, \gamma_0)$ položí Spoilerův kámen na pozici l' ve slově v a použije svoji vítěznou strategii v této hře, aby získal odpovídající pozici l'' ve slově w . Duplicatorova strategie pak je odpovědět v tahu $r + 1$ hry začínající v konfiguraci $\mathcal{G}_m(u, \alpha_0, w, \gamma_0)$ položením kamene x_{r+1} na pozici l'' ve slově w . Po tomto tahu jsou opět na slově v v obou hrách položeny kameny na stejných pozicích.

Tato Duplicatorova strategie je vítězná. Po každém tahu m -tahové hry začínající v konfiguraci $\mathcal{G}_m(u, \alpha_0, w, \gamma_0)$ jsou podmínky (4.1) a (4.2) splněné, protože jsou splněné v obou hrách a ve slově u jsou v obou hrách položeny kameny na stejných pozicích. $\square$

4.2 Souvislost s logikou prvního řádu

Definice 4.5. Pro libovolnou logiku $\text{FO}[R_1, R_2, \dots, R_k](\Sigma)$ definujeme induktivně vzhledem ke struktuře formule $\varphi \in \text{FO}[R_1, R_2, \dots, R_k](\Sigma)$ její kvantifikátorovou složitost $c(\varphi) \in \mathbb{N}$.

- Je-li $\varphi \equiv Q_a(x)$, pak $c(\varphi) = 0$.
- Je-li $\varphi \equiv R_i(x_1, x_2, \dots, x_{j_i})$, pak $c(\varphi) = 0$.
- Je-li $\varphi \equiv \neg\varphi_1$, pak $c(\varphi) = c(\varphi_1)$.
- Je-li $\varphi \equiv \varphi_1 \wedge \varphi_2$, pak $c(\varphi) = \max(c(\varphi_1), c(\varphi_2))$.
- Je-li $\varphi \equiv \exists x \varphi_1$, pak $c(\varphi) = c(\varphi_1) + 1$.

Protože se v dalším textu budeme potřebovat souhrně vyjadřovat o interpretacích libovolné formule z logiky $\text{FO}[R_1, R_2, \dots, R_k](\Sigma)$, zavedeme označení *interpretace proměnných*, které nám takové vyjadřování usnadní.

Definice 4.6. Pro libovolnou abecedu Σ nazveme *interpretací proměnných* dvojici $(w, \mathcal{J})$, kde $w \in \Sigma^*$ a $\mathcal{J}: \mathcal{V} \rightarrow \{1, \dots, |w|\}$ pro nějakou konečnou podmnožinu proměnných prvního řádu $\mathcal{V} \subseteq \mathcal{X}_1$.

Definice 4.7. Pro libovolnou abecedu Σ , logiku $\text{FO}[R_1, R_2, \dots, R_k](\Sigma)$, její realizaci $\mathcal{R}$ a přirozené číslo $m \in \mathbb{N}$ definujeme binární relaci $\equiv_m$ na interpretacích proměnných. Pro interpretace proměnných $(u, \mathcal{J}_u), (v, \mathcal{J}_v)$ klademe $(u, \mathcal{J}_u) \equiv_m (v, \mathcal{J}_v)$ právě tehdy, když každá formule φ kvantifikátorové složitosti nejvýše m splňuje obě následující podmínky.

1. Dvojice $(u, \mathcal{J}_u)$ je interpretace formule φ právě tehdy, když dvojice $(v, \mathcal{J}_v)$ je interpretace formule φ – tedy, když zobrazení $\mathcal{J}_u$ a $\mathcal{J}_v$ mají stejný definiční obor.
2. Pokud jsou obě dvojice $(u, \mathcal{J}_u)$ a $(v, \mathcal{J}_v)$ interpretacemi formule φ , vztah $(u, \mathcal{J}_u) \models^{\mathcal{R}} \varphi$ platí právě tehdy, když platí vztah $(v, \mathcal{J}_v) \models^{\mathcal{R}} \varphi$.

Relace $\equiv_m$ je zřejmě ekvivalence. Naším cílem bude ukázat, že relace $\sim_m$ a $\equiv_m$ jsou totožné. V důkazu tohoto tvrzení budeme potřebovat fakt, že pro libovolnou logiku a konečnou podmnožinu proměnných prvního řádu $\mathcal{V} \subseteq \mathcal{X}_1$ existuje jen konečně mnoho logicky neekvivalentních formulí této logiky s kvantifikátorovou složitostí c , které obsahují jen proměnné z $\mathcal{V}$. Tento fakt nyní dokážeme spolu s jedním pomocným lemmatem, které v důkazu požadovaného tvrzení použijeme jako bázi indukce.

Lemma 4.8. Pro libovolnou logiku $\text{FO}[R_1, R_2, \dots, R_k](\Sigma)$ a libovolnou konečnou podmnožinu proměnných prvního řádu $\mathcal{V} \subseteq \mathcal{X}_1$ existuje jen konečně mnoho logicky neekvivalentních formulí kvantifikátorové složitosti 0, které obsahují pouze proměnné z $\mathcal{V}$.

Důkaz. Viděli jsme, že přidáním operace disjunkce do logiky FO se nezmění její popisná síla, protože ji lze vyjádřit pomocí konjunkce a negace. Po přidání disjunkce jako nové operace lze každou formuli kvantifikátorové složitosti 0 převést do disjunktivní normální formy a díky idempotenci a komutativitě obou operací odstranit opakované výskyty podformulí. Tudíž každá formule kvantifikátorové složitosti 0 je ekvivalentní s formulí, která se skládá z disjunkcí, ve kterých je každý disjunkt konjunkcí literálů.

Atomických formulí, které obsahují jen proměnné z $\mathcal{V}$, je v logice $\text{FO}[R_1, R_2, \dots, R_k](\Sigma)$ jistě konečný počet – jedna formule $Q_a(x)$ pro každé $a \in \Sigma$ a $x \in \mathcal{V}$ a jedna formule pro každý j -ární predikátový symbol R_i a j -tici proměnných z $\mathcal{V}^j$. Označme tento počet jako m . Pak literálů je $2m$. Jejich konjunkcí, ve kterých se neopakují konjunkty, je $2^{2m} - 1$. A konečně disjunkcí takovýchto konjunkcí, ve kterých se neopakují disjunkt, je tedy $2^{2^{2m}-1} - 1$. $\square$

Lemma 4.9. Pro libovolnou logiku $\text{FO}[R_1, R_2, \dots, R_k](\Sigma)$ a libovolnou konečnou podmnožinu proměnných prvního řádu $\mathcal{V} \subseteq \mathcal{X}_1$ existuje jen konečně mnoho logicky neekvivalentních formulí kvantifikátorové složitosti c , které obsahují pouze proměnné z $\mathcal{V}$.

Důkaz. Indukcí dokážeme, že pro každé $c \in \mathbb{N}$ je jen konečně mnoho logicky neekvivalentních formulí formulí, které mají kvantifikátorovou složitost c .

- Je-li $c = 0$, stačí použít předchozí lemma.

- Je-li $c > 0$, pak lze každou takovou formuli přepsat jako disjunkci konjunkcí formulí ve tvaru $\exists x \varphi_1$, $\neg \exists x \varphi_1$ nebo φ_1 , kde $c(\varphi_1) < c$. Z indukčního předpokladu je takových logicky neekvivalentních formulí φ_1 konečný počet. Označme ho jako p . Protože víme, že $x \in \mathcal{V}$, je logicky neekvivalentních formulí ve tvaru $\exists x \varphi_1$, $\neg \exists x \varphi_1$ nebo φ_1 nejvýše $3p|\mathcal{V}|$. Proto logicky neekvivalentních konjunkcí takovýchto formulí je nejvýše $2^{3p|\mathcal{V}|} - 1$. A proto je logicky neekvivalentních disjunkcí takovýchto konjunkcí nejvýše $2^{2^{3p|\mathcal{V}|} - 1} - 1$. $\square$

Věta 4.10. Pro libovolnou abecedu Σ , logiku $\text{FO}[R_1, R_2, \dots, R_k](\Sigma)$, interpretace proměnných $(u, \mathcal{I}_u)$, $(v, \mathcal{I}_v)$ a přirozené číslo $m \in \mathbb{N}$ jsou následující dvě tvrzení ekvivalentní.

1. $(u, \mathcal{I}_u) \sim_m (v, \mathcal{I}_v)$
2. $(u, \mathcal{I}_u) \equiv_m (v, \mathcal{I}_v)$

Důkaz. Implikaci (1) $\Rightarrow$ (2) dokážeme indukcí vzhledem k m .

- Je-li $m = 0$, pak předpokládáme $(u, \mathcal{I}_u) \sim_0 (v, \mathcal{I}_v)$. Chceme ukázat, že platí vztah $(u, \mathcal{I}_u) \equiv_0 (v, \mathcal{I}_v)$ – tedy, že pro každou formuli φ kvantifikátorové složitosti 0 je dvojice $(u, \mathcal{I}_u)$ interpretací formule φ právě tehdy, když dvojice $(v, \mathcal{I}_v)$ je interpretací formule φ , a když jsou obě tyto dvojice interpretacemi formule φ , pak že vztah $(u, \mathcal{I}_u) \models \varphi$ platí právě tehdy, když platí vztah $(v, \mathcal{I}_v) \models \varphi$.

Z předpokladu $(u, \mathcal{I}_u) \sim_0 (v, \mathcal{I}_v)$ vyplývá, že Duplicator má vítěznou strategii pro hru začínající v konfiguraci $\mathcal{G}_0(u, \mathcal{I}_u, v, \mathcal{I}_v)$. Z definice Ehrenfeuchtovy-Fraïsseho hry platí, že funkce $\mathcal{I}_u$ a $\mathcal{I}_v$ mají stejný definiční obor, a tudíž jsou obě dvojice $(u, \mathcal{I}_u)$, $(v, \mathcal{I}_v)$ interpretacemi formule φ , nebo není interpretací formule φ žádná z nich. Navíc z podmínek (4.1) a (4.2) z téže definice vyplývá, že jsou-li interpretacemi libovolné atomické formule obě z dvojic $(u, \mathcal{I}_u)$, $(v, \mathcal{I}_v)$, pak obě tuto formuli splňují, nebo ji obě nesplňují. Tím pádem také obě splňují formuli φ , nebo ji obě nesplňují, protože formule kvantifikátorové složitosti 0 jsou právě formule vzniklé konjunkcemi a negacemi z atomických formulí.

- Je-li $m > 0$, pak předpokládáme $(u, \mathcal{I}_u) \sim_m (v, \mathcal{I}_v)$. Z definice Ehrenfeuchtovy-Fraïsseho hry tedy opět platí, že funkce $\mathcal{I}_u$ a $\mathcal{I}_v$ mají stejný definiční obor, a tudíž jsou interpretacemi libovolné formule obě dvojice $(u, \mathcal{I}_u)$, $(v, \mathcal{I}_v)$, nebo není interpretací žádná z nich. Předpokládejme dále pro důkaz sporem, že existuje taková formule φ kvantifikátorové složitosti m , která od sebe odlišuje jejich interpretace $(u, \mathcal{I}_u)$ a $(v, \mathcal{I}_v)$, a tedy platí $(u, \mathcal{I}_u) \models \varphi$ a $(v, \mathcal{I}_v) \not\models \varphi$, nebo $(u, \mathcal{I}_u) \not\models \varphi$ a $(v, \mathcal{I}_v) \models \varphi$. Pokud by byla taková formule ve tvaru $\varphi \equiv \psi_1 \wedge \psi_2$, platí tato podmínka pro jednu z formulí ψ_1 , ψ_2 . Podobně, pokud by byla taková formule ve tvaru $\varphi \equiv \neg \psi_1$, platí tato podmínka pro formuli ψ_1 . Bez újmy na obecnosti tedy můžeme předpokládat, že taková formule je ve tvaru $\varphi \equiv \exists x \psi$.

Bez újmy na obecnosti dále předpokládejme, že nastala situace, ve které $(u, \mathcal{I}_u) \models \varphi$ a $(v, \mathcal{I}_v) \not\models \varphi$. Protože $\varphi \equiv \exists x \psi$, platí $|u| \geq 1$ a existuje takové $i \in \{1, \dots, |u|\}$, pro které $(u, \mathcal{I}_u[i/x]) \models \psi$. Uvažme hru začínající v konfiguraci $\mathcal{G}_m(u, \mathcal{I}_u, v, \mathcal{I}_v)$, ve které Spoiler položí v prvním tahu kámen na pozici i . Protože předpokládáme $(u, \mathcal{I}_u) \sim_m (v, \mathcal{I}_v)$, má Duplicator vítěznou strategii pro m -tahovou hru začínající tímto tahem. Označme j pozici ve slově v , na kterou Duplicator položí odpovídající kámen. Když přidáme nový

fixní kámen x , který reprezentuje stav po prvním tahu, a první tah vynecháme, dostaneme $(m-1)$ -tahovou hru začínající v konfiguraci $\mathcal{G}_{m-1}(u, \mathcal{J}_u[i/x], v, \mathcal{J}_v[j/x])$. Pro tuto hru má jistě Duplicator též vítěznou strategii. Tedy platí $(u, \mathcal{J}_u[i/x]) \sim_{m-1} (v, \mathcal{J}_v[j/x])$. Z indukčního předpokladu dostáváme vztah $(u, \mathcal{J}_u[i/x]) \equiv_{m-1} (v, \mathcal{J}_v[j/x])$. Protože formule ψ má kvantifikátorovou složitost $m-1$, musí platit $(u, \mathcal{J}_u[i/x]) \models \psi$ právě tehdy, když platí $(v, \mathcal{J}_v[j/x]) \models \psi$. A tudíž, protože $\varphi \equiv \exists x \psi$, platí $(v, \mathcal{J}_v) \models \varphi$, což je spor.

Platnost opačné implikace $(2) \Rightarrow (1)$ dokážeme též indukcí vzhledem k m .

- Je-li $m = 0$, pak předpokládáme $(u, \mathcal{J}_u) \equiv_0 (v, \mathcal{J}_v)$. Tedy funkce $\mathcal{J}_u$ a $\mathcal{J}_v$ mají stejný definiční obor, protože pro každou formuli jsou jejími interpretacemi obě dvojice $(u, \mathcal{J}_u)$, $(v, \mathcal{J}_v)$, nebo ani jedna. Tudíž $(\mathcal{J}_u, \mathcal{J}_v)$ je skutečně 0-konfigurace 0-tahové Ehrenfeuchtovy-Fraïsseho hry. Navíc dvojice $(u, \mathcal{J}_u)$, $(v, \mathcal{J}_v)$ splňují právě stejné formule kvantifikátorové složitosti 0, jejichž jsou obě interpretacemi. Mezi takové formule patří zejména atomické formule. Tudíž podmínky (4.1) a (4.2) jsou splněny, a proto má Duplicator vítěznou strategii v 0-tahové hře.
- Je-li $m > 0$, pak předpokládáme $(u, \mathcal{J}_u) \equiv_m (v, \mathcal{J}_v)$. Pro důkaz sporem předpokládejme, že Spoiler má vítěznou strategii pro m -tahovou hru začínající v konfiguraci $\mathcal{G}_m(u, \mathcal{J}_u, v, \mathcal{J}_v)$, a tedy $(u, \mathcal{J}_u) \not\sim_m (v, \mathcal{J}_v)$. Buď první tah Spoileru v této strategii položení kamene x_1 na pozici i ve slově u . Libovolná odpověď Duplicatoru musí být položení druhého kamene x_1 na pozici j ve slově v . Pokud je Spoilerův tah ve slově v , situace je symetrická.

Opět můžeme ze stavu po prvním tahu odvodit $(m-1)$ -tahovou hru začínající v konfiguraci $\mathcal{G}_{m-1}(u, \mathcal{J}_u[i/x_1], v, \mathcal{J}_v[j/x_1])$. Tudíž Spoiler má také vítěznou strategii pro $(m-1)$ -tahovou hru, a tedy $(u, \mathcal{J}_u[i/x_1]) \not\sim_{m-1} (v, \mathcal{J}_v[j/x_1])$. Z indukčního předpokladu pak vyplývá $(u, \mathcal{J}_u[i/x_1]) \not\equiv_{m-1} (v, \mathcal{J}_v[j/x_1])$. Podle Lemmatu 4.9 je logicky neekvivalentních formulí kvantifikátorové složitosti nejvýše $m-1$, jejichž jsou obě dvojice $(u, \mathcal{J}_u[i/x_1])$ a $(v, \mathcal{J}_v[j/x_1])$ interpretace, jen konečně mnoho. Označme tedy jako A množinu všech takových formulí. Uvažme formuli ψ definovanou jako

$$\psi \equiv \bigwedge_{\substack{\varphi \in A \\ (u, \mathcal{J}_u[i/x_1]) \models \varphi}} \varphi.$$

Pak jistě platí $(u, \mathcal{J}_u[i/x_1]) \models \psi$ a $(v, \mathcal{J}_v[j/x_1]) \not\models \psi$. Tedy z definice sémantiky také platí $(u, \mathcal{J}_u) \models \exists x_1 \psi$. Protože jsme jako j uvažovali libovolnou Duplicatorovu odpověď, platí také $(v, \mathcal{J}_v) \models \exists x_1 \psi$. Ale to je spor s předpokladem $(u, \mathcal{J}_u) \equiv_m (v, \mathcal{J}_v)$, protože formule $\exists x_1 \psi$ má kvantifikátorovou složitost m . $\square$

Příklad 4.11. Ukážeme, co znamenají výsledky předvedené v Příkladu 4.3 v kontextu právě dokázané věty. Opět uvažujme abecedu $\Sigma = \{a, b\}$.

1. Ukázali jsme, že pro logiku $\text{FO}[\exists](\Sigma)$ platí $(ab, \emptyset) \not\sim_1 (aa, \emptyset)$, a tudíž musí existovat formule $\varphi \in \text{FO}[\exists](\Sigma)$ kvantifikátorové složitosti 1 taková, že $ab \models \varphi$ a $aa \not\models \varphi$. Takovou formulí je

$$\varphi \equiv \exists x Q_b(x).$$

2. Ukázali jsme, že pro logiku $\text{FO}[\equiv](\Sigma)$ platí $(babb, \emptyset) \sim_m (baaa, \emptyset)$ pro libovolné $m \in \mathbb{N}$. Tudíž pro každou uzavřenou formuli $\varphi \in \text{FO}[\equiv](\Sigma)$ platí $babb \models \varphi \Leftrightarrow baaa \models \varphi$. Z toho vyplývá, že každý $\text{FO}[\equiv]$ -definovatelný jazyk obsahující slovo $babb$ musí obsahovat také slovo $baaa$ a naopak.
3. Ukázali jsme, že pro logiku $\text{FO}[=](\Sigma)$ platí $(babb, \emptyset) \sim_1 (baaa, \emptyset)$ a $(babb, \emptyset) \not\sim_2 (baaa, \emptyset)$. Tudíž slova $babb$ a $baaa$ od sebe nejsou rozlišitelná žádnou uzavřenou formulí logiky $\text{FO}[=](\Sigma)$ kvantifikátorové složitosti 1, ale jsou rozlišitelná nějakou uzavřenou formulí $\varphi \in \text{FO}[=](\Sigma)$ kvantifikátorové složitosti 2. Takovou formulí je

$$\varphi \equiv \exists x \exists y (Q_a(x) \wedge Q_a(y) \wedge \neg(x = y)).$$

4. Ukázali jsme, že pro logiku $\text{FO}[=](\Sigma)$ platí $(baab, \emptyset) \sim_m (abab, \emptyset)$ pro libovolné $m \in \mathbb{N}$. Tudíž pro každou uzavřenou formuli $\varphi \in \text{FO}[=](\Sigma)$ platí $baab \models \varphi \Leftrightarrow abab \models \varphi$. Z toho vyplývá, že každý $\text{FO}[=]$ -definovatelný jazyk obsahující slovo $baab$ musí obsahovat také slovo $abab$ a naopak.
5. Ukázali jsme, že pro logiku $\text{FO}[<](\Sigma)$ platí $(baab, \emptyset) \sim_1 (abab, \emptyset)$ a $(baab, \emptyset) \not\sim_2 (abab, \emptyset)$. Tudíž slova $baab$ a $abab$ od sebe nejsou rozlišitelná žádnou uzavřenou formulí logiky $\text{FO}[<](\Sigma)$ kvantifikátorové složitosti 1, ale jsou rozlišitelná nějakou uzavřenou formulí $\varphi \in \text{FO}[<](\Sigma)$ kvantifikátorové složitosti 2. Takovou formulí je

$$\varphi \equiv \exists x (Q_b(x) \wedge \forall y (Q_a(y) \Rightarrow x < y)).$$

4.3 Použití na logiku $\text{FO}[<]$

Nyní ukážeme, jak lze právě dokázaný výsledek o souvislosti Ehrenfeuchtových-Fraïssého her s logikou prvního řádu použít k důkazu toho, že existují jazyky, které jsou $\text{MSO}[<]$ -definovatelné, ale nejsou $\text{FO}[<]$ -definovatelné.

Lemma 4.12. *Pro abecedu $\Sigma = \{a\}$, logiku $\text{FO}[<](\Sigma)$ a libovolné přirozené číslo $m \in \mathbb{N}$ platí $(a^k, \emptyset) \sim_m (a^{k+1}, \emptyset)$ pro každé $k \geq 2^m$.*

Důkaz. Tvrzení dokážeme indukcí vzhledem k m .

- Je-li $m = 0$, pak jistě platí $(a^k, \emptyset) \sim_0 (a^{k+1}, \emptyset)$, protože $(\emptyset, \emptyset)$ je vítězná 0-konfigurace pro hráče Duplicator.
- Je-li $m > 0$, pak předpokládáme $(a^k, \emptyset) \sim_{m-1} (a^{k+1}, \emptyset)$. Ukážeme, že Duplicator má vítěznou strategii i pro m -tahovou hru. Označme $u = a^k$ a $v = a^{k+1}$.
 - Jestliže Spoiler v prvním tahu položí kámen na pozici $i \leq 2^{m-1}$ v libovolném slově, pak Duplicator odpoví položením odpovídajícího kamene na i -tou pozici ve druhém slově.

Tím se obě slova rozpadnou na dvě části, které lze znázornit jako

$$\begin{aligned} u &= a^i \cdot a^{k-i}, \\ v &= a^i \cdot a^{k-i+1}. \end{aligned}$$

Strategii pro každou z obou částí lze řešit zvlášť, protože střídáním částí Spoiler nemůže získat žádnou výhodu – pro kámen x_1 z levé části a kámen x_2 z pravé části bude v každém tahu r vždy platit $\alpha_r(x_1) < \alpha_r(x_2)$ a zároveň $\beta(x_1) < \beta(x_2)$. Pokud Spoiler položí následující kámen na pozici $j \leq i$ v libovolném slově, Duplicator odpoví položením odpovídajícího kamene na j -tou pozici ve druhém slově. Touto strategií Duplicator jistě zvítězí, protože levé části obou slov mají stejnou délku.

Zbývá určit Duplicatorovu strategii pro případ, ve kterém Spoiler položí kámen na pozici $j > i$ v libovolném slově. Z indukčního předpokladu vyplývá $(a^{k'}, \emptyset) \sim_{m-1} (a^{k'+1}, \emptyset)$ pro každé $k' \geq 2^{m-1}$. Protože $k \geq 2^m$ a $i \leq 2^{m-1}$, jistě platí $k - i \geq 2^{m-1}$. Celkově tedy platí $(a^{k-i}, \emptyset) \sim_{m-1} (a^{k-i+1}, \emptyset)$, a proto má Duplicator vítěznou strategii pro $(m-1)$ -tahovou hru nad slovy a^{k-i} a a^{k-i+1} . Pokud Spoiler tedy umístí kámen na pozici $j > i$, stačí Duplicatorovi použít tuto strategii.

- Jestliže Spoiler v prvním tahu položí kámen na pozici $i > 2^{m-1}$, je postup Duplicatora analogický. $\square$

Věta 4.13. Jazyk $L = \{aa\}^*$ nad abecedou $\Sigma = \{a\}$ není FO[<]-definovatelný.

Důkaz. Předpokládejme pro důkaz sporem opak – tedy, že existuje uzavřená formule $\varphi \in \text{FO}[\leq](\Sigma)$ taková, že $L(\varphi) = L$. Označme $m = c(\varphi)$ její kvantifikátorovou složitost. Z předchozího lemmatu vyplývá $(a^{2^m}, \emptyset) \sim_m (a^{2^m+1}, \emptyset)$. Pak ale díky Větě 4.10 platí $a^{2^m} \models \varphi$ právě tehdy, když platí $a^{2^m+1} \models \varphi$, což je spor. $\square$

Důsledek 4.14. Jazyk $L = \{aa\}^*$ nad abecedou $\Sigma = \{a\}$ není star-free.

Důkaz. Přímý důsledek předchozí věty a Lemmatu 3.13. $\square$

Důsledek 4.15. Třída FO[<]-definovatelných jazyků je vlastní podtřídou třídy MSO[<]-definovatelných jazyků.

Důkaz. Přímý důsledek předchozí věty, Věty 2.4 a faktu, že jazyk $L = \{aa\}^*$ je regulární. $\square$

Kapitola 5

Lokálně prahově testovatelné jazyky a logika $\text{FO}[+1]$

Stejně, jako jsme v Kapitole 3 ukázali, že třída $\text{FO}[<]$ -definovatelných jazyků lze charakterizovat i bez pomoci logiky, ukážeme v této kapitole alternativní charakterizaci třídy $\text{FO}[+1]$ -definovatelných jazyků. Cílem této kapitoly je ukázat, že třída $\text{FO}[+1]$ -definovatelných jazyků odpovídá přesně třídě *lokálně prahově testovatelných jazyků*.

Lokálně prahově testovatelné jazyky jsou zobecněním konceptu lokálně testovatelného jazyka, který nezávisle na sobě zavedli McNaughton a Papert v [MP71] a Brzozowski a Simon v [BS73]. Lokálně testovatelné jazyky jsou takové jazyky, pro které existuje taková konstanta k , že příslušnost slova w do jazyka lze určit jen podle toho, jaká neprázdná podslova délky nejvýše k slovo w obsahuje. Pro takové jazyky existuje algoritmus pro rozhodování příslušnosti do jazyka, který čte slovo zleva doprava pomocí čtecího okna délky k a ukládá si, na která neprázdná podslova narazil.

Většina definic a důkazů je převzatá z [Str94], upravená pro notaci použitou v této práci a v případě potřeby rozšířená.

5.1 Lokálně prahově testovatelné jazyky

Lokálně prahově testovatelné jazyky jako zobecnění pojmu lokálně testovatelného jazyka zavedli Beauquier a Pin v [BP89]. Lokálně prahově testovatelné jazyky jsou takové jazyky, pro které existují takové konstanty k a r , že příslušnost slova w do jazyka lze určit jen podle toho, kolik různých výskytů neprázdných podslov délky nejvýše k slovo w obsahuje. Přitom počet těchto podslov počítáme nejvýše do r . Zřejmě každý lokálně testovatelný jazyk je také lokálně prahově testovatelný pro $r = 1$.

Definice 5.1. Pro libovolnou abecedu Σ a slova $u \in \Sigma^*$, $v \in \Sigma^+$ označíme $\text{Fact}(u, v) \in \mathbb{N}$ číslo udávající počet různých výskytů slova v jako podslova slova u .

Příklad 5.2. Například platí

$$\begin{aligned}\text{Fact}(\text{abbabbaa}, ab) &= 2, \\ \text{Fact}(\text{baabaabaaa}, aa) &= 4.\end{aligned}$$

Definice 5.3. Pro libovolnou abecedu Σ a libovolná celá čísla $k, r > 0$ definujeme binární relaci $\approx_r^k \subseteq \Sigma^* \times \Sigma^*$.

Pro slova $u, v \in \Sigma^*$ klademe $u \approx_r^k v$ právě tehdy, když $|u| < k$ a $u = v$, nebo $|u| \geq k$ a jsou splněny všechny tři následující podmínky.

1. Slova u a v mají stejný prefix délky $k - 1$.
2. Slova u a v mají stejný sufix délky $k - 1$.

3. Pro každé slovo $w \in \Sigma^+$ takové, že $|w| \leq k$, platí buď

$$\text{Fact}(u, w) = \text{Fact}(v, w) < r,$$

nebo platí $\text{Fact}(u, w) \geq r$ a zároveň $\text{Fact}(v, w) \geq r$.

Definice 5.4. Pro libovolnou abecedu Σ a přirozené číslo k označme $\Sigma^{\leq k} \subseteq \Sigma^*$ jazyk všech slov nad abecedou Σ délky nejvýše k . Tedy

$$\Sigma^{\leq k} \stackrel{\text{def}}{=} \bigcup_{0 \leq i \leq k} \Sigma^i.$$

Věta 5.5. Pro libovolnou abecedu Σ a celá čísla $k, r > 0$ je relace $\approx_r^k$ ekvivalence s konečným indexem.

Důkaz. Pro každou abecedu Σ a celá čísla $k, r > 0$ je relace $\approx_r^k$ zřejmě ekvivalence.

Pro každé slovo $u \in \Sigma^*$ takové, že $|u| < k$, je třída ekvivalence obsahující u jednoprvková. Pro každé slovo $u \in \Sigma^*$ takové, že $|u| \geq k$, je třída ekvivalence obsahující slovo u jednoznačně určena trojicí (prefix, suffix, factors), kde prefix je prefix slova u délky $k - 1$, suffix je sufix slova u délky $k - 1$ a factors: $(\Sigma^{\leq k} \setminus \{\varepsilon\}) \rightarrow \mathbb{N}$ je zobrazení určené pro každé $w \in \Sigma^{\leq k} \setminus \{\varepsilon\}$ předpisem

$$\text{factors}(w) = \min(\text{Fact}(u, w), r).$$

Slov délky nejvýše $k - 1$ i takovýchto trojic je jistě konečně mnoho, a proto má relace $\approx_r^k$ konečný index. $\square$

Definice 5.6. Buď Σ libovolná abeceda. Jazyk $L \subseteq \Sigma^*$ nazveme *lokálně prahově testovatelný*, pokud existují taková celá čísla $k, r > 0$, pro která je jazyk L sjednocení nějakých tříd ekvivalence $\approx_r^k$.

Příklad 5.7. Nechť $\Sigma = \{a, b\}$. Jazyk

$$L = \{w \in \Sigma^+ \mid w \text{ začíná a končí písmenem } a \text{ a } w \text{ obsahuje nejvýše jedno písmeno } b\}$$

je lokálně prahově testovatelný, protože je sjednocením nějakých tříd ekvivalence $\approx_2^2$. Platí totiž

$$L = \{a\} \cup \{aa\} \cup \{aa\}\{a\}^+ \cup \{aba\} \cup \{aaba, abaa\} \cup \\ (\{aa\}\{a\}^+\{ba\} \cup \{ab\}\{a\}^+\{aa\} \cup \{a\}\{a\}^+\{b\}\{a\}\{a\}^+).$$

Ukážeme, že každý z těchto jazyků je třída ekvivalence $\approx_2^2$.

- Podle definice je slovo a ekvivalentní jen samo se sebou, protože $|a| < k = 2$.
- Jazyk $\{aa\}$ je třída ekvivalence určená trojicí

$$(\text{prefix}, \text{suffix}, \text{factors}) = (a, a, \{a \mapsto 2, b \mapsto 0, aa \mapsto 1, ab \mapsto 0, ba \mapsto 0, bb \mapsto 0\}).$$

- Jazyk $\{aa\}\{a\}^+$ je třída ekvivalence určená trojicí

$$(\text{prefix}, \text{suffix}, \text{factors}) = (a, a, \{a \mapsto 2, b \mapsto 0, aa \mapsto 2, ab \mapsto 0, ba \mapsto 0, bb \mapsto 0\}).$$

- Jazyk $\{aba\}$ je třída ekvivalence určená trojicí

$$(\text{prefix}, \text{suffix}, \text{factors}) = (a, a, \{a \mapsto 2, b \mapsto 1, aa \mapsto 0, ab \mapsto 1, ba \mapsto 1, bb \mapsto 0\}).$$

- Jazyk $\{aaba, abaa\}$ je třída ekvivalence určená trojicí

$$(\text{prefix}, \text{suffix}, \text{factors}) = (a, a, \{a \mapsto 2, b \mapsto 1, aa \mapsto 1, ab \mapsto 1, ba \mapsto 1, bb \mapsto 0\}).$$

- Jazyk $\{aa\}\{a\}^+\{ba\} \cup \{ab\}\{a\}^+\{aa\} \cup \{a\}\{a\}^+\{b\}\{a\}\{a\}^+$ je třída ekvivalence určená trojicí

$$(\text{prefix}, \text{suffix}, \text{factors}) = (a, a, \{a \mapsto 2, b \mapsto 1, aa \mapsto 2, ab \mapsto 1, ba \mapsto 1, bb \mapsto 0\}).$$

Poznámka 5.8. Jazyk L z předchozího příkladu však není lokálně testovatelný. Pro každé $k > 0$ totiž platí $a^kba^k \approx_1^k a^kba^kba^k$, protože slova a^kba^k a $a^kba^kba^k$ mají stejný prefix délky $k - 1$, stejný sufix délky $k - 1$ a obsahují právě stejná podslova délky nejvýše k . Takže jazyk L nemůže být sjednocením tříd ekvivalence $\approx_1^k$ pro žádné $k > 0$.

Příklad 5.9. Ukážeme, že každý konečný jazyk je lokálně prahově testovatelný.

Buď Σ libovolná abeceda a $L \subseteq \Sigma^*$ libovolný konečný jazyk nad touto abecedou. Zvolme $k = \max\{|w| \mid w \in L\}$. Pak jazyk L je sjednocením nějakých tříd ekvivalence $\approx_1^{k+1}$, protože každé slovo délky menší než $k + 1$ určuje vlastní jednoprvkovou třídu ekvivalence.

Poznámka 5.10. Z předchozího důkazu vyplývá, že každý konečný jazyk je také lokálně testovatelný.

Příklad 5.11. Nechť $\Sigma = \{a, b, c\}$. Jazyk $L = \{a\}^*\{b\}\{a\}^*\{c\}\{a\}^* \subseteq \Sigma^*$ není lokálně prahově testovatelný.

Předpokládejme pro důkaz sporem, že jazyk L lokálně prahově testovatelný je. Tedy je sjednocením tříd ekvivalence $\approx_r^k$ pro nějaká celá čísla $k, r > 0$. Pro každé r ale platí

$$a^kba^kca^k \approx_r^k a^kca^kba^k,$$

protože obě slova mají jistě délku alespoň k , mají stejný prefix délky $k - 1$, stejný sufix délky $k - 1$ a každé neprázdné slovo délky nejvýše k se v obou slovech $a^kba^kca^k$ a $a^kca^kba^k$ vyskytuje jako podslovo stejněkrát.

Tedy dostáváme $a^kca^kba^k \in L$, což je spor s definicí jazyka L .

5.2 Popisná síla logiky FO[+1]

Jak bylo naznačeno v úvodu této kapitoly, nyní dokážeme, že libovolný jazyk je lokálně prahově testovatelný právě tehdy, když je FO[+1]-definovatelný.

Věta 5.12. Nechť Σ je libovolná abeceda. Každý lokálně prahově testovatelný jazyk $L \subseteq \Sigma^*$ je FO[+1]-definovatelný.

Důkaz. Buď Σ libovolná abeceda a $L \subseteq \Sigma^*$ lokálně prahově testovatelný jazyk. Tedy existují celá čísla $k, r > 0$ taková, že jazyk L je sjednocením nějakých tříd ekvivalence $\approx_r^k$. Podle Věty 5.5 má ekvivalence $\approx_r^k$ konečný index, a tudíž je jazyk L sjednocením konečně

mnoha tříd ekvivalence. Stačí ukázat, že každá třída ekvivalence $\approx_r^k$ je definovatelná v logice FO[+1], protože potom formuli definující jazyk L získáme jako disjunkci všech formulí definujících třídy ekvivalence, jejichž je jazyk L sjednocením.

Nejprve ukážeme, že logikou FO[+1] lze vyjádřit predikáty, které odpovídají vlastnostem „dané slovo má prefix u “, „dané slovo má sufix u “, „dané slovo je stejné jako slovo u “ a „ v daném slově se vyskytuje alespoň r -krát podslovo u “.

Pro libovolné slovo $u \in \Sigma^*$ můžeme sestavit formuli prefix_u takovou, že pro libovolné slovo $w \in \Sigma^*$ bude platit $w \models \text{prefix}_u$ právě tehdy, když slovo u je prefixem slova w . Označme $u = u_1 \dots u_n$. Pokud $u = \varepsilon$, je slovo u prefixem každého slova, a tedy $\text{prefix}_u \equiv \forall x (x = x)$. V opačném případě je formule prefix_u

$$\text{prefix}_u \equiv \exists x_1 \exists x_2 \dots \exists x_n \left(\bigwedge_{1 \leq i \leq n} (Q_{u_i}(x_i)) \wedge \text{first}(x_1) \wedge \bigwedge_{1 \leq i \leq n-1} (x_{i+1} = x_i + 1) \right).$$

Analogicky definujeme pro každé $u \in \Sigma^*$ formuli sufix_u takovou, že pro libovolné slovo $w \in \Sigma^*$ bude platit $w \models \text{sufix}_u$ právě tehdy, když je slovo u sufixem slova w .

Podobně lze definovat formuli equal_u , jejímž modelem bude jen slovo u . Pokud $u = \varepsilon$, je takovou formulí $\text{equal}_u \equiv \forall x (\neg(x = x))$. V opačném případě je takovou formulí

$$\text{equal}_u \equiv \exists x_1 \exists x_2 \dots \exists x_n \left(\bigwedge_{1 \leq i \leq n} (Q_{u_i}(x_i)) \wedge \text{first}(x_1) \wedge \text{last}(x_n) \wedge \bigwedge_{1 \leq i \leq n-1} (x_{i+1} = x_i + 1) \right).$$

Pro libovolné slovo $u \in \Sigma^+$ a celé číslo $r > 0$ můžeme také sestavit formuli $\text{fact}_u^{\geq r}$ takovou, že pro libovolné slovo $w \in \Sigma^*$ bude platit $w \models \text{fact}_u^{\geq r}$ právě tehdy, když $\text{Fact}(w, u) \geq r$. Označme opět $u = u_1 \dots u_n$. Pak taková formule $\text{fact}_u^{\geq r}$ je

$$\begin{aligned} \text{fact}_u^{\geq r} \equiv \exists x_{1,1} \exists x_{1,2} \dots \exists x_{1,n} \exists x_{2,1} \dots \exists x_{r,n} \left(\bigwedge_{1 \leq i \leq r} \bigwedge_{1 \leq j \leq n} (Q_{u_j}(x_{i,j})) \wedge \right. \\ \bigwedge_{1 \leq i < j \leq r} (\neg(x_{i,1} = x_{j,1})) \wedge \\ \left. \bigwedge_{1 \leq i \leq r} \bigwedge_{1 \leq j \leq n-1} (x_{i,j+1} = x_{i,j} + 1) \right). \end{aligned}$$

V této formulí však nemůžeme použít pro $x = y$ námi definovanou syntaktickou zkratku $\neg(x < y) \wedge \neg(y < x)$, protože v logice FO[+1] není predikát $<$. Místo toho $x = y$ definujeme jako $\forall z (z = x + 1 \Leftrightarrow z = y + 1)$. Proměnná $x_{i,j}$ ve formulí $\text{fact}_u^{\geq r}$ intuitivně značí j -tou pozici v i -tém výskytu podslova u . Vyžadujeme, aby na této pozici byl opravdu i -tý znak slova u , aby každé z podslov začínalo na jiné pozici a aby na sebe jednotlivé znaky každého výskytu podslova navazovaly. S použitím této formule můžeme sestavit formule $\text{fact}_u^{< k} \equiv \neg \text{fact}_u^{\geq k}$ a $\text{fact}_u^{=k} \equiv \text{fact}_u^{\geq k} \wedge \text{fact}_u^{< k+1}$. Zamýšlený význam těchto formulí je zřejmý.

Budeme chtít vyjádřit formulí třetí podmínku z definice relace $\approx_r^k$. Proto definujeme pro každá slova $u \in \Sigma^*$, $w \in \Sigma^+$ a celé číslo $r > 0$ formuli $\text{satisfiesFactors}_{u,w}^r$ takovou, že libovolné slovo $v \in \Sigma^*$ je modelem formule $\text{satisfiesFactors}_{u,w}^r$ právě tehdy, když platí $\text{Fact}(u, w) = \text{Fact}(v, w) < r$, nebo $\text{Fact}(u, w) \geq r$ a zároveň $\text{Fact}(v, w) \geq r$. Pokud platí $\text{Fact}(u, v) < r$, takovou formulí je

$$\text{satisfiesFactors}_{u,w}^r \equiv \text{fact}_w^{=\text{Fact}(u,w)}.$$

V opačném případě je takovou formulí

$$\text{satisfiesFactors}_{u,w}^r \equiv \text{fact}_w^{\geq r}.$$

Uvažme třídu ekvivalence $\approx_r^k$ obsahující slovo $u = u_1 \dots u_n \in \Sigma^*$. Pak pokud $|u| < k$, obsahuje tato třída jen slovo u , a lze ji proto popsat formulí $\varphi \equiv \text{equal}_u$. Pokud naopak platí $|u| \geq k$, je hledanou formulí

$$\varphi \equiv \text{prefix}_{u_1 \dots u_{k-1}} \wedge \text{suffix}_{u_{n-k+2} \dots u_n} \wedge \bigwedge_{w \in \Sigma^{\leq k} \setminus \{\varepsilon\}} (\text{satisfiesFactors}_{u,w}^r). \quad \square$$

Opět budeme chtít dokázat i druhou implikaci, podle které je každý FO[+1]-definovatelný jazyk také lokálně prahově testovatelný. K tomu nejprve definujeme některé potřebné pojmy.

Definice 5.13. Nechť Σ je libovolná abeceda, $w = w_1 \dots w_n \in \Sigma^*$ slovo nad touto abecedou a $i, j \in \mathbb{N}$ přirozená čísla splňující podmínku $1 \leq i \leq j \leq n$. Pak zápisem $w[i, j]$ rozumíme výskyt podslova $w_i w_{i+1} \dots w_j$.

Definice 5.14. Nechť Σ je libovolná abeceda a $w \in \Sigma^*$ slovo nad touto abecedou. Dva výskyty podslov $u = w[i, j]$ a $v = w[k, l]$ nazveme *oddělené* pokud $k > j + 1$ nebo $i > l + 1$. Tedy dva výskyty podslov jsou oddělené, pokud je mezi nimi alespoň jedno písmeno.

Definice 5.15. Nechť Σ je libovolná abeceda, $w \in \Sigma^*$ slovo nad touto abecedou a výskyty podslov $u = w[i, j]$ a $v = w[k, l]$ nejsou oddělené. Pak *sjednocením* těchto dvou výskytů podslov rozumíme podslovo $w[\min(i, k), \max(j, l)]$.

Pro libovolnou množinu A výskytů podslov slova w definujeme její *sjednocení* $\bigcup A$ jako množinu výskytů podslov slova w takovou, že všechny výskyty podslov z $\bigcup A$ jsou po dvou oddělené, každý výskyt podslova z A je také podvýskytem nějakého výskytu podslova z $\bigcup A$ a množina $\bigcup A$ má nejmenší součet délek všech výskytů podslov z $\bigcup A$ ze všech množin, které splňují tyto dvě podmínky.

Příklad 5.16. Jako příklad uvažme slovo $w = \text{abbababba}$ nad abecedou $\{a, b\}$. Výskyty podslov $w[1, 2] = ab$ a $w[3, 3] = b$ nejsou oddělené a jejich sjednocením je výskyt podslova $w[1, 3] = abb$. Výskyty podslov $w[1, 2] = ab$, $w[4, 6] = aba$ a $w[8, 9] = ba$ jsou po dvou oddělené.

Sjednocením množiny $A = \{w[1, 2], w[3, 3], w[4, 6], w[8, 9]\}$ těchto výskytů podslov je pak množina výskytů podslov $\{w[1, 6], w[8, 9]\} = \{\text{abbaba}, \text{ba}\}$.

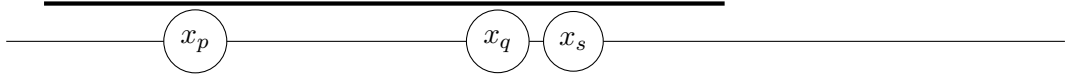
Lemma 5.17. Nechť Σ je libovolná abeceda, $w_1, w_2 \in \Sigma^*$ slova nad touto abecedou a $r > 0$ libovolné kladné celé číslo. Označme $R = 3^r$. Pak pokud platí $w_1 \approx_{3R}^R w_2$, tak také platí $w_1 \sim_r w_2$.

Důkaz. Ukážeme, že v r -tahové Ehrenfeuchtově-Fraïssého hře nad logikou FO[+1] má Duplicator vítěznou strategii.

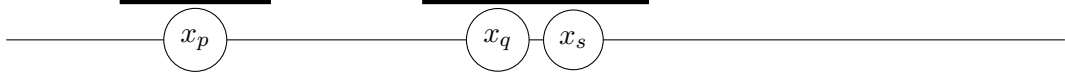
Myšlenkou této Duplicatorovy strategie je udržovat si po každém kroku kolem každého dosud položeného kamene podslovo o určitém poloměru, v jehož středu je příslušný kámen, a hrát tak, aby tato podslova byla v obou slovech stejná a byly na nich položené stejné kameny na stejných pozicích. Pokud nějaká taková dvě podslova nebudou oddělená, budeme pracovat s jejich sjednocením. V každém kroku hry pak bude Duplicator poloměry postupně zmenšovat. Tím se můžou slova původně vzniklá sjednocením více podslov kolem různých

kamenů rozpadnout na více oddělených podslov. Na konci hry pak budou poloměry podslov kolem kamenů zkráceny až na délku 0. Tudíž každé takové podslovo bude obsahovat jen pozici, na které je nějaký kámen, pokud na sousedních pozicích nejsou položeny jiné kameny, nebo takové podslovo bude obsahovat jen několik po sobě následujících pozic, na kterých je položen nějaký kámen. Protože Duplicatorova strategie je volena tak, aby každé podslovo v jednom slově odpovídalo i s pozicemi kamenů nějakému podslovu ve slově druhém, budou na konci hry platit obě podmínky (4.1) a (4.2).

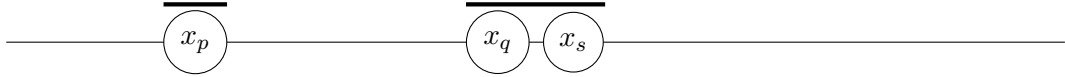
V průběhu hry tedy mohou podslova kolem dosud položených kamenů na jednom ze slov vypadat pro nějaké pozice kamenů x_p, x_q, x_s například následovně. Jednotlivá oddělená podslova znázorníme tučnou čarou.



V dalším tahu hry se mohou díky zkrácení poloměrů slov kolem kamenů původně sjednocená slova rozpadnout na více oddělených podslov.



Na konci hry pak budou poloměry podslov kolem položených kamenů rovny nule.



Tuto intuici nyní zachytíme formálně. Uvažme situaci po k -tém tahu hry. Tedy zatím byly umístěny kameny $x_1, \dots, x_k$. Označme pro každý kámen x_i výskyty podslov $u_{k,1,i}$ a $u_{k,2,i}$ jako

$$u_{k,1,i} = w_1[\max(1, \alpha_k(x_i) - 3^{r-k} + 1), \min(|w_1|, \alpha_k(x_i) + 3^{r-k} - 1)],$$

$$u_{k,2,i} = w_2[\max(1, \beta_k(x_i) - 3^{r-k} + 1), \min(|w_2|, \beta_k(x_i) + 3^{r-k} - 1)].$$

Označme dále pro každé $j \in \{1, 2\}$ zápisem $U_{k,j}$ množinu sjednocení všech výskytů podslov $u_{k,j,i}$ pro každé $1 \leq i \leq k$.

Množiny výskytů $U_{k,1}$ a $U_{k,2}$ nazveme izomorfní, pokud existuje bijekce $f: U_{k,1} \rightarrow U_{k,2}$ taková, že pro libovolný výskyt podslova $u \in U_{k,1}$ jsou podslova u a $f(u)$ stejná a ve výskytu podslova u se vyskytují právě tytéž kameny a na stejných pozicích jako ve výskytu podslova $f(u)$.

Ukážeme-li, že Duplicator má pro každé $0 \leq k \leq r$ takovou strategii pro k -tahovou hru, že na konci hry jsou množiny $U_{k,1}$ a $U_{k,2}$ izomorfní, potom z případu $k = r$ dostaneme tvrzení lemmatu. Pokud totiž jsou na konci hry libovolné kameny x_p, x_q položeny na pozicích, pro které platí $\alpha_r(x_p) = \alpha_r(x_q) + 1$, pak musí být oba kameny položeny ve stejném výskytu podslova z $U_{r,1}$. A tedy ve výskytu podslova $f(u)$ jsou oba kameny položeny na stejných pozicích, a proto platí také $\beta_r(x_p) = \beta_r(x_q) + 1$. Stejný argument lze použít pro opačný případ, ve kterém předpokládáme $\beta_r(x_p) = \beta_r(x_q) + 1$.

Ukážeme tedy indukcí vzhledem k k , že Duplicator má pro každé $0 \leq k \leq r$ takovou strategii pro k -tahovou hru, že na konci hry jsou množiny $U_{k,1}$ a $U_{k,2}$ izomorfní. Pro $k = 0$ tvrzení zřejmě platí. Předpokládejme tedy, že Duplicator má strategii pro k -tahovou hru, na jejímž konci jsou množiny $U_{k,1}$ a $U_{k,2}$ izomorfní. Chceme ukázat, že pak Duplicator má také strategii pro $(k + 1)$ -tahovou hru, na jejímž konci jsou množiny $U_{k+1,1}$ a $U_{k+1,2}$ izomorfní.

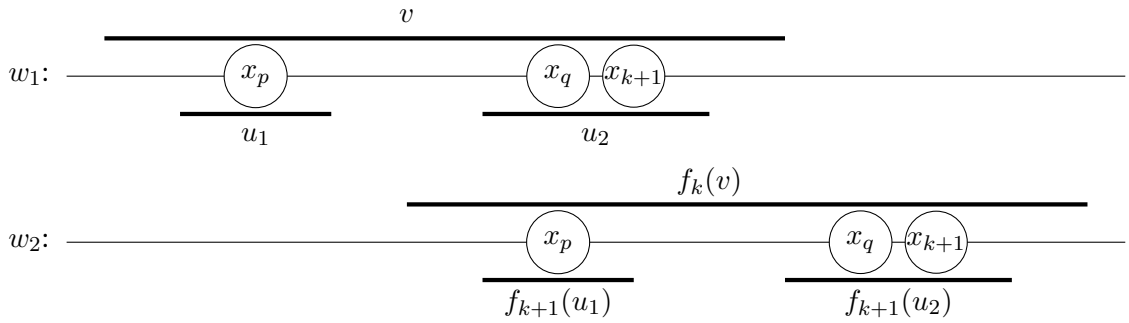
Označme tedy $f_k: U_{k,1} \rightarrow U_{k,2}$ izomorfismus množin $U_{k,1}$ a $U_{k,2}$. Budeme definovat Duplicatorovu strategii pro $(k+1)$ -tahovou hru a izomorfismus $f_{k+1}: U_{k+1,1} \rightarrow U_{k+1,2}$ množin $U_{k+1,1}$ a $U_{k+1,2}$. O definovaném zobrazení f_{k+1} nejprve ukážeme, že splňuje požadované podmínky a poté zdůvodníme, že z nich nutně vyplývá bijektivita zobrazení f_{k+1} .

Pro prvních k tahů použijeme Duplicatorovu strategii z indukčního předpokladu. Tuto strategii rozšíříme o tah $k+1$. Předpokládejme, že Spoiler tahu $k+1$ položil kámen na m -tou pozici ve slově w_1 . Pro případ, ve kterém Spoiler položil kámen na slovo w_2 , je důkaz analogický. Uvažme libovolný výskyt podslova $u \in U_{k+1,1}$.

Pokud výskyt podslova u ve slově w_1 neobsahuje kámen x_{k+1} , musel výskyt slova u vzniknout sjednocením nějakých výskytů podslov $z u_{k+1,1,1}, \dots, u_{k+1,1,k}$. Takže výskyt podslova u ve slově w_1 musí být také podslovem nějakého výskytu podslova $v \in U_{k,1}$. Proto stačí za $f_{k+1}(u)$ zvolit odpovídající výskyt podslova u ve výskytu podslova $f_k(v) \in U_{k,2}$.

Pokud naopak výskyt podslova u obsahuje kámen x_{k+1} , uvažme dvě možnosti. Jestliže výskyt podslova $u \in U_{k+1,1}$ leží celý uvnitř nějakého výskytu podslova $v \in U_{k,1}$, stačí Duplicatorovi položit kámen x_{k+1} na odpovídající pozici ve výskytu podslova $f_k(v)$ ve slově w_2 . Poté stačí opět za $f_{k+1}(u)$ zvolit odpovídající výskyt podslova u ve výskytu podslova $f_k(v)$.

Předchozí dvě situace pro větší přehlednost znázorníme. V následujícím obrázku je u_1 výskyt podslova, který neobsahuje kámen x_{k+1} , a u_2 je výskyt podslova, který kámen x_{k+1} obsahuje a leží celý uvnitř nějakého výskytu podslova v z $U_{k,1}$.



Pokud výskyt podslova u obsahuje kámen x_{k+1} a neleží celý uvnitř žádného výskytu podslova z $U_{k,1}$, ukážeme, že pak musí výskyt podslova u být oddělený od všech výskytů podslov $U_{k+1,1}$, na kterých je položený nějaký kámen z $x_1, \dots, x_k$. Pak bude platit $u = u_{k+1,1,k+1}$. Buď tedy x_i libovolný kámen z $x_1, \dots, x_k$. Předpokládejme, že platí $m > \alpha_{k+1}(x_i)$, opačný případ se ukáže stejně. Pak, protože výskyt podslova $u_{k+1,1,k+1}$ neleží celý uvnitř žádného výskytu podslova z $U_{k,1}$, platí $m + 3^{r-k-1} - 1 > \alpha_{k+1}(x_i) + 3^{r-k} - 1$. Proto také platí

$$m - \alpha_{k+1}(x_i) > 3^{r-k} - 3^{r-k-1} = (3-1)3^{r-k-1} = 2 \cdot 3^{r-k-1},$$

a tudíž výskyt podslova $u_{k+1,1,k+1}$ je oddělený od výskytu podslova z $U_{k+1,1}$, na kterém je položený kámen x_i .

Budeme chtít ukázat, že ve slově w_2 je nějaká pozice, na kterou když Duplicator položí kámen x_{k+1} , bude platit $u_{k+1,1,k+1} = u_{k+1,2,k+1}$ a výskyt podslova $u_{k+1,2,k+1}$ bude také oddělený od všech výskytů podslov z $U_{k+1,2}$, na kterých je položený nějaký kámen z $x_1, \dots, x_k$. Pak stačí zvolit $f_{k+1}(u) = u_{k+1,2,k+1}$ a po tomto tahu Duplicatora budou požadované podmínky také platit.

Součet délek všech podslov z $U_{k,1}$ je nejvýše $k(2 \cdot 3^{r-k} - 1) \leq k \cdot 3^{r-k+1} < 3^{r+1}$, a tedy podslovo $u_{k+1,1,k+1}$ se jako podslovo nějakého z podslov z $U_{k,1}$ vyskytuje méně než 3^{r+1} -krát. Označme tento počet jako n . Protože množiny $U_{k,1}$ a $U_{k,2}$ jsou z indukčního předpokladu

izomorfní, vyskytuje se podslovo $u_{k+1,1,k+1}$ právě n -krát i jako podslovo nějakého z podslov z $U_{k,2}$.

Celkově tedy nejméně jeden výskyt podslova $u_{k+1,1,k+1}$ neleží celý uvnitř žádného výskytu podslova z $U_{k,1}$ a dalších $n < 3^{r+1}$ výskytů slova $u_{k+1,1,k+1}$ jsou jako podslova nějakého podslova z $U_{k,1}$. Protože podslovo $u_{k+1,1,k+1}$ má délku $2 \cdot 3^{r-k-1} - 1 \leq 3^r$, z předpokladu $w_1 \approx_{3R}^R w_2$ vyplývá, že slovo $u_{k+1,1,k+1}$ se jako podslovo slova w_2 vyskytuje také nejméně $(n+1)$ -krát. Tedy nejméně jeden z těchto výskytů není podslovem žádného podslova z $U_{k,2}$, a je proto oddělený od všech výskytů podslov z $U_{k+1,2}$.

Zbývá ukázat, že právě definované zobrazení f_{k+1} je bijekce. Nejprve se přesvědčíme o injektivitě tohoto zobrazení. Uvažme tedy libovolné dva výskyty podslov $u, v \in U_{k+1,1}$ takové, že $f_{k+1}(u) = f_{k+1}(v)$. Pak z podmínek kladených na zobrazení f_{k+1} vyplývá, že podslova u, v a $f_{k+1}(u)$ jsou a stejná a ve výskytu podslova u se vyskytují právě tytéž kameny a na stejných pozicích jako ve výskytu podslova v . Tudíž výskyty podslov u a v jsou stejné. Zobrazení f_{k+1} je také surjektivní, protože libovolný výskyt slova $u \in U_{k+1,2}$ obsahuje alespoň jeden kámen, a tedy pro výskyt podslova $v \in U_{k+1,1}$, který obsahuje tentýž kámen, platí $u = f_{k+1}(v)$. $\square$

Věta 5.18 (Thomas). *Nechť Σ je libovolná abeceda. Jazyk $L \subseteq \Sigma^*$ je lokálně prahově testovatelný právě tehdy, když je FO[+1]-definovatelný.*

Důkaz. Implikace „ $\Rightarrow$ “ vyplývá z Lemmatu 5.12.

Naopak buď L libovolný FO[+1]-definovatelný jazyk. Tedy existuje uzavřená formule $\varphi \in \text{FO}[+1](\Sigma)$, která ho definuje. Označme r kvantifikátorovou složitost této formule. Tedy jazyk L je sjednocením nějakých tříd ekvivalence $\equiv_r$. Z Věty 4.10 pak vyplývá, že jazyk L je také sjednocením nějakých tříd ekvivalence $\sim_r$. Podle Lemmatu 5.17 je tedy jazyk L sjednocením nějakých tříd ekvivalence $\approx_{3^{r+1}}^{3^r}$, a je tedy lokálně prahově testovatelný. $\square$

Důsledek 5.19. *Třída všech FO[+1]-definovatelných jazyků je vlastní podtřída třídy všech FO[<]-definovatelných jazyků.*

Důkaz. V Příkladu 5.11 jsme ukázali jazyk $L = \{a\}^* \{b\} \{a\}^* \{c\} \{a\}^*$, který není lokálně prahově testovatelný. A tedy podle předchozí věty není ani FO[+1]-definovatelný. Avšak jazyk L je FO[<]-definovatelný, protože je definovaný formulí

$$\varphi \equiv \exists x \exists y (x < y \wedge Q_b(x) \wedge Q_c(y) \wedge \forall z ((\neg(z = x) \wedge \neg(z = y)) \rightarrow Q_a(z))). \quad \square$$

Důsledek 5.20. *Buď Σ libovolná abeceda. Predikát $x < y$ nelze vyjádřit jako ekvivalentní formuli logiky FO[+1](Σ).*

Důkaz. Předpokládejme opak. Pak platí, že každý FO[<]-definovatelný jazyk je také FO[+1]-definovatelný. To je spor s předchozím důsledkem. $\square$

Závěr

V práci jsme shrnuli výsledky z několika různých zdrojů o souvislosti třídy regulárních jazyků a několika jejich podtříd s vyjádřením jazyků pomocí logiky druhého řádu. Výsledky ze všech těchto zdrojů jsme vyjádřili stejným formalismem a v případě nutnosti rozšířili a ilustrovali příklady. Některé z důkazů jsou vlastním dílem autora (2.1, 2.2, 3.5, 3.8, 3.13, 4.4), některé jsou značně upravené (2.11, 3.11, 4.10, 5.12) a některé jsou přepsané do jednotného formalismu, popsaného v práci, a v případě potřeby rozšířené a opravené.

Mezi možnými pokračováními této práce je popsání jazyků definovatelných formulemi $\text{FO}[\prec](\Sigma)$ kvantifikátorové složitosti nejvýše c pro pevně dané $c \in \mathbb{N}$. Tím získáme kvantifikační hierarchii jazyků. Dá se ukázat, že tato hierarchie je shodná s dot-depth hierarchií star-free jazyků [Tho82]. Další možností rozšíření je přidání dalších predikátů k logice $\text{FO}[\prec](\Sigma)$ – například jako v [Str94] můžeme přidat modulární predikáty, kterými lze vyjádřit pro libovolná $0 \leq r < q$ vlastnost „počet pozic, které splňují danou formuli, je kongruentní s r modulo q “. Takovou logikou můžeme popsat více jazyků než logikou $\text{FO}[\prec](\Sigma)$, protože je zřejmě možné definovat například jazyk $\{aa\}^*$, o kterém jsme ukázali, že není $\text{FO}[\prec]$ -definovatelný. Je také možné uvážit logiku, ve které bude budou mít pozice dva různé následníky – tím získáme logiku popisující binární stromy.

Literatura

- [BP89] Beauquier, Danièle a Jean Eric Pin: *Factors of Words*. V Ausiello, Giorgio, Mariangiola Dezani-Ciancaglini a Simona Ronchi Della Rocca (editoři): *ICALP*, svazek 372 z *Lecture Notes in Computer Science*, strany 63–79. Springer, 1989, ISBN 3-540-51371-X.
- [BS73] Brzozowski, J.A. a Imre Simon: *Characterizations of locally testable events*. *Discrete Mathematics*, 4(3):243 – 271, 1973, ISSN 0012-365X.
- [Bü60] Büchi, J. Richard: *Weak Second-Order Arithmetic and Finite Automata*. *Mathematical Logic Quarterly*, 6(1-6):66–92, 1960, ISSN 1521-3870.
- [ČKK06] Černá, Ivana, Mojmír Křetínský a Antonín Kučera: *Formální jazyky a automaty I [online]*. Dostupné z WWW <<http://is.muni.cz/elportal/?id=703389>>, 2006. ISSN 1802-128X. [cit. 2014-05-25].
- [CMMP12] Choffrut, Christian, Andreas Malcher, Carlo Mereghetti a Beatrice Palano: *First-order logics: some characterizations and closure properties*. *Acta Inf.*, 49(4):225–248, 2012.
- [DG08] Diekert, Volker a Paul Gastin: *First-order definable languages*. V *Logic and Automata: History and Perspectives, Texts in Logic and Games*, strany 261–306. Amsterdam University Press, 2008.
- [Ehr61] Ehrenfeucht, Andrzej: *An Application of Games to the Completeness Problem for Formalized Theories*. *Fundamenta Mathematicae*, 49:129–141, 1961.
- [Esp12] Esparza, Javier: *Automata and Logic*. V *Automata Theory: An Algorithmic Approach*. 2012.
- [Fra54] Fraïsse, Roland: *Sur quelques classifications des systemes de relations*. *Publications Scientifiques de l'Université d'Algerie, Series A*, 1:35–182, 1954.
- [Gö34] Gödel, Kurt: *On undecidable propositions of formal mathematical systems*. Lecture notes taken by Kleene and Rosser at Princeton, reprinted in *The Undecidable*, ed. Martin Davis, pp. 39–74., 1934.
- [Imm99] Immerman, Neil: *Ehrenfeucht-Fraïssé Games*. V *Descriptive Complexity, Graduate Texts in Computer Science*, strany 91–112. Springer New York, 1999, ISBN 978-1-4612-6809-3.
- [MP71] McNaughton, Robert a Seymour A. Papert: *Counter-Free Automata (M.I.T. Research Monograph No. 65)*. The MIT Press, 1971, ISBN 0262130769.

- [PST01] Pin, Jean Eric, H. Straubing a D. Thérien: *Some Results on the Generalized Star-Height Problem*, 2001.
- [Reg09] Reghizzi, Stefano Crespi: *Formal Languages and Compilation*. Springer Publishing Company, Incorporated, 2009, ISBN 978-1-4471-5514-0.
- [Sch65] Schützenberger, Marcel Paul: *On Finite Monoids Having Only Trivial Subgroups*. Information and Control, 8(2):190–194, 1965.
- [Str94] Straubing, Howard: *Finite Automata, Formal Logic, and Circuit Complexity*. Birkhäuser, Basel, 1994.
- [Tho82] Thomas, Wolfgang: *Classifying regular events in symbolic logic*. Journal of Computer and System Sciences, 25(3):360 – 376, 1982, ISSN 0022-0000.
- [Tho06] Thomas, Michael: *Logics for Context-Free Languages*. diplomová práce, Universität Hannover, Institut für Theoretische Informatik, Hannover, 2006.
- [Tur36] Turing, Alan M.: *On Computable Numbers, with an application to the Entscheidungsproblem*. Proceedings of the London Mathematical Society, 42:230–265, 1936.