

UNIVERZITA PALACKÉHO V OLMOUCI  
KATEDRA INFORMATIKY

M. Rotter, T. Kukučka, J. Zehnula

KMI/FJAA – Formální jazyky a automaty



25. dubna 2012

### **Abstrakt**

Tento dokument je pouze přepisem zápisků a poznámek z přednášek předmětu KMI/FJAA. Přednášel doc. Vilém Vychodil PhD.

## Obsah

<b>1. Historie</b>	<b>1</b>
<b>2. Kódová analýza</b>	<b>1</b>
2.1. Lexikální analýza . . . . .	1
2.2. Syntaktická analýza . . . . .	1
<b>3. Základní pojmy</b>	<b>1</b>
<b>4. Operace s řetězcí</b>	<b>2</b>
<b>5. Formální jazyk</b>	<b>4</b>
<b>6. Lexikografické uspořádání</b>	<b>4</b>
<b>7. Operace nad jazyky</b>	<b>4</b>
7.1. Množinové . . . . .	4
7.2. Ostatní . . . . .	5
<b>8. Gramatiky</b>	<b>5</b>
8.1. Přepisovací generovací pravidla . . . . .	5
8.1.1. Vlastnosti pravidel . . . . .	5
8.1.2. Příklady pravidel . . . . .	6
8.1.3. Přímé odvozování řetězců pomocí pravidel . . . . .	6
8.2. Formální gramatiky . . . . .	7
8.3. Hierarchie gramatik . . . . .	8
8.4. Gramatika nezkracující . . . . .	9
8.5. Základní vlastnosti bezkontextových gramatik . . . . .	10
<b>9. Automaty</b>	<b>13</b>
9.1. Reprezentace KNA . . . . .	14
9.2. Nedeterministický výpočet . . . . .	15
9.3. Rozšířená přechodová funkce . . . . .	15
9.4. Řetězce přijímané KNA . . . . .	16
9.5. Determinizace KNA . . . . .	16
9.6. Algoritmus pro převod KNA na KDA . . . . .	17
<b>10. Vztah regulárních jazyků a konečných automatů</b>	<b>18</b>
10.1. Regulární jazyky jsou rozpoznatelné KDA (implikace zleva) . . . . .	19
10.2. Jazyky rozpoznatelné KDA jsou regulární (implikace zprava) . . . . .	20
10.3. Regulární gramatiky . . . . .	21

<b>11.Nedeterministický konečný automat s <math>\varepsilon</math>-přechody</b>	<b>23</b>
11.1. Reprezentace $\varepsilon$ KNA . . . . .	24
11.2. Nedeterministický výpočet . . . . .	24
11.3. $\varepsilon$ -uzávěry množin stavů . . . . .	24
11.4. Rozšířená přechodová funkce . . . . .	25
11.5. Ekvivalence s KDA . . . . .	26
<b>12.Algoritmus na převod <math>\varepsilon</math> KNA na KDA</b>	<b>26</b>
<b>13.Regulární výrazy</b>	<b>27</b>
<b>14.Jazyky generované regulárními výrazy</b>	<b>28</b>
<b>15.Uzávěrové vlastnosti regulárních jazyků</b>	<b>28</b>
15.1. Základní uzávěrové vlastnosti . . . . .	29
15.2. Další uzávěrové vlastnosti . . . . .	31
15.3. Pumping lemma . . . . .	32
<b>16.Minimalizace KDA</b>	<b>33</b>
16.1. Zprava invariantní ekvivalence . . . . .	33
16.2. FaktORIZACE automatu . . . . .	34
16.3. Algoritmus pro hledání redukovaného automatu . . . . .	35
<b>17.Izomorfismus automatů</b>	<b>37</b>
<b>18.Bezkontextové gramatiky</b>	<b>38</b>
18.1. Derivační stromy . . . . .	40
18.2. Jednoznačné a nejednoznačné bezkontextové gramatiky . . . . .	42
18.3. Uzávěrové vlastnosti bezkontextových jazyků . . . . .	42
18.4. Bezkontextové gramatiky v programátorské praxi . . . . .	43

## Seznam obrázků

1.	Grafické znázornění komutativity zřetězení řetězců. . . . .	4
2.	Vychodilovo „vajíčko.“ . . . . .	9
3.	Pseudokód pro převod KNA na KDA. . . . .	18
4.	Pseudokód pro převod $\varepsilon$ KNA na KDA. . . . .	27

## Seznam tabulek

1.	Přechodová tabulka s množinami stavů . . . . .	14
2.	Přechodová tabulka pro 28. příklad . . . . .	24
3.	Tabulkový popis rozkladů stavů . . . . .	37

## 1. Historie

Počátek úvah, jež byly později základem seriózního zkoumání formálních jazyků potažmo automatů se datuje do 30. let. Průkopníkem této oblasti byl Noam Chomsky <sup>1</sup>.

Jako příklad selhání autora programovacího jazyka si uveďme jazyk **Fortran**, jehož konstrukce byla po syntaktické stránce špatná, což vedlo ke **gramatické ne jednoznačnosti** tohoto jazyka.

## 2. Kódová analýza

### 2.1. Lexikální analýza

Dělení kódu na tokeny<sup>2</sup>, jež se zapisují například ve stylu  $\langle \text{znak}, \text{identifikátor} \rangle$ . Příkladem je tedy i token  $\langle =, \text{assignment} \rangle$  a jiné.

### 2.2. Syntaktická analýza

Syntaktická analýza vytváří stromovou závislost jednotlivých tokenů, jejíž reprezentace se nazývá *syntaktický-derivační strom*. V rámci této analýzy rozlišme:

1. Teorii jazyků, jenž se zabývá stavbou jazyka (respektive jeho syntaxí) a poskytuje tzv. **generativní aparát**. Dodejme, že gramatika říká, v jakém tvaru může být zapsán validní program.
2. Teorii automatů, jež poskytuje tzv. **analytický aparát**. Dodejme, že automatem se rozumí de-facto jednoduchý algoritmus.

## 3. Základní pojmy

- **Symbol** (případně znak). Jedná se o syntaktický pojem (význam tedy nehraje roli), který představuje *jméno* (analogicky k *písmenu* z přirozeného jazyka). Mezi symboly počítejme například **0**, **+**, **Š**, **while**.
- **Abeceda**. Abecedou rozumíme množinu (například množinu  $X$ ) všech přípustných *symbolů* (znaků), přičemž taková množina je neprázdná (tedy  $|x| > 0$ ) a konečná. Konečnost množiny je omezení dané reprezentovatelností množiny v rámci počítačové techniky. Abecedy značíme řeckými písmeny. Například  $\Sigma, \Sigma', \Gamma, \dots, \Omega$ . Například  $\Sigma = \{a, b, c\}$ .
- **Řetězec** (případně slovo). Jedná se o konečnou posloupnost symbolů (znaků) vybraných z nějaké dané abecedy. Například  $\langle a_1, a_2, \dots, a_n \rangle \in \Sigma, n$  nazvěme *délkou řetězce*. Formálně definujme řetězec jakožto *zobrazení*

$$x : \{a, b, c, d, \dots, i, j, \dots\} \rightarrow \Sigma$$

kde

$$1 \rightarrow a, 2 \rightarrow b, 3 \rightarrow c$$

a tak podobně. Délku řetězce označme  $|x|$ .

<sup>1</sup>Jméno této osoby čti [čomski] a zapamatuj si ke státnicím, že Chomsky byl *nebezpečný levicový intelektuál*.

<sup>2</sup>Překládej jako *část, díl nebo také fráze*.

- **Prázdný řetězec.** Jedná se o řetězec, pro který platí, že  $|x| = 0$  a značíme jej  $\varepsilon$ , přičemž platí následující zápis:

$$\varepsilon \subseteq \emptyset \rightarrow \Sigma$$

Prázdný řetězec **není** symbolem, tedy  $\varepsilon \notin \Sigma$ .

**Věta 1:** Nad  $k$ -prvkovou abecedou je právě  $k^n$  řetězců délky  $n$ .

**Poznámka 1:** Uvedme si rovněž značení pro dva důležité pojmy:

- $\Sigma^*$  označuje množinu všech řetězců nad abecedou  $\Sigma$ .
- $\Sigma^+$  označuje množinu všech řetězců nad abecedou  $\Sigma$  vyjma  $\varepsilon$ .

## 4. Operace s řetězci

- **Zřetězení** (konkatenace). Jde v podstatě o spojení<sup>3</sup> dvou řetězců v daném pořadí do jednoho řetězce.

**Příklad 1:** Mějme dva řetězce  $a, b$ :

$$a_1 \dots a_n \text{ a } b_1 \dots b_m$$

Pak jejich zřetězení má tvar:

$$a_1 \dots a_n b_1 \dots b_m$$

Identifikátorem<sup>4</sup> operace zřetězení je  $\circ$ , například  $x \circ y$  je zřetězením řetězců  $x$  a  $y$ . Formálně takto:

$$\begin{aligned} x &: \{1, \dots, n\} \rightarrow \Sigma \\ y &: \{1, \dots, m\} \rightarrow \Sigma \\ x \circ y &: \{1, \dots, n + m\} \rightarrow \Sigma \end{aligned}$$

**Poznámka 2:** Algebraicky je tatáž operace zapsána jako  $\langle \Sigma^*, \circ, \varepsilon \rangle$ .

- **Rovnost řetězců** Pro prohlášení dvou řetězců za sobě rovné v žádaném smyslu je třeba splnit obecně dvě následující podmínky:

1. Oba řetězce mají stejnou délku, tedy  $|x| = |y|$ .
2. Bude-li délka označena jako  $n$ , pak musí platit, že  $\forall i | i \in \{1, \dots, n\}, x(i) = y(i)$ . Tedy každé dva k sobě náležící symboly z daných řetězců jsou si rovny.

Uvažujeme-li rovnost řetězců, pak je záhodno uvažovat následující pojmy:

- **Prefix** řetězce. Označme jej  $Pfx(x) = \{y | \exists z \text{ tak, že } yz = x\}$ .
- **Infix** řetězce. Označme jej  $Ifx(x) = \{y | \exists z_1, z_2 \text{ tak, že } z_1 y z_2 = x\}$ .
- **Suffix** řetězce. Označme jej  $Sfx(x) = \{y | \exists z \text{ tak, že } zy = x\}$ .

<sup>3</sup>Pro milovníky jazyka Scheme můžeme tuto operaci přirovnat k proceduře *append*

<sup>4</sup>Identifikátor zřetězení se velmi často v zápisech zřetězení vynechává.



**Věta 2:**

$$xy = xz \implies y = z$$

$$yx = zx \implies y = z$$

Algebraicky je operace zapsána jako  $\langle \Sigma^*, \cdot, \varepsilon \rangle$ .

**Věta 3:** Vyslovme předpoklad, že platí  $xy = uv$ . Pak platí právě jedno z těchto tvrzení:

$$x = u, y = v$$

$$|x| > |u| \text{ a } \exists w |w| \neq \varepsilon, \text{ tak že } x = uw \text{ a } v = wy$$

$$|x| < |u| \text{ a } \exists w |w| \neq \varepsilon, \text{ tak že } u = xw \text{ a } y = wv$$

• **N-tá mocnina** řetězce.

$$x^n = \left\{ \begin{array}{ll} x & \text{pro } n = 1 \\ xx^{n-1} & \text{v ostatních případech} \end{array} \right\}$$

respektive

$$x^n = \left\{ \begin{array}{ll} \varepsilon & \text{pro } n = 0 \\ xx^{n-1} & \text{v ostatních případech} \end{array} \right\}$$

**Poznámka 3:** Mějme na paměti, že operace mocnění má vyšší prioritu než-li operace konkaténace (zřetězení).

**Věta 4:** Mějme  $u$  a  $v \in \Sigma^*$ , pak platí  $uv = vu$  (komutativita), právě tehdy, když  $\exists z |z| \in \Sigma^*$  a nezáporná celá čísla  $p, q$  tak, že  $u = z^p$  a  $v = z^q$ .

Předpokládejme, že po  $p, z, q$  máme  $u = z^p, v = z^q$ . Pak obecně platí následující zápis:

$$uv = z^p z^q = z^{p+q} = z^q z^p = vu$$

Předpokládejme, že  $uv = vu$ . Indukcí přes  $|uv|$  předpokládáme, že tvrzení platí pro libovolné dva řetězce, jejichž délka zřetězení je menší než-li  $|uv|$ . Mohou nastat tyto případy:

1.  $|u| = |v|$ , pak  $u = v$ , pak  $z = u, p = q = 1$
2.  $|u| < |v|$

Berme v potaz také následující zápis doplněný obrázkem: 1.

$$uw = v$$

$$wu = v$$

$$uw = wu$$

$$|uw| < |uv|, \text{ tedy } \exists z, p, q \text{ tak, že } u = z^p, w = z^q, v = z^{p+q}$$



Obrázek 1. Grafické znázornění komutativity zřetězení řetězců.

## 5. Formální jazyk

Zavedme si pojem *formální jazyk nad množinou všech řetězců*  $\Sigma^*$ . Označme tento jazyk jako  $L$ . Pak platí tato tvrzení:

$$\begin{aligned} L &\subseteq \Sigma^* \text{ (každá podmnožina abecedy je jazykem)} \\ L &= \emptyset \text{ (prázdný jazyk)} \\ L &= \{\varepsilon\} \text{ (jazyk s prázdným řetězcem)} \\ L &= \text{jazyk } C++ \text{ (jazyk C++)} \\ &\vdots \end{aligned}$$

**Pozor,** obecně platí že **prázdný jazyk**  $\neq$  **jazyk s prázdným řetězcem**.

## 6. Lexikografické uspořádání

Předpokládejme uspořádání na množině  $\Sigma^*$ . Nazvěme toto uspořádání *striktním totálním*. Pak toto uspořádání například pro  $\Sigma = \{a_1, \dots, a_n\}$  je  $a_1 < a_2 < a_3 < \dots < a_n$ .

Totální striktní uspořádání označme  $<_l$ .

Položme  $x <_l y$  pro  $x, y \in \Sigma^*$ . To ale platí pokud platí alespoň jedno z následujících dvou tvrzení:

1.  $|x| < |y|$
2.  $|x| = |y|$  a  $\exists i$  tak, že  $x(i) < y(i)$  a zároveň  $x(j) = y(j)$  pro  $\forall j | j < i$

**Příklad 2:**  $\Sigma = \{0, 1\}$ . Triviálně tedy  $0 < 1$ . Následně striktně  $\varepsilon <_l 0 <_l 1 <_l 00 <_l 01 <_l 10 <_l 11$ .

**Věta 5:** Striktní totální uspořádání je asymetrické a tranzitivní. A pro  $x \neq y$  platí buď  $x <_l y$  nebo  $y <_l x$ .

**Důsledek 1:** Důsledkem věty 5 je tvrzení, že množina  $\Sigma^*$  je spočetně nekonečná. Dodejme, že jazyk je (obvykle) spočetná množina.

## 7. Operace nad jazyky

### 7.1. Množinové

Množinové operace nad jazyky jsou prakticky totožné operacím na kterýchkoliv jiných množinách. Můžeme tedy použít množinový průnik, sjednocení, komplement (doplňěk) nebo rozdíl.

## 7.2. Ostatní

- **Zřetězení** (produkt) množin. Vyjádřeme produkt takto:

$$L_1 L_2 = \{xy | x \in L_1, y \in L_2\}$$

Produkt množin není obecně komutativní, ale je asociativní, přičemž prázdná množina tuto operaci anihiluje. Uvedme si rovněž monoid  $\langle 2^{\Sigma^*}, \circ, \{\varepsilon\} \rangle$ .

- **Mocnina** jazyka. Mocninu vyjádříme takto:

$$L^n = \begin{cases} \{\varepsilon\} & \text{pro } n = 0 \\ LL^{n-1} & \text{pro } n \geq 1 \end{cases}$$

- **Kleeneho**<sup>5</sup> uzávěr neboli **iterace**. Tento uzávěr vyjádříme takto:

$$L^* = \bigcup_{i=0}^{\infty} L^i$$

- **Pozitivní** uzávěr neboli pozitivní iterace. Tento uzávěr vyjádříme takto:

$$L^+ = \bigcup_{i=1}^{\infty} L^i$$

Všimněte si podobností mezi těmito dvěma uzávěry. Pozitivní uzávěr vynechává prázdný řetězec.

## 8. Gramatiky

Jak víme, tak jazyky mohou být *nekonečné* ve smyslu, že obsahují nekonečný počet slov. Nabízí se tedy otázka, jak tyto jazyky rozumně popsat, jak je reprezentovat resp. jak vytvořit *konečnou* sadu pravidel, jejichž aplikace by vedla k opětovné generaci původního jazyka.

### 8.1. Přepisovací generovací pravidla

Pravidlem rozumíme zpravidla každou takto definovanou dvojici.

$$\langle x, y \rangle \in \Sigma^* \times \Sigma^*$$

Pak neformálně tvrdíme, že  $x$  se přepisuje na  $y$ . Nutno dodat, že předchozí zápis lze zapsat i například takto.

$$x \rightarrow y, \text{ kde symbol } \rightarrow \notin \Sigma \text{ můžeme prohlásit za tzv. metasymbol.}$$

#### 8.1.1. Vlastnosti pravidel

- *Nezkracující* pravidlo je pravidlo, o kterém platí, že  $|x| \leq |y|$ . Tedy aplikaci tohoto pravidla na vstupní řetězec určité nevznikne řetězec kratší, než-li jeho předloha.
- $\varepsilon$  - pravidlo je pravidlo tvaru  $x \rightarrow \varepsilon$ .

---

<sup>5</sup>Stephen Cole Kleene je známý matematik, jenž se významně podílel na položení základů teoretických počítačových věd.

### 8.1.2. Příklady pravidel

**Příklad 3:** Mějme zadání abecedy  $\Sigma = \{a, b, c\}$ . Pravidla s využitím této abecedy by mohla být například tato.

$$\begin{aligned} aa &\rightarrow bc \\ bb &\rightarrow abba \\ c &\rightarrow \varepsilon \end{aligned}$$

**Příklad 4:** Mějme další zadání abecedy  $\Sigma = \{expr, +, \times\}$ . Pravidla s využitím této abecedy by mohla být například tato.

$$\begin{aligned} expr &\rightarrow expr + expr \\ expr &\rightarrow expr \times expr \end{aligned}$$

### 8.1.3. Přímé odvozování řetězců pomocí pravidel

Uvažujme odvozovací pravidlo  $x \rightarrow y$  nad abecedou  $\Sigma$ , pak řekneme, že řetězec  $v$  je **přímo odvozen** z řetězce  $u$  pomocí pravidla  $x \rightarrow y$ , pokud  $\exists p, q \in \Sigma^*$  tak, že

$$\begin{aligned} u &= pxq \\ v &= pyq \end{aligned}$$

Značení předchozí operace je následující:

$$u \Rightarrow_{x \rightarrow y} v$$

Slovně bychom tento zápis vystihli jako „přímý přepis dle pravidla  $x \rightarrow y$ “.

Řetězec  $v$  vznikne přímým přepisem z  $u$  pomocí pravidel  $P \subseteq \Sigma^* \times \Sigma^*$ , pokud  $\exists \pi \in P$  tak, že  $u \Rightarrow_{\pi} v$ .

Značme  $u \Rightarrow_P v$ .  $P$  je množinou užitých pravidel.  $P$  i  $\Rightarrow_P$  jsou binární relace na  $\Sigma^*$  a  $P \subseteq \Rightarrow_P$ , tedy „ $P$  je podmnožinou šipky  $\Rightarrow_P$ “. Platí, že  $x \rightarrow y \in P$  a  $x \Rightarrow_{x \rightarrow y} y$ .

**Příklad 5:** Mějme abecedu  $\Sigma = \{a, b, c\}$  a soubor pravidel  $P = \{aa \rightarrow bc, a \rightarrow cab, bb \rightarrow \varepsilon\}$ . Pak by odvození v jednom kroku mohla vypadat například takto:

$$\begin{aligned} baaa &\rightarrow bbca \\ bac &\rightarrow bcabc \end{aligned}$$

**Definice 1:** Definujme pojem **derivace**. Jedná se o posloupnost řetězců ve tvaru:

$$x_0, \dots, x_k, \text{ kde } k \geq 0 \text{ a kde } \{x_0, \dots, x_k\} \in \Sigma^*$$

se nazývá **P-derivace délky  $k$** , pokud  $x_{i-1} \Rightarrow_P x_i, \forall 1 \leq i \leq k$ . Symbolicky totéž  $x_0 \Rightarrow_P x_1 \Rightarrow_P \dots \Rightarrow_P x_k$ . Počet odvození tedy značí *délku* derivace.

Pokud pro  $u, v \in \Sigma^*$   $\exists$  P-derivace  $u = x_0 \dots x_k = v$ , pak říkáme, že  $v$  je odvozeno z  $u$  pomocí pravidel z  $P$ , což značíme například  $u \Rightarrow_P^+ v$ , tímto je pochopitelně myšleno odvození ve více krocích. Platí, že  $P \subseteq \Rightarrow_P \subseteq \Rightarrow_P^+$ .

**Příklad 6:** Mějme abecedu  $\Sigma = \{a, \dots, z\}$  a pravidla stejná jako v příkladu 5. Nyní odvozujeme například takto:

$$\underline{baaa}, \underline{bbca}, \underline{ca}, \underline{ccab}$$

## 8.2. Formální gramatiky

Mějme následující entity:

- $\Sigma$  - abeceda terminálních symbolů (tyto symboly tvoří řetězce daného jazyka).
- $N$  - abeceda neterminálních symbolů (tyto symboly se užívají k řízení průběhu odvozování).

Dodejme, že obě množiny by měly být neprázdné a konečné.

**Definice 2:** Odvozovací pravidlo  $x \rightarrow y$  se nazývá *generativní*, pokud  $x$  obsahuje alespoň jeden neterminální symbol.

**Definice 3:** Mějme strukturu  $G = \langle N, \Sigma, P, S \rangle$ , kde  $N$  je abecedou neterminálních symbolů,  $\Sigma$  je abecedou terminálních symbolů,  $P$  je množinou odvozovacích pravidel a  $S \in N$  je tzv. *počátečním* resp. *startovním* neterminálem. Pak tuto čtveřici nazveme **gramatikou**.

**Poznámka 4:** Pokud chceme vyjádřit, že z jednoho symbolu odvozujeme několik možných alternativ, tak to zapíšeme místo klasického dlouhého zápisu  $y \rightarrow x_1, y \rightarrow x_2, \dots$  pomocí zkrácené notace např.  $y \rightarrow x_1 | x_2 | \dots$ .

**Příklad 7:** Gramatika může vypadat třeba takto:

$$\begin{aligned} N &= \{\varepsilon, S, D, I\} \\ \Sigma &= \{0, \dots, 9, +, -\} \\ P &= \{S \rightarrow -I \mid I \mid I, I \rightarrow DI \mid D, D \rightarrow 0 \mid 1 \mid \dots \mid 9\} \\ G &= \langle N, \Sigma, P, S \rangle \end{aligned}$$

**Příklad 8:** Nebo takto:

$$\begin{aligned} N &= \{S, X, Y\} \\ \Sigma &= \{a, b, c\} \\ P &= \{S \rightarrow XcYcX, X \rightarrow aX, X \rightarrow bX, X \rightarrow cX, X \rightarrow \varepsilon, Y \rightarrow abY, Y \rightarrow ab\} \\ G &= \langle N, \Sigma, P, S \rangle \end{aligned}$$

**Definice 4:** Každý řetězec  $x \in (N \cup \Sigma)^*$ , pro který platí  $S \rightarrow^* x$ , je **větná forma** gramatiky  $G = \langle N, \Sigma, P, S \rangle$ . Větná forma se nazývá **větou**, pokud  $x \in \Sigma^*$ .

**Definice 5:** Jazyk generovaný gramatikou definujeme jako:

$$L(G) = \{x \in \Sigma^* \mid S \Rightarrow_G^* x\}$$

Vidíme tedy, že takový jazyk obsahuje *věty*, které lze odvodit ze startovacího neterminálu pomocí pravidel této gramatiky.

**Příklad 9:** Tento příklad čerpá gramatiku z příkladu 8.

$$\begin{aligned} S &\Rightarrow_G^* abbccYcX \\ S &\Rightarrow_G^* Xcababababc \\ S &\Rightarrow_G^* cYcbaX \\ S &\Rightarrow_G^* abbccabca \\ S &\Rightarrow_G^* cabababc \end{aligned}$$

**Definice 6:** Gramatiky  $G_1$  a  $G_2$  jsou **ekvivalentní**, pokud generují stejný jazyk.

### 8.3. Hierarchie gramatik

- **Gramatiky typu 0** – jedná se o gramatiky bez omezení.
- **Gramatiky typu 1** – jedná se o tzv. *kontextové* nebo *kontextově závislé* gramatiky. Ty splňují následující omezení na tvar pravidel. Pro každé pravidlo gramatik tohoto typu platí, že:

1. Buď je (pravidlo) ve tvaru  $pAq \rightarrow p \times q$ , kde  $p, q \in (\Sigma \cup N)^*$ ,  $A \in N$ ,  $x \in (\Sigma \cup N)^*$ , kde  $p$  a  $q$  se nazývají levým resp. pravým **kontextem**.
2. Nebo je (pravidlo) ve tvaru  $S \rightarrow \varepsilon$ , kde  $S$  je startovní terminál gramatiky, ale pouze za předpokladu, že  $S$  se nevyskytuje na pravé straně žádného pravidla.

Zároveň platí pro každé pravidlo (s výjimkou pravidla  $S \rightarrow \varepsilon$ ), že délka odvozeného řetězce je minimálně stejně velká jako délka vstupního řetězce. Gramatika tedy zároveň obsahuje pouze tzv. *nezkracující* pravidla.

- **Gramatiky typu 2** – jedná se o tzv. *bezkontextové* gramatiky, jenž obsahují pravidla ve tvaru:

$$A \rightarrow x, \text{ kde } A \in N, x \in (\Sigma \cup N)^*$$

Na levých stranách pravidel tedy očekáváme pouze neterminální symbol a na pravé straně očekáváme minimálně jeden symbol (s výjimkou pravidla  $S \rightarrow \varepsilon$ ).  $S$  se navíc nesmí vyskytovat na pravých stranách pravidel.

**Příklad 10:** Mějme tuto gramatiku:

$$\begin{aligned} G &= \langle N, \Sigma, P, S \rangle \\ N &= \{A, S\} \\ \Sigma &= \{0, 1\} \\ P &= \{S \rightarrow 0A, A \rightarrow \varepsilon\} \end{aligned}$$

- **Gramatiky typu 3** – jedná se o tzv. *regulární* resp. *pravolineární* gramatiky, které obsahují pravidla ve třech následujících tvarech:

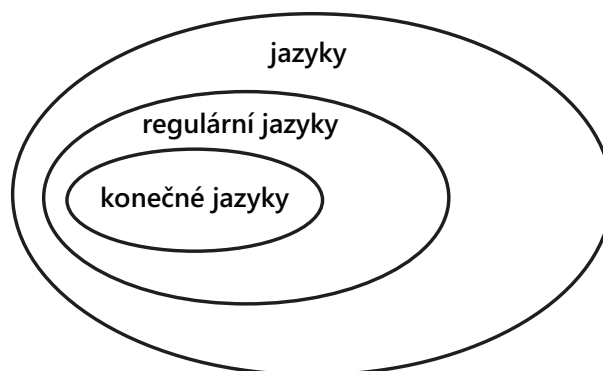
1.  $A \rightarrow bB$ , kde  $A, B \in N, b \in \Sigma$
2.  $A \rightarrow a$
3.  $S \rightarrow \varepsilon$

**Poznámka 5:** Každý konečný jazyk je regulární.

**Důkaz 1:** Mějme jazyk  $L = \{x_1, \dots, x_n\}$ . Abychom tento jazyk prohlásili za regulární, tak je třeba najít regulární gramatiku, která tento jazyk generuje.

Mějme tedy nějaké dané  $\Sigma$  a  $S$  a zvolme  $N$ . Následně platí  $\forall x_i \in L$  je dvojího typu:

1.  $x_i = \varepsilon$  a následně  $S \rightarrow \varepsilon$
2.  $x_i = a_1 \dots a_k$  a následně  $S \rightarrow a_{i1}A', A' \rightarrow a_{i2}A'', \dots, A^{k-1} \rightarrow a_{ik}A^k$



Obrázek 2. Vychodilovo „vajíčko.“

**Příklad 11:**

$$\begin{aligned}
 N &= \{S\} \\
 \Sigma &= \{a, b\} \\
 P &= \{S \rightarrow aSb | \varepsilon\} \\
 L(G) &= \{a^n b^n \mid n \geq 0\}
 \end{aligned}$$

Máme tedy *bezkontextový* jazyk.**Příklad 12:**

$$\begin{aligned}
 N &= \{S\} \\
 \Sigma &= \{a, b\} \\
 P &= \{S \rightarrow SS | aSb | bSa | \varepsilon\}
 \end{aligned}$$

 $L(G)$  je *bezkontextový* jazyk.**Příklad 13:**

$$\begin{aligned}
 N &= \{S, V\} \\
 \Sigma &= \{p, ), (, \Rightarrow, !\} \\
 P &= \{S \rightarrow V | (S \Rightarrow S) | !S, V \Rightarrow pV | p\}
 \end{aligned}$$

 $L(G)$  je jazyk všech výrokových formulí.

## 8.4. Gramatika nezkracující

Gramatika  $G$  se nazývá nezkracující, pokud má pouze nezkracující pravidla a může mít pravidlo ve tvaru  $S \rightarrow \varepsilon$ , přičemž  $S$  se nenachází na žádné z pravých stran.

**Příklad 14:**

$$\begin{aligned}
 N &= \{S, A, B, C\} \\
 \Sigma &= \{a, b, c\} \\
 P &= \{S \rightarrow \varepsilon | abc | Ac, A \rightarrow aBcb, Bcb \rightarrow bBc, Bcc \rightarrow Ccc, bc \rightarrow Cb, aC \rightarrow aab | aA\}
 \end{aligned}$$

**Věta 6:** Gramatiky typu 1(8.3.) a 3(8.3.) jsou nezkracující.

**Věta 7:** Ke každé gramatice  $G$ , existuje ekvivalentní gramatika  $G'$ , ve které jsou všechna pravidla obsahující terminální symboly ve tvaru  $A \rightarrow a$ , kde  $A \in N, a \in \Sigma$ .

**Důkaz 2:** Pro každý terminál  $a \in \Sigma$ , zavedeme terminál  $N_a$  a pravidlo  $N_a \rightarrow a$ . Všechny výskyty terminálů ve výchozích pravidlech nahradíme příslušnými pomocnými neterminály.

$$Bcb \rightarrow bBc \text{ se změní na } BN_cN_b \rightarrow N_bBN_c, N_c \rightarrow c, N_b \rightarrow b$$

**Věta 8:** Ke každé nezkracující gramatice existuje ekvivalentní gramatika, která je kontextově závislá.

**Důkaz 3:** Předpokládejme, že  $G = \langle N, \Sigma, P, S \rangle$  je nezkracující gramatika. Dle věty 7 můžeme předpokládat, že všechna pravidla jsou buď ve tvaru  $A \rightarrow a$  (nevadí) nebo ve tvaru obecně.  $A_1A_2 \cdots A_m \rightarrow B_1B_2 \cdots B_n$ , kde  $A_1, \dots, A_m, B_1, \dots, B_n \in N$  a navíc  $m \leq n$ . Tj. taková pravidla lze psát ve tvaru  $A_1A_2 \cdots A_m \rightarrow B_1B_2 \cdots B_{my}$ , kde  $y = B_{m+1} \cdots B_n$ . Budeme uvažovat nové pomocné neterminály  $X_1, \dots, X_m$ <sup>6</sup>. A zavedeme následující pravidla:

$$\begin{aligned} A_1A_2 \cdots A_m &\rightarrow X_1A_2 \cdots A_m \\ X_1A_2 \cdots A_m &\rightarrow X_1X_2A_3 \cdots A_m \\ &\vdots \\ X_1X_2 \cdots X_{m-1}A_m &\rightarrow X_1 \cdots X_{m-1}X_{my} \\ X_1X_2 \cdots X_{my} &\rightarrow B_1X_2X_3 \cdots X_{my} \\ &\vdots \\ B_1B_2 \cdots B_{m-1}X_{my} &\rightarrow B_1B_2 \cdots B_{m-1}B_{my} \end{aligned}$$

Tento postup se aplikuje pro všechna pravidla. Hledaná gramatika  $G'$  se skládá z  $\Sigma, N +$  všechny pomocné terminály + všechna odvozená pravidla.

### 8.5. Základní vlastnosti bezkontextových gramatik

- Levé strany pravidel obsahují jediný neterminál.
- Odvozování nezávisí na kontextu.

**Věta 9:** Mějme bezkontextovou gramatiku  $G = \langle N, \Sigma, P, S \rangle$  a necht'  $X_1 \cdots X_k, \dots, z$  je P-derivace délky  $n$ , kde  $X_1, \dots, X_k \in (N \cup \Sigma)$  a  $z \in (N \cup \Sigma)^*$  a potom pro každé  $i = 1, \dots, k$  existuje řetězec  $z_i \in (N \cup \Sigma)^*$  a P-derivace  $X_i, \dots, z_i$  délky  $n_i$  tak, že  $z = z_1z_2 \cdots z_k$  a  $n = n_1 + n_2 + \cdots + n_k$

**Důkaz 4:** Tvrzení prokážeme indukcí přes délku výchozí derivace  $X_1 \cdots X_k, \dots, z$ . Pro  $n = 0$ : Triviální  $z = X_1 \cdots X_k, z_i = X_i, n_i = 0$ . Každé  $X_i$  je derivace délky 0. Necht' tvrzení platí pro libovolnou derivaci délky  $n$  a dokážeme, že  $X_1 \cdots X_k$  je P-derivace délky  $n+1$ . Jelikož má uvažovaná P-derivace délku  $n+1$ , lze ji psát ve tvaru:

$$X_1 \cdots X_k, \dots, y^7, z$$

Máme  $y \Rightarrow_G z$ . Můžeme aplikovat indukční předpoklad: Existují řetězce  $y_1, \dots, y_k$  a P-derivace  $X_1, \dots, y_1$  až  $X_k, \dots, y_k$  délek  $n_1 \cdots n_k$  tak, že  $y = y_1y_2 \cdots y_k$  a  $n = n_1 + n_2 + \cdots + n_k$ . Z faktu, že  $y \Rightarrow_G z$  a z toho, že gramatika je bezkontextová plyne, že  $y$  je ve tvaru  $y = y''y'Aw''w''$  pro

<sup>6</sup>pro každé pravidlo se uvažují zvlášť

<sup>7</sup> $X_1 \cdots X_k, \dots, y$  má délku  $n$



$i = 1, \dots, k$ . Pak  $z$  je ve tvaru  $z = y'' y' u w' w''$  a  $A \rightarrow n \in P$ , to jest  $X_i, \dots, y_i, y' u w'$  je P-derivace délky  $n_{i+1}$ . Hledané derivace jsou:

$$\begin{array}{c} X_1, \dots, y_1 \\ \vdots \\ X_{i-1}, \dots, y_{i-1} \\ X_i, \dots, y_i y' u w' \\ X_{i+1}, \dots, y_{i+1} \\ X_k, \dots, j_k \end{array}$$

**Příklad 15:**

$$\begin{array}{lcl} N & = & \{S\} \\ \Sigma & = & \{a, b\} \\ P & = & \{S \rightarrow SS|aSb|bSa|\varepsilon\} \end{array}$$

Posloupnost:  $SbSaS, SbSa, SbaSba, aSbbaSba, abbaSba$  je P-derivace délky 4. Hledáme P-derivace:

1.  $S, aSb, ab$  (délka 2)
2.  $b$  (délka 0)
3.  $S, aSb$  (délka 1)
4.  $a$  (délka 0)
5.  $S, \varepsilon$  (délka 1)

**Příklad 16:** Gramatika s jediným pravidlem  $aBc \rightarrow abc$

**Poznámka 6:** U regulárních a kontextových gramatik lze hned vidět, jestli  $\varepsilon \in L(G)$ .

Pro bezkontextovou gramatiku  $G = \langle N, \Sigma, P, S \rangle$  zavedeme následující podmnožiny

$$\begin{aligned} E_0 &= \{A \in N | A \rightarrow \varepsilon \in P\} \\ E_{i+1} &= E_i \cup \{A \in N | A \rightarrow x, \text{ kde } x \in E_i^*\} \end{aligned}$$

**Příklad 17:**

$$\begin{aligned} A &\rightarrow \varepsilon \\ B &\rightarrow \varepsilon \\ E_0 &= \{A, B\} \\ E_1 &= \{A, B, F\} \\ E_2 &= \{A, B, F, G\} \\ E_i &\subseteq N, E_N = \bigcup_{i=0}^{\infty} E_i \end{aligned}$$

Jelikož je  $N$  konečná, musí platit:

$$E_0 \subseteq E_1 \subseteq E_2 \subseteq \dots \subseteq E_i = E_{i+1} = E_{i+2} \\ E_N = E_i$$

**Věta 10:** Pro každou bezkontextovou gramatiku  $G = \langle N, \Sigma, P, S \rangle$  a pro příslušné  $E_N$  platí následující  $A \Rightarrow_G^* \varepsilon$ , pak  $A \in E_N$ . Speciálně  $\varepsilon \in L(G)$ , pak  $S \in E_N$ .

**Důkaz 5:** Prokážeme obě implikace:

Pokud  $A \Rightarrow_G^* \varepsilon$ , pak prokážeme indukci přes délku P-derivace, tj. triviální případ je  $A \Rightarrow_G \varepsilon$ , tj. existuje pravidlo  $A \rightarrow \varepsilon \in P$  tj.  $A \in E_0$ . Předpokládejme, že tvrzení platí pro všechny P-derivace délky  $n$ . Mějme  $A, \dots, \varepsilon$  P-derivace délky  $n + 1$ . Použitím předchozí věty  $(A, X_1 \dots X_k, \dots, \varepsilon)$   $A, X_i \dots X_n, \dots, \varepsilon$ . Tzn. existují derivace  $X_i, \dots, \varepsilon$  délek nejvýše  $n$ . Z předpokladu  $X_i \in E_n$ , pro každé  $i$  tj.  $A \in E_N$ .  $\Leftarrow$  Dokáže, že pro každé  $E_i$  platí, pokud  $E \in E_i$  pak  $A \Rightarrow_G^* \varepsilon$ . Pro  $E_0$  zřejmé.  $A \rightarrow X_0 \dots X_k, A \in E_j$ .

**Věta 11:** Pro každou bezkontextovou gramatiku  $G$ , existuje bezkontextová gramatika  $G'$  neobsahující  $\varepsilon$  pravidla tak, že  $L(G) \setminus \{\varepsilon\} = L(G')$ .

**Důkaz 6:**  $G = \langle N, \Sigma, P, S \rangle$  - výchozí gramatika.

Stanovíme množinu  $E_n$  dle předchozího postupu  $G' = \langle N, \Sigma, P', S \rangle$ .  $P' = \{A \rightarrow y \mid A \rightarrow x \in P \text{ a } y \in D_{(x)}\}$ , kde  $D_{(x)}$  značí množinu řetězců, které jsou neprázdné a vznikly z řetězce  $x$  vynecháním libovolného množství neterminálů z  $E_N$ .

**Příklad 18:**

$$\begin{aligned} E_n &= \{A, B\} \\ X &\rightarrow aAbAB \\ &\dots \\ X &\rightarrow aAbAB \\ X &\rightarrow abAB \\ X &\rightarrow aAbB \\ X &\rightarrow aAbB \\ X &\rightarrow aAbA \\ X &\rightarrow abB \\ X &\rightarrow abA \\ X &\rightarrow aAb \\ X &\rightarrow ab \end{aligned}$$

**Věta 12:** Pro každou bezkontextovou gramatiku existuje ekvivalentní bezkontextová gramatika, která je navíc kontextová (a tudíž nezkracující)

**Důkaz 7:** Vstupní gramatika  $G$ . Dle předchozí věty existuje  $G'$  tak, že  $L(G) \setminus \{\varepsilon\} = L(G')$ .  $G'$  je nezkracující a kontextová, protože nemá  $\varepsilon$  pravidla. Pokud  $\varepsilon$  nepatří do  $L(G)$ , pak jsme hotovi. Pokud  $\varepsilon \in L(G)$ . Pak  $G'$  rozšíříme tak, že přidáme startovní symbol  $S'$  a pravidlo  $S' \rightarrow \varepsilon$  a  $S' \rightarrow S$ .

dopsat jednu stránku

## 9. Automaty

Gramatiky x automaty

generativní formalismus

Automaty - analytické formalismy

Konečné automaty: neformální výpočetní formalismus „jednoduchý počítač“ omezená paměť vstup: řetězec nad vstupní abecedou  $\Sigma$ . Řídící jednotka. Skládá se z konečně mnoha stavů. **Počátek činnosti:** Vstup = celý vstupní řetězec. Řídící jednotka je v počátečním (iniciálním) stavu. **Činnost automatu:** Na základě prvního symbolu na vstupu a na základě aktuálního stavu se řídící jednotka přepne do jiného stavu a odebere vstupní symbol.

**Konec činnosti:** Byl přečten celý vstupní řetězec. Podle toho v jaké končí automat stavu říkáme, že buď přijímá nebo zamítá vstupní řetězec. Některé stavy jsou označeny jako přijímací.

**Příklad 19:** sešit - automat (obr. 4.1)

Formalizace: Konečný deterministický automat (s úplnou přechodovou funkcí) (nad vstupní abecedou  $\Sigma$ ) je struktura:

$\langle \Sigma, Q, d, q_0 \rangle$

$\Sigma \dots$  vstupní abeceda

$Q \dots$  konečná množina stavů, která je neprázdná

$q_0 \in Q \dots$  počáteční stav

$F \subseteq Q \dots$  množina koncových stavů (přijímacích)

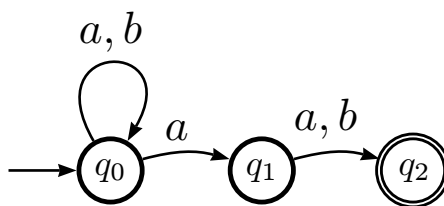
$\delta$  je zobrazení  $\delta : Q \times \Sigma \rightarrow Q$

$\delta(r, a) = q$  čteme: automat A při vstupním symbolu  $A \in \Sigma$  a aktuálním stavu  $r \in Q$  přejde do stavu  $q \in Q$

Pozn.:  $Q$  je konečná  $\delta \dots$  zobrazení

**Definice 7:** Za *determinismus* považujeme takovou konfiguraci, pro kterou platí, že je v každém jejím kroku jasné, co bude následovat. Naopak u *nedeterministických* konfigurací není v určitých případech možné další krok přesně vyjádřit na základě znalostí aktuálního kroku.

**Příklad 20:**



Vstupní řetězce: *abba* (nepřijat), *baba* (nepřijat), *baab* (přijat), *bbaa* (přijat).

V případě řetězce *baab* máme dokonce 3 možnosti výpočtu:

1.  $\langle q_0, baab \rangle, \langle q_0, aab \rangle, \langle q_0, ab \rangle, \langle q_0, b \rangle, \langle q_0, \varepsilon \rangle$  – končí neúspěchem.
2.  $\langle q_0, baab \rangle, \langle q_0, aab \rangle, \langle q_1, ab \rangle, \langle q_2, b \rangle$  – končí neúspěchem.
3.  $\langle q_0, baab \rangle, \langle q_0, aab \rangle, \langle q_0, ab \rangle, \langle q_1, b \rangle, \langle q_2, \varepsilon \rangle$  – končí úspěchem.

Předchozí zápisy můžeme pojmenovat také jako „nedeterministický výpočet.“

Jiným zápisem téhož může být také ten následující.

$$\langle \{q_0\}, baab \rangle, \langle \{q_0\}, aab \rangle, \langle \{q_0, q_1\}, ab \rangle, \langle \{q_0, q_1, q_2\}, b \rangle, \langle \{q_0, q_2\}, \varepsilon + \rangle$$

**Definice 8:** Strukturu  $A = \langle \Sigma, Q, \delta, I, F \rangle$  nazvěme **konečným nedeterministickým automatem** nad abecedou  $\Sigma$ . Pro tuto strukturu následně platí tato tvrzení:

- $\Sigma, Q$  a  $F$  jsou stejné jako u konečného deterministického automatu.
- $I$  označuje množinu počátečních stavů, která by měla být obecně neprázdná.
- $\delta$  označuje přechodovou funkci ve tvaru  $\delta : Q \times \Sigma \rightarrow 2^Q$ , tedy  $\delta(q, a) = \{r_1, \dots, r_k\}$ . Totéž slovně: „Automat může při stavu  $q$  při symbolu  $a$  přejít do kteréhokoliv stavu z  $\{r_1, \dots, r_k\}$ .“

**Příklad 21:**

$$\begin{aligned}\Sigma &= \{a, b\} \\ P &= \{q_0, q_1, q_2, q_3\} \\ I &= \{q_0, q_3\} \\ F &= \{q_2\}\end{aligned}$$

Následně přechodová funkce:

$$\delta = \{ \langle q_0, a, \{q_0, q_1\} \rangle, \langle q_0, b, \{q_0\} \rangle, \langle q_1, a, \{q_2\} \rangle, \langle q_1, b, \{q_2\} \rangle, \\ \langle q_2, a, \emptyset \rangle, \langle q_2, b, \emptyset \rangle, \langle q_3, a, \emptyset \rangle, \langle q_3, b, \emptyset \rangle \}$$

### 9.1. Reprezentace KNA

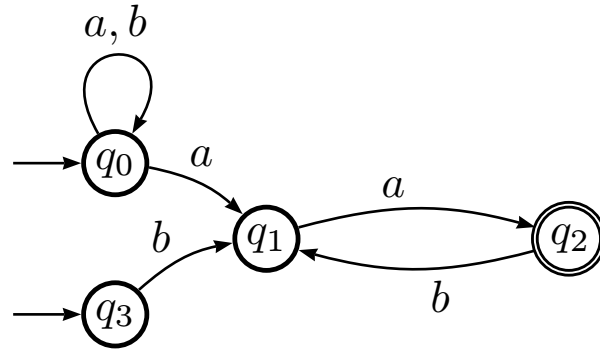
Předchozí příklad číslo 21 lze reprezentovat několika způsoby:

1. **Přechodová tabulka**, která ve svém těle obsahuje množiny stavů.

	a	b
$\rightarrow q_0$	$\{q_0, q_1\}$	$\{q_0\}$
$q_1$	$\{q_2\}$	$\{q_2\}$
$q_2^*$	$\emptyset$	$\emptyset$
$\rightarrow q_3$	$\emptyset$	$\{q_1\}$

Tabulka 1. Přechodová tabulka s množinami stavů

2. **Diagram**, který automat demonstruje v grafičtější podobě.



### 9.2. Nedeterministický výpočet

Nyní si popíšme **nedeterministický výpočet**, který je definován následujícími věcmi:

- Konfigurace, což je dvojice ve tvaru  $\langle stav, řetězec \rangle$ .
- Počáteční konfigurace ve tvaru  $\langle q, w \rangle$  kde  $q \in I$ .
- Koncová konfigurace ve tvaru  $\langle q, \varepsilon \rangle$ .
- Koncová přijímací konfigurace  $\langle q, \varepsilon \rangle$  kde  $q \in F$ .

**Definice 9:** Mějme  $A = \langle \Sigma, Q, \delta, I, F \rangle$  a  $w \in \Sigma^*$ . Pak posloupnost konfigurací  $\langle r_i, w_i \rangle$  pro  $i = \{0, \dots, n\}$  splňující podmínky:

$$R_0 \in I \quad (1)$$

$$w_0 = w \quad (2)$$

$$w_n = \varepsilon \quad (3)$$

$$w_i = a_i w_{i+1} \text{ a } r_{i+1} \in \delta(r_i, a_i) \text{ pro } i = \{0, \dots, n-1\} \quad (4)$$

nazveme **nedeterministický výpočet**.

### 9.3. Rozšířená přechodová funkce

**Definice 10:** Rozšířená přechodová funkce má tvar:

$$\delta^* : \Sigma^Q \times \Sigma^* \rightarrow \Sigma^Q$$

$$\delta^*(R, w) = \begin{cases} R & \text{pokud } w = \varepsilon \\ \delta^*(\bigcup_{q \in R} \delta(q, w), u) & \text{pokud } w = au, \text{ kde } a \in \Sigma, u \in \Sigma^q \end{cases}$$

**Věta 13:** Platí  $\delta^*(R, w) = \delta^*(\delta^*(R, u), v), \forall R \subseteq Q, uv \in \Sigma^*$ .

**Důkaz 8:** Předchozí tvrzení dokazujeme indukcí přes délku řetězce.

1. Pro  $u = \varepsilon$  je situace triviální.
2. Pokud  $u = ay, |y| < |u|$ , pak  $\delta^*(R, w) = \delta^*(R, ayv) = \delta^*(R, a(yv))$ .

3. Nyní aplikujme definici.

$$\begin{aligned}\delta^*\left(\bigcup_{q \in R} \delta(q, a), yv\right) &= \text{indukční předpoklad} \\ \delta^*\left(\delta^*\left(\bigcup_{q \in R} \delta(q, a), y\right), v\right) &= \text{definice } \delta^* \\ \delta^*(\delta^*(R, ay), v) &= \delta^*(\delta^*(R, u), v)\end{aligned}$$

**Věta 14:** Platí následující tvrzení:

$$\delta^*\left(\bigcup_{i=1}^k R_i, w\right) = \bigcup_{i=1}^k \delta^*(R_i, w) \text{ pro každé } R_i \subseteq Q, w \in \Sigma^*$$

**Důkaz 9:** Předchozí tvrzení dokazujeme indukcí přes délku řetězce  $w$ .

$$\begin{aligned}\delta^*\left(\bigcup_{i=1}^k R_i, w\right) &= \delta^*\left(\bigcup_{i=1}^k R_i, au\right) = \delta^*\left(q \in \bigcup_{i=1}^k \delta(q, a), u\right) \\ &= \delta^*\left(\bigcup_{i=1}^k \bigcup_{q \in R_i} \delta(q, a), u\right) \dots \text{indukční předpoklad} \\ &= \bigcup_{i=1}^k \delta^*\left(\bigcup_{q \in R_i} \delta(q, a), u\right) \\ &= \bigcup_{i=1}^k \delta^*(R_i, au) = \bigcup_{i=1}^k \delta^*(R_i, w)\end{aligned}$$

#### 9.4. Řetězce přijímané KNA

KNA  $A$  přijímá řetězec  $w$ , pokud  $\delta^*(I, w) \cap F \neq \emptyset$ . Navíc jazyk, přijímaný KNA  $A$  si definujeme jako  $L(A) = \{w \in \Sigma^* \mid \delta^*(I, w) \cap F \neq \emptyset\}$ .

**Věta 15:** Platí, že  $w \in L(A)$  právě tehdy, když KNA  $A$  má přijímací výpočet pro  $w$ .

**Důkaz 10:** Předchozí tvrzení lze dokázat indukcí přes délku řetězce  $w$ .

$$q \in \delta^*(I, w) \text{ právě tehdy, když existuje výpočet pro } w, \text{ končící ve stavu } q \\ \text{dekompozice navíc } w = ua$$

#### 9.5. Determinizace KNA

**Věta 16:** Pro každý KDA  $A = \langle \Sigma, Q, \delta, q_0, F \rangle \exists$  KNA  $A'$  tak, že  $L(A) = L(A')$ .

**Důkaz 11:** Pro výchozí  $A$  uvažujme  $A' = \langle \langle \Sigma, Q, \delta', q_0, F \rangle, \text{pak } \delta'(q, a) = \{\delta(q, a)\}$ . Zbytek důkazu je zřejmý.

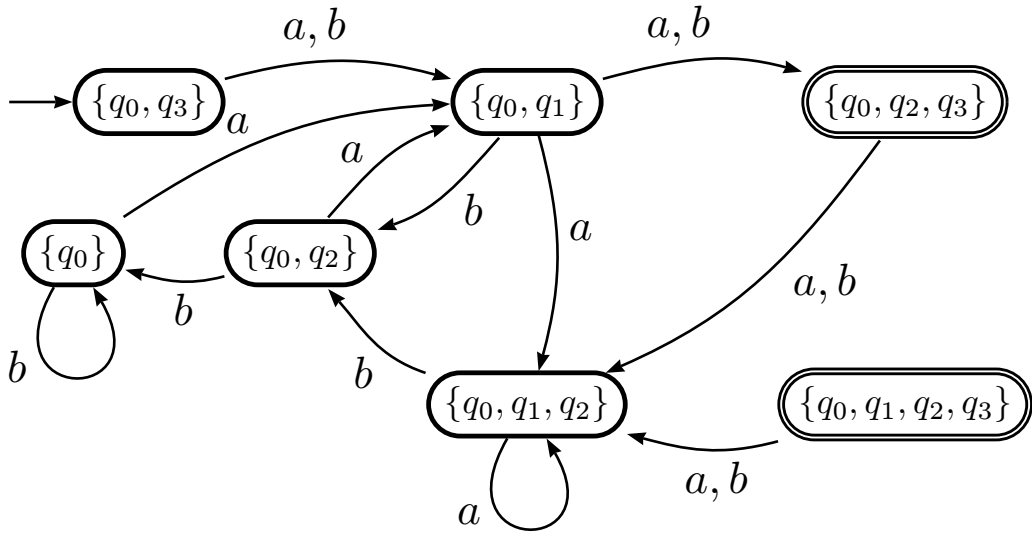
**Věta 17:** Pro každý KNA  $A$  existuje KDA  $A^D$  tak, že  $L(A) = L(A^D)$ .

**Důkaz 12:** Předchozí větu lze dokázat následujícím způsobem:

1. Uvažujme  $A^D = \langle \Sigma, 2^Q, \delta^D, I, F^D \rangle$ , kde  $F^D = \{R \subseteq Q \mid R \cap F \neq \emptyset\}$ ,  $\delta^D(R, a) = \delta^*(R, a)$ . Nyní zbývá ukázat, že  $\delta^*(I, w) \cap F \neq \emptyset$  právě tehdy, když  $(\delta D)^*(I, w) \in F^D$ , což prokážeme indukcí přes délku řetězce  $w$ .
2. Pro  $w = \varepsilon$  je situace zřejmá. Jinak  $(\delta D)^*(R, w) = (\delta D)^*(R, \varepsilon) = R = \delta^*(R, \varepsilon) = \delta^*(R, w)$ .
3. Předpokládejme, že tvrzení platí pro řetězce délky  $n$  a necht'  $w$  má délku  $n+1$  a  $w = au$  pro  $a \in \Sigma, |u| < |v|$ . Pak:

$$\begin{aligned} (\delta^D)^*(R, w) &= (\delta^D)^*(R, au) = (\delta^D)^*(\delta^D(R, a), u) = \\ &= \delta^*(\delta^D(R, a)u) = \delta^*(\delta^*(R, a), u) = \delta^*(R, au) = \delta^*(R, w) \end{aligned}$$

**Příklad 22:** Vemme KNA z příkladu 21.



Ještě jeden automat, nepřečtu to dobře ze sešitu. :)

## 9.6. Algoritmus pro převod KNA na KDA

Nyní si ukažme pseudokód algoritmu pro převod konečných nedeterministických automatů na konečné deterministické automaty, pro které platí, že akceptují řetězce stejného jazyka.

**Definice 11:** *Trie* je prefixový strom, který umožňuje „rychlé hledání ve slovníku.“

**Definice 12:** Jako *slovník* označujeme konečný neprázdný jazyk který neobsahuje  $\varepsilon$ .





### 10.1. Regulární jazyky jsou rozpoznatelné KDA (implikace zleva)

**Věta 18:** Pro každou regulární gramatiku  $G = \langle N, \Sigma, P, S \rangle$  existuje konečný deterministický automat  $A$  tak, že jazyk generovaný gramatikou je totéž, jako jazyk rozpoznatelný automatem, tj.  $L(G) = L(A)$

**Důkaz 13:** Nejprve uvažujeme situaci, že  $\varepsilon \notin L(G)$ . Uvažujme konečný nedeterministický automat  $A = \langle \Sigma, N \cup \{\#\}, \delta, \{S\}, \{\#\} \rangle$ .

$$\delta(A, b) = \begin{cases} \{B \in N \mid A \rightarrow bB \in P\} & \text{pokud } A \in N \wedge A \rightarrow b \notin P \\ \{B \in N \mid A \rightarrow bB \in P\} \cup \{\#\} & \text{pokud } A \in N \wedge A \rightarrow b \in P \\ \emptyset & \text{jinak} \end{cases}$$

Pro důkaz  $L(G) = L(A)$  stačí prokázat, že pro každé  $A \in N$  a  $x \in \Sigma^*$  platí, že  $A \Rightarrow_G^* x$  právě když  $\# \in \delta^*(\{A\}, x)$ .

Důkaz provedeme indukcí přes délku řetězce  $x$ .

1. Pro  $|x| = 1$  zřejmé.  $A \Rightarrow_G^* x$  právě když  $A \Rightarrow_G x$ , tj. existuje pravidlo  $A \rightarrow x \in P$  tj. z definice  $\delta^*$  platí, že  $\# \in \delta^*(\{A\}, x)$ . Nechť  $|x| = n$  a nechť tvrzení platí pro všechny řetězce kratší délky. Jelikož gramatika  $G$  je regulární, má  $P$ -derivace  $A, \dots, x$  právě  $n$  kroků. Pokud  $|x| > 1$  pak  $A \Rightarrow_G bB \Rightarrow_G^* by = x$  pro nějaké  $A \rightarrow bB \in P$ .
2. Pro  $|y| < n$  z indukčního předpokladu platí, že  $\# \in \delta^*(\{B\}, y)$ . Tím spíš  $\delta^*(\{A\}, x) = \delta^*(\{A\}, by) = \delta^*(\delta(A, b), y) = \delta^*(\{B, \dots\}, y) \supseteq \delta^*(\{B\}, y)$  tj.  $\# \in \delta^*(\{A\}, \#)$  protože  $A \rightarrow bB \in P$  tj.  $B \in \delta(A, b)$ .
3. Tím jsme prokázali, že pokud  $A \Rightarrow_G^* x$  pak  $\# \in \delta^*(\{A\}, x)$ .
4. Obráceně, pokud  $\# \in \delta^*(\{A\}, x)$  pak pro  $x = by, b \in \Sigma$  máme:  $\# \in \delta^*(\{A\}, by) = \delta^*(\delta(A, b), y) = \delta^*(\bigcup_{B \in \delta(A, b)} \{B\}, y) = \bigcup_{B \in \delta(A, b)} \delta^*(\{B\}, y)$  Tj. existuje  $B \in \delta(A, b)$  tak, že  $\# \in \delta^*(\{B\}, y)$ . Ze zavedení  $\delta$  plyne, že  $A \rightarrow bB \in P$ .
5. Aplikací indukčního předpokladu, existuje  $P$ -derivace  $B, \dots, y$ . Hledaná  $P$ -derivace je ve tvaru:  $A, bB, \dots, by = x$ , tj.  $A \Rightarrow_G^* x$ .

V případě, že  $\varepsilon \in L(G)$ , rozšíříme automat následovně, jednou ze tří možností:

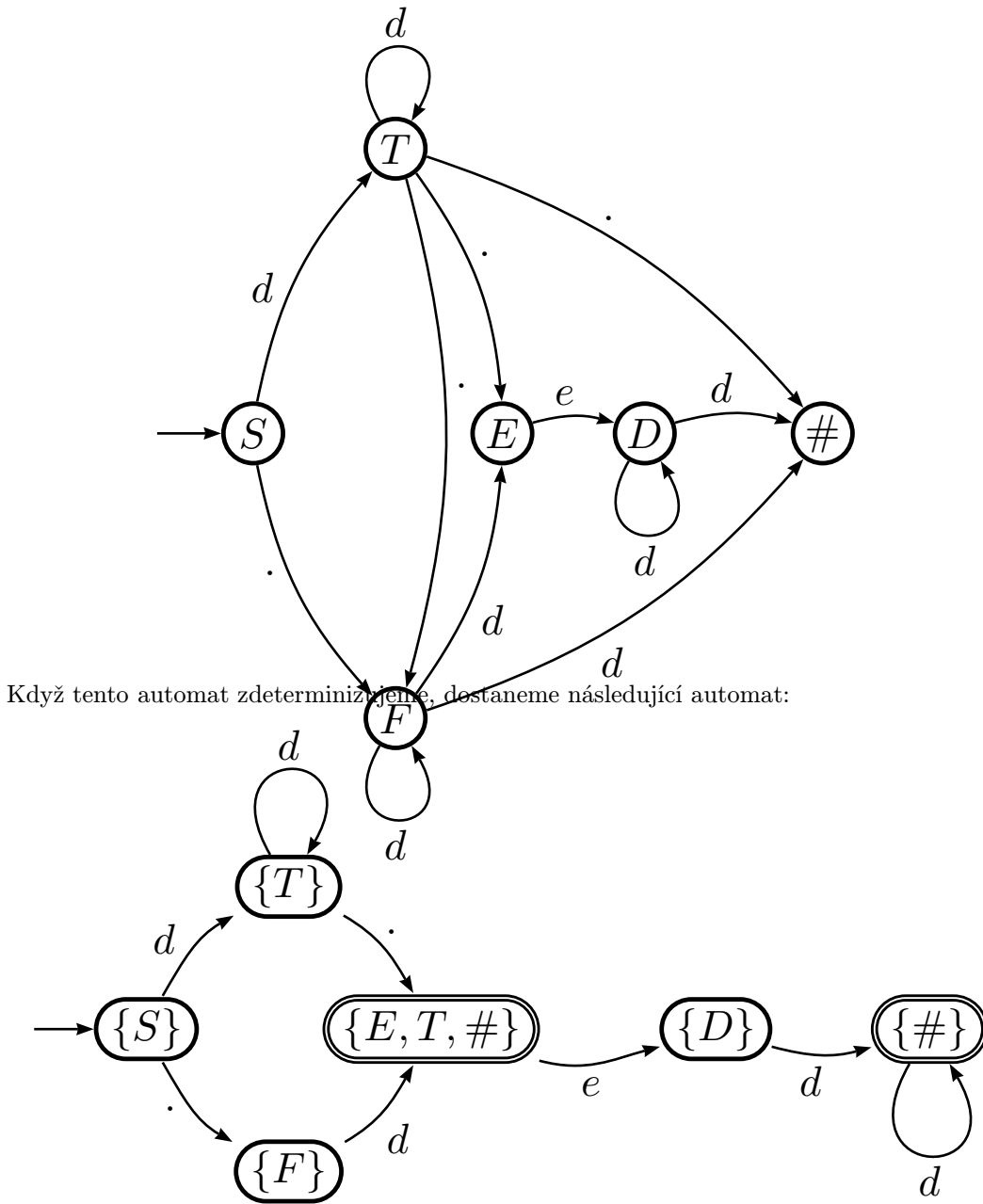
1. Přidáme  $S$  do množiny koncových stavů.
2. Přidáme  $\#$  mezi počáteční stavy
3. Zavedeme nový stav, který bude počáteční a zároveň koncový a nevedou z něj žádné přechody jinam.

**Poznámka 7:** Nyní zbývá automat pouze determinizovat.

**Příklad 24:** Máme gramatiku  $G$ .

$$\begin{aligned} G &= \langle N, \Sigma, P, S \rangle \\ \Sigma &= \{e, d, .\} \\ P &= \{S \rightarrow .F|dT, T \rightarrow .E|.F|dT|., D \rightarrow dD|d, E \rightarrow eD, F \rightarrow dE|dF|d\} \end{aligned}$$

Automat rozpoznávající jazyk, generovaný gramatikou  $G$ , bude vypadat následovně:



## 10.2. Jazyky rozpoznatelné KDA jsou regulární (implikace zprava)

**Věta 19:** Pro každý konečný deterministický automat  $A = \langle \Sigma, Q, \delta, q_0, F \rangle$  existuje regulární gramatika  $G$  tak, že  $L(A) = L(G)$ .

**Důkaz 14:** Za neterminální symboly  $G$  vezmeme stavy automatu. Startovní neterminál bude  $q_0$ . Uvažujeme gramatiku:  $G = \langle Q, \Sigma, P, q_0 \rangle$

$P = \{q \rightarrow ar \mid \text{pokud } \delta(q, a) = r, \text{ pro } q, r \in Q \text{ a } a \in \Sigma\}$   
 $\cup \{q \rightarrow a \mid \text{pokud } \delta(q, a) \in F\}$

Prokážeme že:  $q \Rightarrow_G^* x$  právě když  $\delta^*(q, x) \in F$

Pro  $|x| = 1$  platí:  $q \Rightarrow_G^* x$  právě když existuje pravidlo  $q \rightarrow x \in P$ , tj. z definice  $P$  platí  $\delta(q, x) \in F$

Pro  $x = by$ , kde  $b \in \Sigma^*$  předpokládejme, že tvrzení platí pro  $y$ . Platí, že  $q \Rightarrow_G br \Rightarrow_G^* by = x$  právě když  $\delta(q, b) = r$  a  $\delta^*(r, y) \in F$

To znamená  $\delta^*(q, by) = \delta^*(\delta(q, b), y) \in F$

Předchozí dokazuje, že  $x \in L(G)$  právě když  $x \in L(A)$  pro každý neprázdný  $x$ .

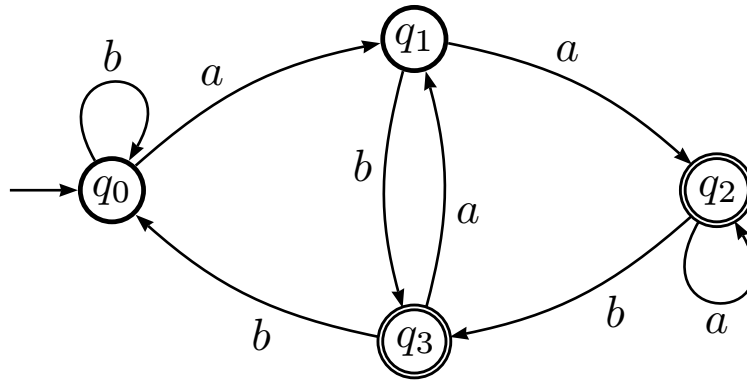
Pokud  $A$  nepřijímá  $\varepsilon$ , pak jsme hotovi.

Uvažujeme nový neterminál  $S$ , který bude nový startovní symbol, tj. místo  $G$  uvažujeme  $G' = \langle Q \cup \{S\}, \Sigma, P', S \rangle$

$P' = \{S \rightarrow \varepsilon\} \cup \{S \rightarrow x \mid q_0 \rightarrow x \in P\} \cup P$

Pak  $L(A) = L(G)$ .

**Příklad 25:** Mějme abecedu  $\Sigma = \{a, b\}$  a automat zadaný diagramem:



Odvozovací pravidla gramatiky, generující tento jazyk budou:

$$\begin{aligned}
 q_0 &\rightarrow aq_1 \mid bq_0 \\
 q_1 &\rightarrow aq_2 \mid a \mid bq_3 \mid b \\
 q_2 &\rightarrow aq_2 \mid a \mid bq_3 \mid b \\
 q_3 &\rightarrow aq_1 \mid bq_0
 \end{aligned}$$

### 10.3. Regulární gramatiky

Co jsou to regulární gramatiky a jaké podmínky jejich odvozovací pravidla splňují již víme, ale můžeme si je ještě rozdělit na dva druhy, právě podle tvaru odvozovacích pravidel.

1. **Zprava regulární gramatiky:** Obsahují pravidla ve tvaru  $A \rightarrow bB$  tj. neterminál na pravé straně je napravo od terminálního symbolu.
2. **Zleva regulární gramatiky:** Obsahují pravidla ve tvaru  $A \rightarrow Bb$ . Analogicky se neterminál nachází vlevo od terminálního symbolu.

**Věta 20:** Pro každou zleva regulární gramatiku  $G = \langle N, \Sigma, P, S \rangle$  existuje konečný deterministický automat  $A$  tak, že  $L(A) = L(G)$ .

**Důkaz 15:** Budeme konstruovat automat, jehož stavy budou  $N$ , nový pomocný počáteční stav  $\#$  a jediný koncový stav je  $S$ .

Hledaný KNA  $A = \langle \Sigma, N \cup \{\#\}, \delta, \{\#\}, \{S\} \rangle$  s následovně definovanou přechodovou funkcí  $\delta$

$$\delta(q, a) = \begin{cases} \{A \in N \mid A \rightarrow a \in P\} & \text{pokud } q = \# \\ \{A \in N \mid A \rightarrow Ba \in P\} & \text{pokud } q = B \end{cases}$$

Ekvivalence  $L(A) = L(G)$  se dokazuje vzájemně jednoznačnou korespondencí P-derivace a nedeterministického výpočtu.

Pro derivaci:

$$x_0 = S, x_1, x_2, \dots, x_{n-1}, x_n = x$$

jsme schopni sestavit posloupnost

$$\langle \#, x_n \rangle, \langle A_{n-1}, y_{n-1} \rangle, \dots, \langle A_1, y_1 \rangle, \langle S, \varepsilon \rangle \text{ kde } x_i = A_i y_i$$

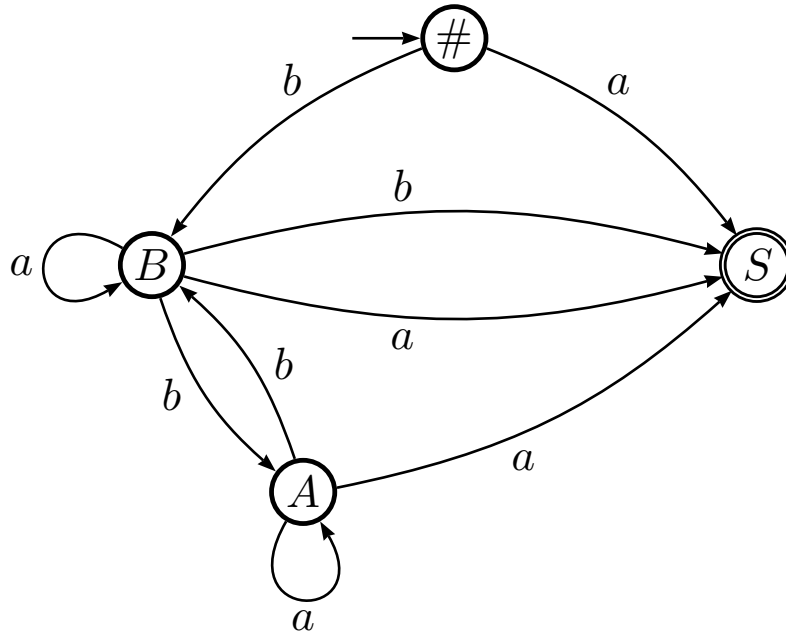
**Příklad 26:** Máme gramatiku  $G$  s následovně definovanými pravidly.

$$S \rightarrow Aa|Ba|Bb|a$$

$$A \rightarrow Aa|Bb$$

$$B \rightarrow Ab|Ba|b$$

Automat rozpoznávající jazyk generovaný touto gramatikou bude vypadat následovně:



**Věta 21:** Pro každý konečný deterministický automat  $A$  existuje zleva regulární gramatika taková, že  $L(A) = L(G)$

**Důkaz 16:** Neterminály gramatiky jsou stavy automatu a budeme uvažovat dodatečný startovní neterminál  $S$ .

$$P = \{ \delta(q, a) \rightarrow qa \mid q \in Q \wedge a \in \Sigma \} \cup \\ \{ \delta(q_0, a) \rightarrow a \mid q_0 \text{ je počáteční stav} \} \cup \\ \{ S \rightarrow w \mid w \text{ je pravá strana každého pravidla } q \rightarrow w, \text{ kde } q \in F \}$$

**Příklad 27:** Vezmeme KDA z příkladu 25. Odvozovací pravidla budou vypadat takto:

$$\begin{aligned} q_0 &\rightarrow q_0b \mid b \mid q_3b \\ q_1 &\rightarrow q_0a \mid a \mid q_3a \\ q_2 &\rightarrow q_1a \mid q_2a \\ q_3 &\rightarrow q_1b \mid q_2b \\ S &\rightarrow q_1a \mid q_1b \mid q_2a \mid q_2b \end{aligned}$$

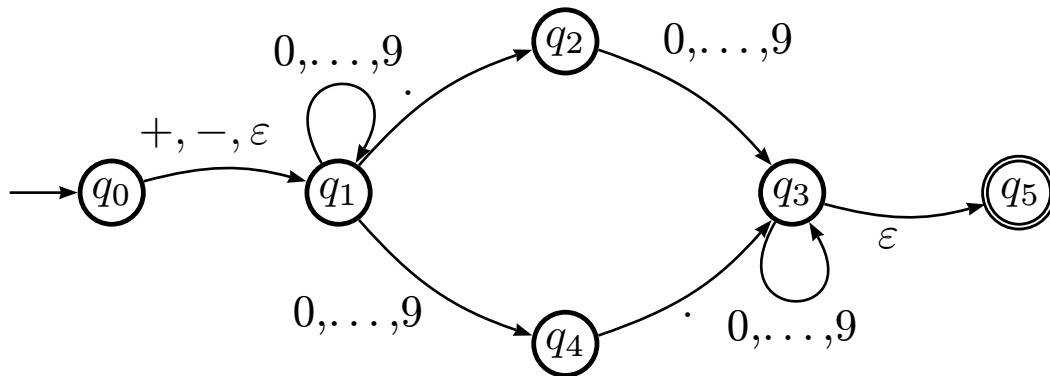
**Definice 14:** Regulární jazyky jsou jazyky, generované zprava (zleva) regulárními gramatikami, tj. jsou rozpoznatelné konečnými ne/deterministickými automaty.

**Poznámka 8:** Pravidla zprava a zleva nelze míchat.

## 11. Nedeterministický konečný automat s $\varepsilon$ -přechody

Značíme  $\varepsilon$ KNA.

**Příklad 28:** Zde je jeden motivační příklad na úvod.



**Definice 15:** Nedeterministický konečný automat s  $\varepsilon$ -přechody je struktura  $\langle \Sigma, Q, \delta, I, F \rangle$ , kde  $\Sigma, Q, \delta, I, F$  mají stejná význam jako u KNA.  $\delta$  je přechodová funkce  $\delta : Q \times (\Sigma \cup \{\varepsilon\}) \rightarrow 2^Q$ .

Fakt  $\delta(q, a) = \{r_1, \dots, r_k\}$  čteme: „automat  $A$  při čtení symbolu  $a$  přejde ze stavu  $q$  do některého ze stavů  $r_1, \dots, r_k$ “

Fakt  $\delta(q, \varepsilon) = \{r_1, \dots, r_k\}$  čteme: „automat  $A$  přejde samovolně ze stavu  $q$  do některého ze stavů  $r_1, \dots, r_k$ “

### 11.1. Reprezentace $\varepsilon$ KNA

1. **Přechodová tabulka**, vypadá stejně jako u KNA s tím, že přidáme jeden sloupec, ve kterém budeme zaznamenávat  $\varepsilon$ -přechody.

Takto bude vypadat předchozí 28. příklad, reprezentovaný pomocí tabulky.

	$+, -$	$.$	$0, \dots, 9$	$\varepsilon$
$\rightarrow q_0$	$\{q_1\}$	$\emptyset$	$\emptyset$	$\{q_1\}$
$q_1$	$\{q_2\}$	$\emptyset$	$\{q_1, q_2\}$	$\emptyset$
$q_2$	$\emptyset$	$\emptyset$	$\{q_3\}$	$\emptyset$
$q_3$	$\emptyset$	$\emptyset$	$\{q_3\}$	$\{q_5\}$
$q_4$	$\emptyset$	$\{q_3\}$	$\emptyset$	$\emptyset$
$q_5$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$

Tabulka 2. Přechodová tabulka pro 28. příklad

2. **Přechodový diagram**, u kterého mohou být některé hrany ohodnoceny  $\varepsilon$ . Viz příklad 28.

### 11.2. Nedeterministický výpočet

Pojem **konfigurace** pro nás zůstává stejný, jedná se stále o dvojici  $\langle stav, řetězec \rangle$ .

**Definice 16:** Mějme automat  $A = \langle \Sigma, Q, \delta, I, F \rangle$  a řetězec  $w \in \Sigma^*$ . Posloupnost konfigurací  $\langle r_i, w_i \rangle$  pro  $i = 0, \dots, n$  splňující podmínky:

1.  $r_0 \in I, w_0 = w$
2.  $w_n = \varepsilon$
3. pro každé  $i = 0, \dots, n$ 
  - (a)  $w_i = aw_{i+1}, r_{i+1} \in \delta(r_i, a)$
  - (b)  $w_i = w_{i+1}, r_{i+1} \in \delta(r_i, \varepsilon)$

### 11.3. $\varepsilon$ -uzávěry množin stavů

Je dána množina stavů  $R \subseteq Q$

$R$  se nazývá  $\varepsilon$ -uzavřená, pokud  $\delta(q, \varepsilon) \subseteq R$  pro každý stav  $q \in Q$ .

**Příklad 29:** Lze ukázat na našem příkladě 28.

$\{q_0\}$  není  $\varepsilon$ -uzavřená protože  $\delta(q_0, \varepsilon) = \{q_1\} \not\subseteq \{q_0\}$

$\{q_0, q_1\}$  je  $\varepsilon$ -uzavřená

$\varepsilon$ -uzávěr  $R$

vstup:  $R \subseteq Q$

$$E_0 = R$$

a pro  $i \geq 1$

$$E_i = E_{i-1} \cup \{\delta(q, \varepsilon) \mid q \in E_{i-1}\}$$

$$E_A(R) = \bigcup_{i=0}^{\infty} E_i$$

$$E_0 \subseteq E_1 \subseteq E_2 \subseteq \dots \subseteq E_A(R)$$

**Poznámka 9:** Vzhledem ke konečnosti množiny  $E_A(R)$  musí existovat index  $k$ , pro který platí  $E_k = E_{k+1} = \dots = E_A(R)$

Můžeme pozorovat, že  $E_A(R)$  není jen  $\varepsilon$ -uzavřená, ale je také **nejmenší**  $\varepsilon$ -uzavřená. Z toho můžeme vyvodit, že

$$E_A : 2^Q \rightarrow 2^Q$$

je **uzávěrový** operátor.

#### 11.4. Rozšířená přechodová funkce

Pro automat  $A = \langle \Sigma, Q, \delta, I, F \rangle$  je rozšířená přechodová funkce definovaná jako

$$\delta^* : 2^Q \times \Sigma^* \rightarrow 2^Q$$

$$\delta^*(R, w) = \begin{cases} E_A(R) & \text{pokud } w = \varepsilon \\ \delta^*(E_A(\bigcup_{q \in E_A(R)} \delta(q, a), u)) & \text{pokud } w = au, a \in \Sigma, u \in \Sigma^* \end{cases}$$

**Věta 22:** Pro libovolné množiny  $R_i \subseteq Q$  ( $i = 1, \dots, k$ ) platí:

$$\bigcup_{i=1}^k E_A(R_i) = E_A(\bigcup_{i=1}^k R_i)$$

**Důkaz 17:** Z monotonie  $E_A$  dostáváme

$$E_A(R_i) \subseteq E_A(\bigcup_{i=1}^k R_i) \text{ pro všechna } i$$

$$\bigcup_{i=1}^k E_A(R_i) \subseteq E_A(\bigcup_{i=1}^k R_i)$$

Opačná inkluze

$$\bigcup_{i=1}^k E_A(R_i) \text{ je } \varepsilon\text{-uzavřená a obsahuje } \bigcup_{i=1}^k R_i$$

z extenzivity  $E_A$  plyne, že  $\bigcup_{i=1}^k R_i \subseteq \bigcup_{i=1}^k E_A(R_i)$

Stačí tedy ukázat, že  $\bigcup_{i=1}^k E_A(R_i)$  je  $\varepsilon$ -uzavřená.

$$q \in \bigcup_{i=1}^k E_A(R_i) \Rightarrow \exists k \ q \in E_A(R_i)$$

Protože  $E_A(R_i)$  je  $\varepsilon$ -uzavřená,  $\delta(q, \varepsilon) \in E_A(R_i) \Rightarrow \delta(q, \varepsilon) \in \bigcup_{i=1}^k E_A(R_i) \Rightarrow \bigcup_{i=1}^k (E_A(R_i))$  je  $\varepsilon$ -uzavřená množina.

Nechť  $A = \langle \Sigma, Q, \delta, I, F \rangle$  je  $\varepsilon$ KNA. Řetězec  $w \in \Sigma^*$  nazýváme řetězec **přijímaný**  $A$ , pokud

$$\delta^*(I, w) \cap F \neq \emptyset$$

jinak  $w$  nazýváme řetězec **zamítaný**  $A$ .

### 11.5. Ekvivalence s KDA

**Věta 23:** Ke každému  $\varepsilon$ KNA  $A = \langle \Sigma, Q, \delta, I, F \rangle$  existuje KNA  $A^S = \langle \Sigma, Q, \delta^S, I^S, F \rangle$  takový, že  $L(A) = L(A^S)$ .

**Důkaz 18:**  $I^S = E_A(I)$

$$\delta^S(q, a) = \delta^*(\{q\}, a) = E_A(\bigcup_{u \in E_A(\{q\})} \delta(u, a))$$

Indukcí přes délku řetězce  $w \in \Sigma^*$  prokážeme, že  $\delta^{S*}(E_A(R), w) = \delta^*(R, w)$

1.  $w = \varepsilon$  platí triviálně
2. Předpokládejme, že tvrzení platí pro  $w$  délky  $n$  a dokážeme pro slova  $w$  délky  $n + 1$ .

$$w = av, \quad a \in \Sigma, \quad v \in \Sigma^*, \quad |v| = n, \quad R \subseteq Q$$

$$\begin{aligned} \delta^{S*}(E_A(R), w) &= \delta^{S*}(E_A(R), av) \\ &= \delta^{S*}\left(\bigcup_{q \in E_A(R)} \delta^S(q, a), u\right) \\ &= \delta^{S*}\left(\bigcup_{q \in E_A(R)} E_A\left(\bigcup_{u \in E_A(\{q\})} \delta(u, a)\right), u\right) \\ &= \delta^{S*}\left(E_A\left(\bigcup_{q \in E_A(R)} \bigcup_{u \in E_A(\{q\})} \delta(u, a)\right), u\right) \\ &= \delta^{S*}\left(E_A\left(\bigcup_{q \in E_A(R)} \delta(q, a)\right), u\right) \\ &= \delta^{S*}\left(E_A\left(E_A\left(\bigcup_{q \in E_A(R)} \delta(q, a)\right)\right), u\right) \\ &= \delta^*\left(E_A\left(\bigcup_{q \in E_A(R)} \delta(q, a)\right), u\right) \\ &= \delta^*(R, w) \end{aligned}$$

## 12. Algoritmus na převod $\varepsilon$ KNA na KDA

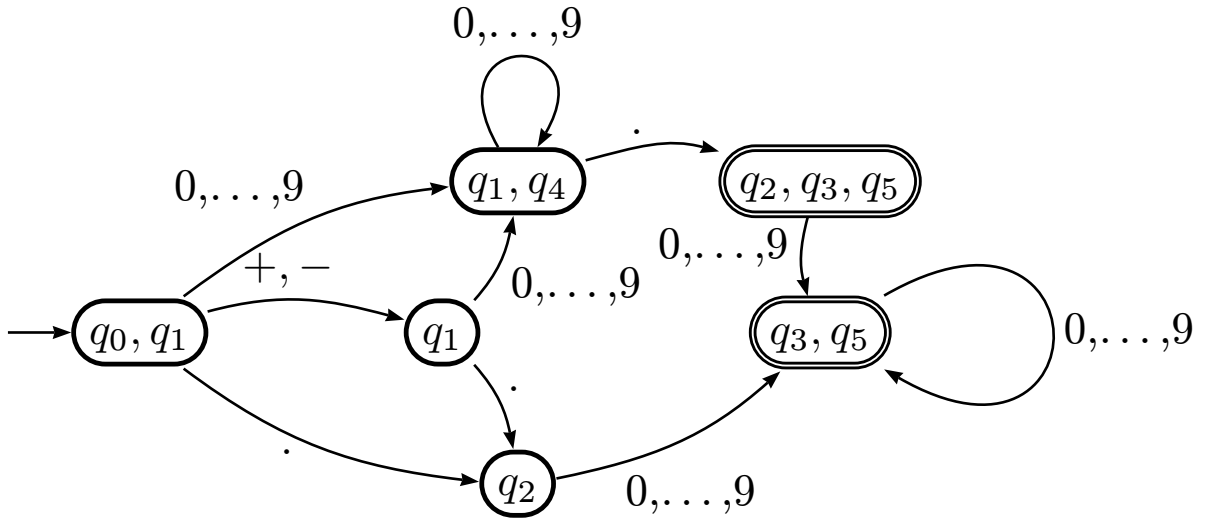
Máme  $A = \langle \Sigma, Q, \delta, I, F \rangle$  a víme, že  $W$  je neprázdná množina dosud nezpracovaných stavů.



```

 $\delta^s \leftarrow \emptyset$ ;  $Q^s \leftarrow \emptyset$ ;  $F^s \leftarrow \emptyset$ ;  $W \leftarrow E\_A(I)$ 
while  $W \neq Q$  do
  select  $R \in W$ 
   $W \leftarrow W \setminus R$ ;  $Q^s \leftarrow Q^s \cup R$ 
  if  $R \cap F \neq \emptyset$  then
     $F^s \leftarrow F^s \cup R$ 
  endif
  foreach  $a \in \Sigma$  do
     $N \leftarrow \delta^*(R, a)$ 
    if  $N \neq \emptyset$  then
      if  $N \notin W \cup Q^s$  then
         $W \leftarrow W \cup N$ 
      endif
      %% je to R, W, N ???
       $\delta^s \leftarrow \delta^s \cup \langle R, W, N \rangle$ 
    endif
  end
end
end
return  $\langle \Sigma, Q^s, \delta^s, I, F^s \rangle$ 

```

Obrázek 4. Pseudokód pro převod  $\varepsilon$  KNA na KDA.

### 13. Regulární výrazy

**Definice 17:** Necht' je dána  $\Sigma = \{a_1, \dots, a_k\}$ . Pak regulární výraz na  $\Sigma$  je:

1.  $\emptyset$
2.  $\varepsilon$
3. symboly  $a_1, \dots, a_k$
4. pokud  $R_1, R_2$  jsou RV, pak  $(R_1|R_2)$  je RV

5. pokud  $R_1, R_2$  jsou RV, pak  $(R_1 R_2)$  je RV
6. pokud  $R$  je RV, pak  $(R^*)$  je RV

**Příklad 30:** Podívejme se na následující výrazy a rozhodněme, zda-li jsou regulární:

- $((ab)c)^*, ((a|b)c)^*$  – jsou RV
- $a^*b), a||b$  – nejsou RV

## 14. Jazyky generované regulárními výrazy

**Definice 18:** Necht  $R$  je RV nad  $\Sigma$ . Pak  $L(R) \subseteq \Sigma^*$ . Pak také platí:

1.  $L(R) = \emptyset$ , pokud  $R = \emptyset$ .
2.  $L(R) = \{\varepsilon\}$ , pokud  $R = \varepsilon$ .
3.  $L(R) = \{a_i \mid \text{pokud } R = a_i\}$ .
4.  $L(R) = L(R_1) \cup L(R_2)$ , pokud  $R = \{R_1 | R_2\}$ .
5.  $L(R) = L(R_1) \circ L(R_2)$ , pokud  $R = \{R_1 \circ R_2\}$ .
6.  $L(R) = L(R_1)^*$ , pokud  $R = R_1^*$ .

**Věta 24:** Každý regulární výraz lze převést na konečný automat.

**Věta 25:** Každý jazyk generovaný regulárním výrazem je regulární.

**Důkaz 19:** Předchozí body dokážeme indukcí dle složitosti regulárního výrazu.

- Pro body 1 až 3 je vše zřejmé.
- Pro bod 4 –  $R = R_1 | R_2$ , kde  $R_1, R_2$  jsou regulární výrazy. Předpokládáme, že existuje KDA, rozpoznávající jazyky  $L(R_1)$  a  $L(R_2)$ ,  $L(R_1) = L(A_1)$ ,  $L(R_2) = L(A_2)$ .  
Pak vytvoříme KNA, který má tvar  $\langle \Sigma, Q_1 \cup Q_2, \delta_1 \cup \delta_2, \{q_0, q'_0\}, F_1 \cup F_2 \rangle$ .
- Pro bod 5 –  $R = R_1 R_2$ . Z koncových stavů  $A_1$  vytvoříme  $\varepsilon$ -přechody do počátečního stavu automatu  $A_2$  a počátečním stavem bude stav  $q_0$  z automatu  $A_1$ .
- Pro bod 6 –  $R = R_1^*$ ,  $L(A_1) = L(R_1)$ . Následně sestavujeme  $\varepsilon$  KNA tak, že před počátečním stavem vytvoříme nový koncový stav, který bude navíc novým počátečním stavem a do kterého vedeme  $\varepsilon$  přechody z již existujících koncových stavů a z našeho nového koncového stavu navíc vedeme  $\varepsilon$  přechod do původního počátečního stavu.

**Věta 26:** Každý regulární jazyk lze generovat regulárním výrazem.

## 15. Uzávěrové vlastnosti regulárních jazyků

### 15.1. Základní uzávěrové vlastnosti

- **Komplement**

Pokud je  $L$  regulární, pak je  $\Sigma^* \setminus L$  regulární.

**Důkaz 20:** Pro  $L$  existuje KDA s úplnou přechodovou funkcí tak, že  $L(A) = L$ . Na základě tohoto automatu zkonstruujeme  $A' = \langle \Sigma, Q, \delta, q_0, Q \setminus F \rangle$ ,  $A$  i  $A'$  mají stejnou rozšířenou přechodovou funkci  $\delta^*$ .

Tj.  $\Sigma^* \setminus L = L(A')$

- **Sjednocení**

Jsou-li  $L_1$  a  $L_2$  regulární, pak je  $L_1 \cup L_2$  regulární.

**Důkaz 21:** Pro  $L_1$  existuje  $A_1 = \langle \Sigma, Q_1, \delta_1, q_{01}, F_1 \rangle$ , tak že  $L_1 = L(A_1)$

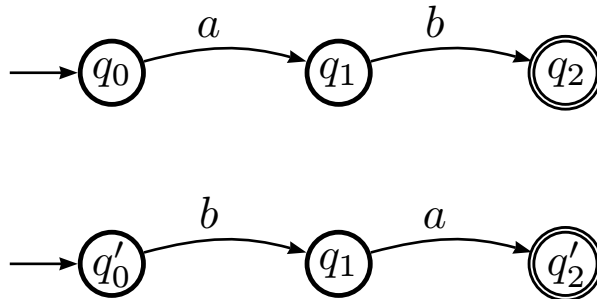
Pro  $L_2$  existuje  $A_2 = \langle \Sigma, Q_2, \delta_2, q_{02}, F_2 \rangle$ , tak že  $L_2 = L(A_2)$

Předpokládáme-li, že  $A_1$  a  $A_2$  mají disjunktní množiny stavů, tj.  $Q_1 \cap Q_2 = \emptyset$ , sestavíme následující KNA:

$A = \langle Q_1 \cup Q_2, \delta, \{q_{01}, q_{02}\}, F_1 \cup F_2 \rangle$  s přechodovou funkcí  $\delta$  definovanou jako

$$\delta(q, a) = \begin{cases} \{\delta_1(q, a)\} & \text{pokud } q \in Q_1 \\ \{\delta_2(q, a)\} & \text{pokud } q \in Q_2 \end{cases}$$

**Příklad 31:** Nyní si uvedeme protipříklad, co by se stalo, kdyby množiny stavů nebyly disjunktní.



První automat přijímá řetězec  $ab$ , druhý automat přijímá řetězec  $ba$ , čili od jejich sjednocení očekáváme, že bude přijímat  $ab$  i  $ba$ . Jelikož množiny stavů nejsou disjunktní ( $q_1$  je společný pro oba), sjednocení těchto automatů může stejně dobře přijímat i řetězce  $aa$  nebo  $bb$ , což je nežádoucí.

- **Průnik**

S použitím De Morganových zákonů, dostáváme:

$$L_1 \cap L_2 = \Sigma^* \setminus (\Sigma^* \setminus L_1 \cup \Sigma^* \setminus L_2)$$

tj.  $L_1 \cap L_2$  je regulární.

**Důkaz 22:** Uvažujeme konečné deterministické automaty s úplnou přechodovou funkcí  $A_1 = \langle \Sigma, Q_1, \delta_1, q_{01}, F_1 \rangle$  a  $A_2 = \langle \Sigma, Q_2, \delta_2, q_{02}, F_2 \rangle$

Zkonstruujeme automat  $A_1 \times A_2$  (direktní součin):

$$A_1 \times A_2 = \langle \Sigma, Q_1 \times Q_2, \delta, \langle q_{01}, q_{02} \rangle, F_1 \times F_2 \rangle$$

s přechodovou funkcí  $\delta$ :

$$\delta(\langle q, r \rangle, a) = \langle \delta_1(q, a), \delta_2(r, a) \rangle$$

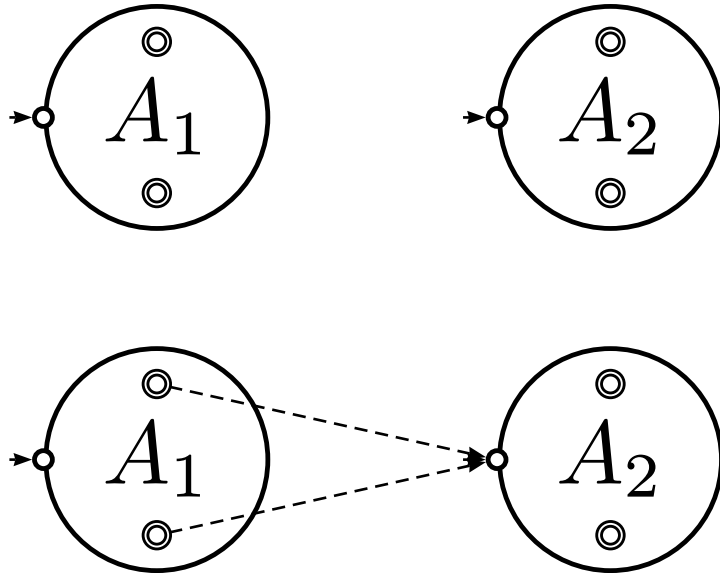
platí:

$$L_1 \cap L_2 = L(A_1 \times A_2)$$

- **Produkt (zřetězení)**

$$L_1 \cdot L_2$$

Předpokládáme existenci automatů  $A_1 = \langle \Sigma, Q_1, \delta_1, q_{01}, F_1 \rangle$  a  $A_2 = \langle \Sigma, Q_2, \delta_2, q_{02}, F_2 \rangle$  takových, že  $L_1 = L(A_1)$  a  $L_2 = L(A_2)$ .



Sestavíme  $\varepsilon$ KNA

$$A = \langle \Sigma, Q_1 \cup Q_2, \delta, \{q_{01}\}, F_2 \rangle$$

s přechodovou funkcí  $\delta : (Q_1 \cup Q_2) \times (\Sigma \cup \{\varepsilon\}) \rightarrow 2^{Q_1 \cup Q_2}$

$$\delta(q, a) = \begin{cases} \{\delta_1(q, a)\} & \text{pokud } q \in Q_1 \\ \{\delta_2(q, a)\} & \text{pokud } q \in Q_2 \end{cases}$$

$$\delta(q, \varepsilon) = \begin{cases} \{q_{02}\} & \text{pokud } q \in F_1 \\ \emptyset & \text{pokud } q \notin F_1 \end{cases}$$

- **Kleeneho uzávěr**

Pokud je  $L$  regulární, pak je  $L^*$  regulární.

Předpokládáme, že existuje automat  $A$  tak, že  $L = L(A)$

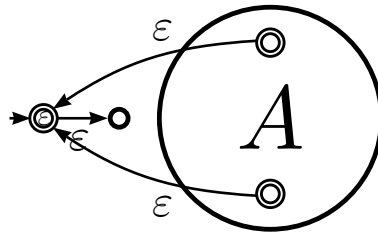
**Poznámka 10:** Automat rozpoznávající  $L^*$  musí mít možnost dostat se z koncového stavu zpět na začátek.

Pro automat  $A = \langle \Sigma, Q, \delta, q_0, F \rangle$  je potřeba zavést nový počáteční stav  $q_T \notin Q$ . Konstruovaný  $\varepsilon$ KNA bude vypadat následovně:

$$A' = \langle \Sigma, Q \cup \{q_T\}, \delta', \{q_T\}, F \cup \{q_T\} \rangle$$

s přechodovou funkcí  $\delta'$ :

$$\begin{array}{ll} \delta'(q, a) = \{\delta(q, a)\} & \text{pokud } q \in Q \\ \delta'(q, \varepsilon) = \emptyset & \text{pokud } q \in Q \setminus F \\ \delta'(q, \varepsilon) = \{q_T\} & \text{pokud } q \in F \\ \delta'(q_T, \varepsilon) = \{q_0\} & \\ \delta'(q_T, a) = \emptyset & a \in \Sigma \end{array}$$



## 15.2. Další uzávěrové vlastnosti

- **Množinový rozdíl**

Jsou-li  $L_1$  a  $L_2$  regulární, pak  $L_1 \setminus L_2 = L_1 \cap (\Sigma^* \setminus L_2)$  je regulární.

- **Kleeneho pozitivní uzávěr**

Je-li  $L$  regulární, pak  $L^+ = (L^* \setminus \{\varepsilon\}) \cup L$  je regulární.

- **N-tá mocnina jazyka**

$L^n \dots$  plyne z uzavření na produkt

- **Jazyk reverzních řetězců**

$$L^R = \{w^R \mid w \in L\}$$

Zdůvodníme pomocí konstrukce automatu rozpoznávající  $L^R$  (viz cvičení 6)

- **Jazyk sufixů**

$$Sfx(L) = \bigcup_{w \in L} Sfx(w)$$

**Důkaz 23:** Pro  $L$  uvažujeme  $A$  tak, že  $L = L(A)$ . Navíc  $A$  je KDA s úplnou přechodovou funkcí takový, že všechny jeho stavy jsou dosažitelné.

Námi hledaný automat je KNA definován jako:

$$A' = \langle \Sigma, Q, \delta, Q, F \rangle$$

s přechodovou funkcí

$$\delta(q, a) = \{\delta(q, a)\}$$

**Poznámka 11:** Všechny stavy jsou označeny za počáteční, aby měl automat možnost skočit do libovolné fáze výpočtu a tím "uhádnout" vynechané znaky řetězce, jehož sufix zkoumáme.

- **Jazyk prefixů**

$$\begin{aligned} Pfx(L) &= \bigcup_{w \in L} Pfx(w) \\ Pfx(L) &= (Sfx(L^R))^R \end{aligned}$$

Důkaz plyne z uzavření na Sfx a reverzní řetězec.

- **Jazyk infixů**

$$Ifx(L) = Pfx(Sfx(L))$$

Důkaz je taktéž zřejmý.

### 15.3. Pumping lemma

**Poznámka 12:** Jen drobné upozornění na začátek: jedná se o tvrzení ve tvaru když  $\rightarrow$  pak, tj. "Pokud je  $L$  regulární, pak ..."

**Věta 27:** Necht  $L$  je regulární jazyk nad  $\Sigma$ . Pak existuje  $n \in \mathbb{N}$  tak, že pro každý řetězec  $w \in L$  délky alespoň  $n$  platí, že existují  $x, y, z \in \Sigma^*$  tak, že jsou splněny podmínky:

1.  $w = xyz$
2.  $|xy| \leq n$
3.  $y \neq \varepsilon$
4. pro každé  $i \geq 0$  platí, že  $xy^iz \in L$

**Důkaz 24:** Rozlišíme dva případy.

- **$L$  je konečný**

pak je tvrzení triviální. Hledané  $n$  je ve tvaru  $l + 1$ , kde  $l$  je délka nejdelšího řetězce z  $L$ . Pak není v  $L$  žádný řetězec delší než  $n$  a tvrzení 1. – 4. platí triviálně.

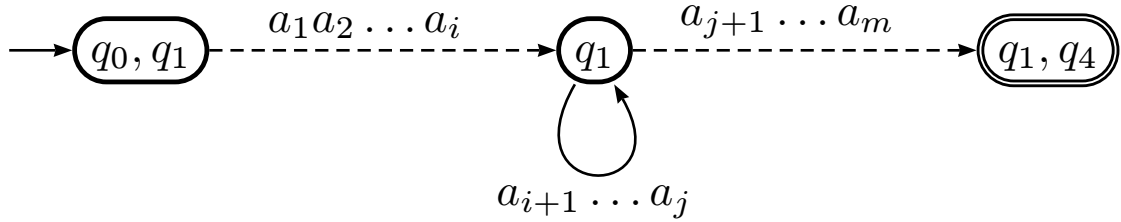
- **$L$  je nekonečný**

pak pro něj existuje KDA s množinou stavů  $Q$  tak, že  $L(A) = L$ . Položíme  $n = |Q|$ .

Pro každý řetězec  $w = a_1a_2 \cdots a_m$  kde  $m \geq n$  existuje přijímací výpočet délky  $m$ :

$$\langle q_0, w \rangle = \langle r_0, a_1 \cdots a_m \rangle, \langle r_1, a_2 \cdots a_m \rangle \cdots \langle r_{m-1}, a_m \rangle, \langle r_m, \varepsilon \rangle$$

Jelikož je  $m + 1 > n$  musí existovat alespoň 1 stav, který je v tomto přijímacím výpočtu zopakován.



Položme

$$\begin{aligned}
 x &= a_1 a_2 \dots a_i \\
 y &= a_{i+1} a_{i+2} \dots a_j \\
 z &= a_{j+1} a_{j+2} \dots a_m
 \end{aligned}$$

**Příklad 32:** Jazyk  $L = \{a^n b^n \mid \text{kde } n \in N\}$  není regulární.

**Důkaz 25:** Předchozí příklad 32 není regulární, což dokážeme sporem. Nechť je tedy  $L$  regulární a dle předchozí věty existuje číslo  $n$  tak, že vezmeme řetězec  $a^n b^n = xyz$  tak, že  $x = a^k, y = a^l, z = a^{n-k-l} b^n$ .

## 16. Minimalizace KDA

**Poznámka 13:** Pro regulární jazyk  $L$  více, než jeden automat  $A$  tak, že  $L = L(A)$  a navíc můžeme mít  $A_1, A_2$  tak, že  $L(A_1) = L(A_2)$ , ale  $|Q_1| < |Q_2|$ .

### 16.1. Zprava invariantní ekvivalence

**Definice 19:** Předpokládejme, že máme  $A = \langle \Sigma, Q, \delta, q_0, F \rangle$  a relaci ekvivalence  $\Theta \subseteq Q \times Q$  nazveme **zprava invariantní ekvivalencí** vzhledem k  $\delta$ , pokud platí, že  $\langle q_r \rangle \in \Theta$  a  $a \in \Sigma$ , pak  $\langle \delta(q, a) \delta(r, a) \rangle \in \Theta$ .

Pravá invariance reprezentuje přirozenou vlastnost, kterou musí mít relace nerozlišitelnosti stavů. Mezními případy invariantních relací zprava jsou:

1. identita  $\Theta = \{ \langle p, q \rangle \mid q \in Q \}$
2.  $\Theta = Q \times Q$

## 16.2. Faktorizace automatu

**Definice 20:** Mějme KDA  $A = \langle \Sigma, Q, \delta, q_0, F \rangle$  a zprava invariantní ekvivalenci  $\Theta$  vzhledem k  $\delta$ , pak zavedeme  $A/\theta = \langle \Sigma, Q/\Theta, \delta^{A/\Theta}, [q_0]_\Theta, F^{A/\Theta} \rangle$ , kde

$$\delta^{A/\Theta} = ([q]_\Theta, a) = [\delta(q, a)]_\Theta$$

a

$$F^{A/\Theta} = \{[q]_\Theta \mid q \in F\}$$

.

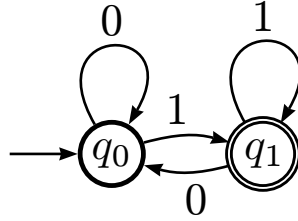
**Věta 28:** Automat  $A$  je dobře definovaný KDA.

**Důkaz 26:** Větu 28 si ověříme tak, že definice  $\delta^{A/\Theta}$  nezávisí na výběru prvků z tříd rozkladu dle  $\Theta$ .

Mějme  $[q]_\Theta = [r]_\Theta$ , to jest  $\langle q, r \rangle \in \Theta$ . Potom pro  $a \in \Sigma$  platí  $\langle \delta(q, a), \delta(r, a) \rangle \in \Theta$ , to jest  $[\delta(q, a)]_\Theta = [\delta(r, a)]_\Theta$ .

Obecně  $L(A/\Theta) \neq L(A)$ . Snažíme se najít co největší  $\Theta$ , tak aby tato rovnost platila.

**Příklad 33:**  $\Theta = Q \times Q$



**Věta 29:** Platí, že  $(\delta^{A/\Theta})^*([q]_\Theta, x) = [\delta^*(q, x)]_\Theta$  pro každý  $x \in \Sigma^*$ .

**Důkaz 27:** Dokážeme indukcí přes délku řetězce.

- Tedy  $x = \varepsilon$ , pak

$$(\delta^{A/\Theta})^*([q]_\Theta, \varepsilon) = [q]_\Theta = [\delta^*(q, \varepsilon)]_\Theta$$

- pak nechť toto tvrzení platí pro  $u \in \Sigma^*$  a řetězec  $x = au$ , kde  $a \in \Sigma$ .

$$\begin{aligned} (\delta^{A/\Theta})^*([q]_\Theta, au) &= (\delta^{A/\Theta})^*(\delta^{A/\Theta}([q]_\Theta, a), u) = \\ &= (\delta^{A/\Theta})^*([\delta(q, a)]_\Theta, u) = [\delta^*(\delta(q, a), u)]_\Theta = [\delta^*(q, au)]_\Theta \end{aligned}$$

**Definice 21:** Mějme KDA  $A = \langle \Sigma, Q, \delta, q_0, F \rangle$  s úplnou přechodovou funkcí. Pro stavy  $q, r \in Q$  položme  $q \equiv_A r$ , pokud pro každý řetězec  $x \in \Sigma^*$  platí, že  $\delta^*(q, x) \in F$ , pokud  $\delta^*(r, x) \in F$ .

**Věta 30:**  $\equiv_A$  je zprava invariantní operace.



**Důkaz 28:** Dokazujeme větu 30. Důkaz reflexivity je zřejmý, stejně tak symetrie i tranzitivita.

Nechť  $q \equiv_A r$  a máme  $a \in \Sigma$ . Máme dokázat, že  $\delta(q, a) \equiv_A \delta(r, a)$ . Pro každé  $x \in \Sigma^*$  platí:

$$\begin{aligned}\delta^*(\delta(q, a), x) &= \delta^*(q, ax) \\ \delta^*(\delta(r, a), x) &= \delta^*(r, ax)\end{aligned}$$

Užitím faktu, že  $\delta^*(q, ax) \in F$  dostaneme  $\delta^*(r, ax) \in F$ , tedy  $\delta^*(\delta(q, a), x) \in F$  právě tehdy, když  $\delta^*(\delta(r, a), x) \in F$ , to jest  $\delta(q, a) \equiv_A \delta(r, a)$  a tedy  $A / \equiv_A$ .

**Věta 31:** Pro  $A / \equiv_A$  a stav  $q \in Q$  platí, že  $q \in F$  právě, když  $[q]_{\equiv_A} \in F^{A/\equiv_A}$ .

**Důkaz 29:** Předchozí větu dokážeme tak, že dokážeme implikace z obou stran.

- Implikace zleva doprava („ $\Rightarrow$ “) plyne z definice.
- Implikace zprava doleva („ $\Leftarrow$ “): pokud  $[q]_{\equiv_A} \in F^{A/\equiv_A}$ , pak z definice  $\exists r \in F$  tak, že  $r \in [q]_{\equiv_A}$ . To znamená, že  $r \equiv_A q$  pro každý  $x \in \Sigma^*$  platí, že  $\delta^*(r, x) \in F$  právě tehdy, když  $\delta^*(q, x) \in F$ , speciálně pro  $x = \varepsilon$ , navíc  $\delta^*(r, \varepsilon) = r \in F$ , to jest  $\delta^*(q, \varepsilon) = q \in F$ .

**Důsledek 2:** KDA  $A$  nazveme redukovaný, pokud je  $\equiv_A$  identita.

**Věta 32:** KDA  $A$  je identita.

**Důkaz 30:** Automat  $A \equiv_A$  označme jako  $B$ , následně prokážeme, že  $\equiv_B$  je identita, tozn., že pokud  $[q]_{\equiv_A} \equiv_B [r]_{\equiv_A}$ , tak pak jsou si rovny.

Předpokládejme, že platí  $[q]_{\equiv_A} \equiv [r]_{\equiv_A}$ .

- Pak podle definice  $\equiv_B$  tozn, že pro  $x \in \Sigma^*$  platí  $\delta^{A/\equiv_A}([q]_{\equiv_A}, x) \in F$  právě, když  $(\delta^{A/\equiv_A})^*([r]_{\equiv_A}, x) \in F^{A/\equiv_A}$ .
- S využitím věty 29 pro každé  $x \in \Sigma^*$ :

$$[\delta^*(q, x)] \in F^{A/\equiv_A} \text{ právě tehdy, když } [\delta^*(r, x)]_{\equiv_A} \in F^{A/\equiv_A}$$

Následně aplikujeme větu 31:

$$\delta^*(q, x) \in F \text{ právě tehdy, když } \delta^*(r, x) \in F, \text{ tozn., že } q \equiv_A r \text{ a tedy } [q]_{\equiv_A} = [r]_{\equiv_A}$$

**Věta 33:**  $L(A) = L(A / \equiv_A)$ , což plyne užitím vět 31 a 29.

**Důkaz 31:** Máme dokázat, že  $\delta^* \in F$ , právě, když  $(\delta^{A/\equiv_A})([q_0]_{\equiv_A}, x) \in F^{A/\equiv_A}$ .

Z věty 29:  $\delta^*(q_0, x)$  právě, když  $[\delta^*(q_0, x)]_{\equiv_A} \in F^{A/\equiv_A}$ , to ale platí z věty 31.

**Důsledek 3:** Obecně platí, že  $L(A) \subseteq L(A/\Theta)$ .

**Věta 34:** Pokud je každý stav automatu  $A$  dosažitelný, pak má  $A / \equiv_A$  také každý stav dosažitelný.

### 16.3. Algoritmus pro hledání redukovaného automatu

1. Označíme  $A / \equiv_A$  jako  $A^R$ , konstruujeme posloupnost rozkladů na  $Q$  tak, že  $\varphi_1, \varphi_2, \dots, \varphi_i = \varphi_{i+1}$ , kde  $\varphi_1 = \{F, Q \setminus F\}$

2. Při odvození rozkladů  $\varphi_1$  z  $\varphi_{i-1}$  postupujeme následovně:

- (a) Vezmeme libovolný stav  $r \in R$ .
- (b) Položíme  $S = \{s \in R \mid \text{pro každé } a \in \Sigma \text{ platí také, že } \delta(S, a) \in [\delta(R, a)]\}$ .
- (c) Pokud  $S = R$ , pak vložíme  $R$  do  $\varphi_i$ , pokud  $S \subsetneq R$ , pak vložíme  $S$  a  $R \setminus S$  do  $\varphi_i$ .

**Věta 35:** Korektnost algoritmu pro nalezení  $\equiv_A$ : pokud  $\varphi_i = \varphi_{i+1}$ , pak  $\varphi_i = Q / \equiv_A$ .

**Důkaz 32:** Ověříme, že  $q \equiv_A r$  právě, když  $q \in [r]_{\varphi_i}$ .

Pokud  $q \in [r]_{\varphi_i}$ , pak z toho, jak jsme zavedli posloupnost rozkladů, plyne, že nemůže platit  $q \equiv_A r$ .

Pokud  $q \equiv_A r$ , pak  $q \in [r]_{\equiv_A}$  a zbývá dokázat opačnou implikaci, což provedeme indukcí přes délku řetězce:

- Máme libovolná  $q, r \in Q$ .
  - Pokud  $q \in [r]_{\varphi_i}$ , pak  $\delta^*(q, x) \in F$  právě, když  $\delta^*(r, x) \in F$ .
1. Pro  $x = \varepsilon$  je situace triviální. Máme dokázat, že  $q \in F$  právě, když  $r \in F$ , ale to obecně platí, protože  $\varphi_i = \{F, Q \setminus F\}$
  2. Pro  $x \in \Sigma^*$ ,  $x = au$ , kde  $a \in \Sigma, u \in \Sigma^*$ . Z předpokladu, že  $q \in [r]_{\varphi_i}$  a  $\varphi_i = \varphi_{i+1}$  platí, že  $[\delta(q, a)]_{\varphi_i} = [\delta(r, a)]_{\varphi_i}$ , to jest  $\delta(q, a) \in [\delta(r, a)]_{\varphi_i}$  - zde použijeme indukční předpoklad.

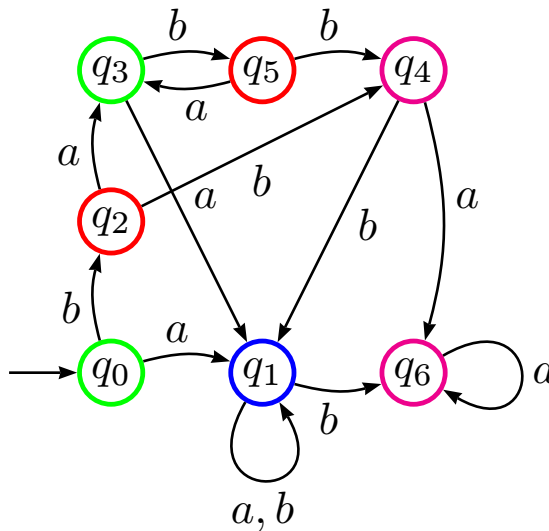
$$\delta^*(\delta(q, a)r) \in F \text{ právě tehdy, když } \delta^*(q, x) \in F$$

$$\delta^*(\delta(r, a)r) \in F \text{ právě tehdy, když } \delta^*(r, x) \in F$$

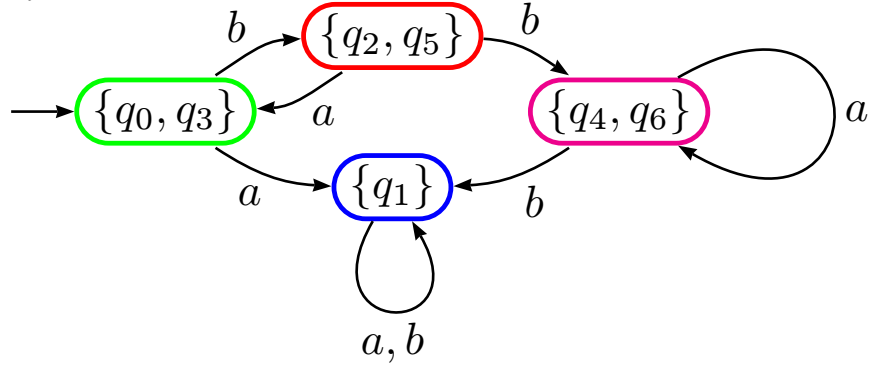
Tvrzení dokázáno pro každé  $x$ , tedy  $q \equiv_A r$ .

**Příklad 34:** Mějme výraz  $b(ab)^*ba^*$ . Dále

$$\begin{aligned} \varphi_1 &= \{\{q_0, q_1, q_2, q_3, q_5\}, \{q_4, q_6\}\} \\ \varphi_2 &= \{\{q_0, q_1, q_3\}, \{q_2, q_5\}, \{q_4, q_6\}\} \\ \varphi_3 &= \{\{q_0, q_3\}, \{q_1\}, \{q_2, q_5\}, \{q_4, q_6\}\} \end{aligned}$$



Následně rozklady:



	$q_0$	$q_1$	$q_2$	$q_3$	$q_4$	$q_5$	$q_6$
$q_0$	*	X	X	*	X	X	X
$q_1$	-	*	X	X	X	X	X
$q_2$	-	-	*	X	X	*	X
$q_3$	-	-	-	*	X	X	X
$q_4$	-	-	-	-	*	X	*
$q_5$	-	-	-	-	-	*	X
$q_6$	-	-	-	-	-	-	*

Tabulka 3. Tabulkový popis rozkladů stavů

V tabulce 3. se objevují znaky **X** na pozicích, pro které platí, že:

$$T[q, x] \quad \text{kde platí} \quad q \in F, x \notin F \\ \text{nebo} \quad q \notin F, x \in F$$

Projdeme prázdná místa,  $T[q, r] = \text{„prázdnno, pokud“ } \exists a \in \Sigma \text{ tak, že:}$

- buď  $T[\delta(q, a), \delta(r, a)] = X$
- nebo  $T[\delta(r, a), \delta(q, a)] = X$

Pokud ano, tak „X“ na pozici  $[q, r]$ .

## 17. Izomorfismus automatů

Pro KDA  $A_1$  a  $A_2$  mějme zobrazení

$$h : Q_1 \rightarrow Q_2$$

a toto zobrazení označme jako **izomorfismus**, pokud platí:

1. Zobrazení  $h$  je bijektivní.
2. Počáteční stav automatu  $A_1$  e zobrazí na počáteční stav automatu  $A_2$ .
3. Pro všechna  $q \in Q$  platí, že  $Q \in F_1$  právě, pokud  $q \in F_2$ .
4. Zobrazení  $h$  je kompatibilní s přechodovou funkcí.

**Věta 36:** Jsou-li dva automaty izomorfní, pak  $L(A_1) = L(A_2)$ .

**Definice 22:** Mějme regulární jazyk  $L$ , pak KDA s úplnou přechodovou funkcí je **minimálním automatem** pro  $L$ , pokud  $L(A) = L$  a pro každý KDA  $B$  takový, že  $L(B) = L$  platí, že  $B$  má tolik stavů jako  $A$ .

**Věta 37:** Je-li automat  $A$  minimální pro jazyk  $L$ , pak je **redukovaný** a nemá nedosažitelné stavy.

**Věta 38:** Pokud jsou automaty  $A, B$  KDA bez nedosažitelných stavů a pokud jsou tyto automaty navíc redukované a existuje stejný jazyk, který generují, pak jsou izomorfní.

**Věta 39:**  $A$  je minimální automat pro jazyk  $L$  právě tehdy, když  $A$  neobsahuje nedosažitelné stavy a je redukovaný.

## 18. Bezkontextové gramatiky

V této části se vrátíme k problematice bezkontextových gramatik. Před dalším čtením je potřeba ovládat základní pojmy, zejména

- Bezkontextové gramatiky
- Bezkontextové jazyky
- P-derivace
- Odvozování řetězců
- Věty a větné formy

Navíc si zavedeme duální pojem k derivaci - **redukci**.

**Definice 23:** Řetězec  $v$  lze redukovat na řetězec  $u$ , pokud  $u \Rightarrow_G^* v$ . Značíme  $v \Leftarrow_G^* u$ .

V následujícím příkladě si ukážeme problém nejednoznačnosti bezkontextových gramatik.

**Příklad 35:** Mějme gramatiku:

$$\begin{aligned} G &= \langle N, \Sigma, P, S \rangle \\ \Sigma &= \{a, b, c, 0, 1, +, -, *, /, (, )\} \\ N &= \{S, E, C, V\} \\ P &= \{ \begin{array}{l} S \rightarrow E, \\ E \rightarrow E * E \mid E / E \mid E + E \mid E - E \mid - E \mid C \mid V \mid (E), \\ C \rightarrow 0C \mid 1C \mid 0 \mid 1, \\ V \rightarrow aV \mid bV \mid cV \mid a \mid b \mid c \end{array} \} \end{aligned}$$

Uvažujme větu  $w = ac * 1 - c$ . K ní se lze dostat buď:

$$S \Rightarrow_G E \Rightarrow_G E * E \Rightarrow_G V * E \Rightarrow_G V * E - E \Rightarrow_G V * E - V \Rightarrow_G aV * E - V \Rightarrow_G ac * E - V \Rightarrow_G ac * C - V \Rightarrow_G ac * C - c \Rightarrow_G \mathbf{ac*1-c}$$

nebo

$$S \Rightarrow_G E \Rightarrow_G E * E \Rightarrow_G V * E \Rightarrow_G aV * E \Rightarrow_G ac * E \Rightarrow_G ac * E - E \Rightarrow_G ac * C - E \Rightarrow_G ac * 1 - E \Rightarrow_G ac * 1 - V \Rightarrow_G \mathbf{ac*1-c}$$

Ačkoliv odvozujeme stejnou větu, můžeme zde pozorovat jakousi nejednoznačnost, způsobenou tím, že neterminály derivujeme v libovolném pořadí.

Tento problém bychom mohli vyřešit použitím tzv. lineární bezkontextové gramatiky, tedy takové, jejíž odvozovací pravidla obsahují pouze jeden neterminální symbol na pravé straně.

**Příklad 36:** Lineární gramatika může vypadat např. takto:

$$\begin{aligned} G &= \langle N, \Sigma, P, S \rangle \\ P &= \{ \begin{array}{l} S \rightarrow abB, \\ A \rightarrow aaBb|\varepsilon, \\ B \rightarrow bbAa \end{array} \} \\ L(G) &= \{ ab(bba)^n bba (ba)^n \mid n \geq 0 \} \end{aligned}$$

Ovšem k nejednoznačnosti může stejně dojít, pokud by se z různých neterminálních symbolů daly odvodit stejná slova.

**Příklad 37:** Příklad nejednoznačné lineární gramatiky:

$$\begin{aligned} G &= \langle N, \Sigma, P, S \rangle \\ P &= \{ \begin{array}{l} S \rightarrow aA \mid aB, \\ A \rightarrow bA|a, \\ B \rightarrow bB|a \end{array} \} \end{aligned}$$

Uvažujme slovo  $w = abba$ , ke kterému lze dojít buď:

$$S \Rightarrow_G aA \Rightarrow_G abA \Rightarrow_G abbA \Rightarrow_G abba$$

nebo:

$$S \Rightarrow_G aB \Rightarrow_G abB \Rightarrow_G abbB \Rightarrow_G abba$$

Mějme bezkontextovou gramatiku  $G = \langle N, \Sigma, P, S \rangle$ .

**Definice 24:** P-derivace  $x_0, \dots, x_k$  se nazývá **nejlevější** P-derivace, pokud pro každé  $i \in \{1, \dots, k\}$  platí, že  $x_{i-1}$  je ve tvaru  $uAv$ , kde  $u \in \Sigma^*$ ,  $A \in N$ ,  $v \in (\Sigma \cup N)^*$  a  $x_i$  je ve tvaru  $uyv$ , kde  $A \rightarrow y \in P$ .

**Poznámka 14:** Řetězci  $u$  se říká uzavřená forma a řetězci  $Av$  otevřená forma (včetně formy  $uAv$ ). Odvození pomocí nejlevější derivace značíme  $u \Rightarrow_{G,l} v$ .

**Věta 40:** Mějme  $v \in \Sigma^*$  a  $X \in N$ . Pokud existuje P-derivace  $X, \dots, v$ , pak existuje nejlevější P-derivace  $X, \dots, v$  používající stejnou množinu pravidel jako výchozí P-derivace.

**Důkaz 33:** Tvrzení dokážeme indukcí přes délku P-derivace

- Pro délku 0 platí triviálně.
- Předpokládejme, že tvrzení platí pro P-derivaci délky  $\leq n$ .

Uvažujme P-derivaci délky  $n+1$ , ve tvaru  $x_0, x_1, \dots, x_{n+1}$ . Pokud  $x_1$  vzniklo z  $x_0$  použitím pravidla  $X \rightarrow w_0 X_{i_1} w_1 X_{i_2} \dots X_{i_k} w_k$  kde  $w_0, \dots, w_j \in \Sigma^*$ ,  $X_{i_j} \in N$ ,  $1 \leq j \leq k$  pak  $X_{n+1}$  je ve tvaru  $w_0 u_1 w_1 u_2 w_2 \dots w_k$  tak, že  $x_{i_j} \Rightarrow_G^* u_j$ .

To znamená, že existují P-derivace délek  $\leq n$ :  $X_{i_j}, \dots, u_j$

Z indukčního předpokladu: existují nejlevější P-derivace  $X_{i_j}, \dots, u_j$  používající stejnou množinu pravidel.

Dále platí:

$$\begin{aligned}
X &\Rightarrow_{G,l} w_0 X_{i_1} w_1 X_{i_2} \dots X_{i_k} w_k \\
&\Rightarrow_{G,l} w_0 u_1 w_1 X_{i_2} \dots X_{i_k} w_k \\
&\Rightarrow_{G,l} w_0 u_1 w_1 u_2 \dots X_{i_k} w_k \\
&\dots \\
&\Rightarrow_{G,l} w_0 u_1 w_1 u_2 w_2 \dots u_k w_k
\end{aligned}$$

To jest, existuje nejlevější P-derivace  $X, \dots, w_0 \dots w_k$

**Příklad 38:** Uvažujme gramatiku z příkladu 35

$S \Rightarrow_G^* E + C$  tady problém není

$S \Rightarrow_{G,l}^* E + C$  tento výraz už smysl nedává. Pomocí nejlevější derivace bychom zákonitě museli nejdříve derivovat  $E$  na levé straně výrazu  $E + E$

**Poznámka 15:** Lze zavést duálně nejpravější derivaci.

**Příklad 39:** Zase se odkážeme na příklad 35.

Výraz  $10 + (ca * 110)$  má jedinou nejlevější derivaci

Naopak  $a + 10 * c$  jich má několik:

$$\begin{aligned}
S &\Rightarrow_{G,l} E \Rightarrow_{G,l} E + E \Rightarrow_{G,l} V + E \Rightarrow_{G,l} a + E \Rightarrow_{G,l} a + E * E \Rightarrow_{G,l} a + C * E \Rightarrow_{G,l} a + 1C * E \Rightarrow_{G,l} \\
&a + 10 * E \Rightarrow_{G,l} a + 10 * V \Rightarrow_{G,l} a + 10 * c
\end{aligned}$$

nebo:

$$S \Rightarrow_{G,l} E \Rightarrow_{G,l} E * E \Rightarrow_{G,l} \dots \Rightarrow_{G,l} a + 10 * c \text{ (už ve druhém odvození můžeme pozorovat rozdíl)}$$

Je to způsobeno jinou nejednoznačností než tou, kterou jsme eliminovali pomocí nejlevější derivace.

### 18.1. Derivační stromy

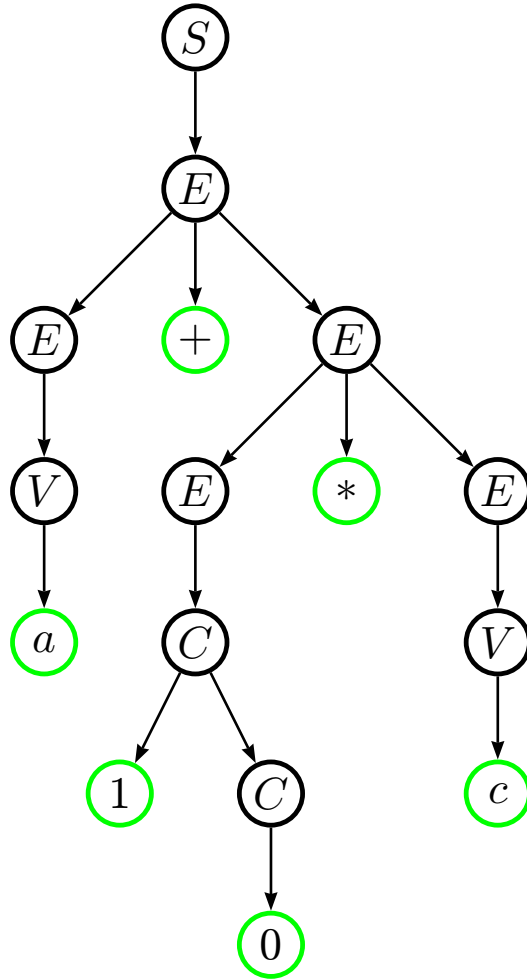
Slouží ke grafickému znázornění nejlevějších derivací. Mějme gramatiku  $G = \langle N, \Sigma, P, S \rangle$ .

**Definice 25:** Derivačním stromem slova  $x \in (\Sigma \cup N)^*$  z  $X \in N$  podle pravidel z  $G$  je každý strom splňující:

1. každý vnitřní uzel je ohodnocen neterminálem
2. kořen je ohodnocen  $X$
3. pokud je vnitřní uzel označený  $A \in N$  a jeho potomci jsou zleva doprava označeni  $X_1, \dots, X_k \in N \cup \Sigma$  pak existuje pravidlo  $A \rightarrow X_1 \dots X_k \in P$
4. zřetěžením hodnot v listech při průchodu stromu do hloubky a zleva doprava získáme řetězec  $x$

**Příklad 40:** Opět pracujeme s gramatikou z příkladu 35.

$S \Rightarrow_{G,l}^* a + 10 * c$  Derivační strom bude vypadat následovně:



**Věta 41:** Pokud  $X \Rightarrow_{G,l}^* x$  pak existuje derivační strom  $x$  z  $X$  podle  $G$ .

**Důkaz 34:** Tvrzení dokážeme jak jinak, než indukcí přes délku P-derivace.

Předpokládejme, že tvrzení platí pro derivace délky  $n$  a méně.

Mějme derivaci  $X = x_0, \dots, x_{n+1}$  délky  $n + 1$  pak existuje pravidlo  $x \rightarrow x_1 \in P$

$x_1 = w_0 X_1 w_1 X_2 w_2 \dots X_k w_k$  kde  $w_i \in \Sigma^*$ ,  $X_i \in N$

$x_{n+1}$  je ve tvaru  $w_0 u_1 w_1 u_2 \dots u_k w_k$

$X_i \Rightarrow_{G,l}^* u_i \dots$  délky nejvýše  $n$ . Z indukčního předpokladu vyplývá, že existují derivační stromy  $X_i \Rightarrow_{G,l}^* u_i$ .

**Věta 42:** Pokud existuje derivační strom  $x$  z  $X$  podle  $G$ , pak  $X \Rightarrow_{G,l}^* x$

**Důkaz 35:** Dokážeme indukcí přes výšku stromu.

- Pro 0 je to jasné

- Předpokládejme, že tvrzení platí pro stromy výšky  $n$  a méně.

Mějme strom výšky  $n + 1$ , podle tohoto víme, jak vypadal první krok derivace  $X \rightarrow w_0 X_0 w_1 \dots$

Z indukčního předpokladu existují derivace  $X_i \Rightarrow_{G,l}^* u_i$  a proto

$$\begin{aligned} X &\Rightarrow_{G,l} w_0 X_1 w_1 X_2 \dots X_{i_k} w_k \\ &\Rightarrow_{G,l} w_0 u_1 w_1 X_2 \dots X_{i_k} w_k \\ &\dots \\ &\Rightarrow_{G,l} w_0 u_1 w_1 u_2 w_2 \dots u_k w_k \end{aligned}$$

**Věta 43:**  $S \Rightarrow_{G,l}^* u$  právě když existuje derivační strom  $u$  z  $S$  podle  $G$ .

## 18.2. Jednoznačné a nejednoznačné bezkontextové gramatiky

**Definice 26:** Bezkontextová gramatika  $G = \langle N, \Sigma, P, S \rangle$  se nazývá **nejednoznačná**, pokud existuje věta  $x \in L(G)$ , která má více než jeden derivační strom z  $S$  podle  $G$ . V opačném případě se nazývá **jednoznačná** bezkontextová gramatika.

**Definice 27:** Bezkontextový jazyk  $L$  se nazývá **jednoznačný** pokud existuje jednoznačná gramatika  $G$  tak, že  $L = L(G)$ .

**Definice 28:** Bezkontextový jazyk  $L$  se nazývá **inherentně nejednoznačný**, pokud neexistuje žádná jednoznačná gramatika  $G$  taková, že  $L = L(G)$ .

**Věta 44:** Každý regulární jazyk je jednoznačný.

**Důkaz 36:** Pokud  $L$  je regulární, pak existuje KDA  $A$  s úplnou přechodovou funkcí  $\delta$ . Pak pro  $A$  lze sestavit gramatiku  $G$  tak, že bude platit  $L(A) = L(G)$ .

Z toho jak byla gramatika  $G$  konstruována plyne, že pro každý  $A \in N$  a  $a \in \Sigma$  existuje nejvýše jedno pravidlo  $A \rightarrow aB$ . To jest, pro  $x \in L$  existuje právě jedna P-derivace.

**Důkaz 37:**

**Příklad 41:**  $L = \{a^n b^n c^p d^p \mid n, p \geq 0\}$  je inherentně nejednoznačný.

## 18.3. Uzávěrové vlastnosti bezkontextových jazyků

- **Sjednocení**

Mějme dány bezkontextové jazyky  $L_1, L_2$  a korespondující bezkontextové gramatiky  $G_1$  a  $G_2$ . Lze předpokládat, že množiny stavů obou gramatik jsou disjunktní (tj.  $N_1 \cap N_2 = \emptyset$ ).

Dále uvažujeme gramatiku  $G = \langle N, \Sigma, P, S \rangle$ , kde

$$N = N_1 \cup N_2 \cup \{S\}$$

$$\Sigma = \Sigma_1 \cup \Sigma_2$$

$$P = P_1 \cup P_2 \cup \{S \rightarrow S_1 \mid S_2\}$$

Platí, že  $L(G) = L(G_1) \cup L(G_2)$ .

- **Produkt**

$L_1$  a  $L_2$  jsou bezkontextové jazyky,  $G_1$  a  $G_2$  jsou odpovídající bezkontextové gramatiky.



Zavedeme gramatiku  $G = \langle N, \Sigma, P, S \rangle$ , kde  
 $N = N_1 \cup N_2 \cup \{S\}$   
 $\Sigma = \Sigma_1 \cup \Sigma_2$   
 $P = P_1 \cup P_2 \cup \{S \rightarrow S_1 S_2\}$   
 Platí, že  $L(G) = L(G_1) \cdot L(G_2)$ .

- **Kleeneho uzávěr**

Mějme bezkontextový jazyk  $L$  a k němu odpovídající gramatiku  $G$ .

Vytvoříme novou gramatiku  $G' = \langle N \cup \{S'\}, \Sigma, P \cup \{S' \rightarrow \varepsilon \mid SS'\}, S' \rangle$ .

- **Pozitivní uzávěr**

Podobné jako u Kleeneho uzávěru, akorát výsledná gramatika se bude lišit v množině pravidel.

$G' = \langle N \cup \{S'\}, \Sigma, P \cup \{S' \rightarrow S \mid SS'\}, S' \rangle$

- **Průnik**

$\mathcal{L}_2$  není uzavřená na průnik.

**Příklad 42:** Uvedeme si příklad, který nám tvrzení potvrdí.

$\Sigma = \{a, b, c\}$

$L_1 = \{a^n b^n c^* \mid n \geq 0\}$

$L_2 = \{a^* b^n c^n \mid n \geq 0\}$

$L_1 \cap L_2 = \{a^n b^n c^n \mid n \geq 0\}$

Z toho také plyne, že není uzavřená ani na Komplement či Množinový rozdíl (De Morganovy zákony).

## 18.4. Bezkontextové gramatiky v programátorské praxi

**BNF**(Backus-Naurova forma): Vytvořili ji John Backus a Peter Naur. Zápis se řídí několika pravidly:

- neterminální symboly se zapisují do  $\langle \dots \rangle$
- používá se  $::=$  místo  $\rightarrow$
- terminální symboly se píší do uvozovek (" \* ")
- pravidla se oddělují pomocí |

**Příklad 43:** Uvedeme si příklad zápisu gramatiky pomocí BNF.

$\langle expr \rangle ::= "(" \langle expr \rangle ")" \mid \langle expr \rangle \langle op \rangle \langle expr \rangle \mid \langle number \rangle$

$\langle op \rangle ::= "+" \mid "-" \mid "*" \mid "/"$

$\langle number \rangle ::= \langle digit \rangle \langle number \rangle \mid \langle digit \rangle$

$\langle digit \rangle ::= "0" \mid \dots \mid "9"$

**Extended BNF:** Zavedl Niklaus Wirth v roce 1977. Odpovídá BNF, pouze zjednodušuje její notaci a to tak že:

- každé pravidlo je ukončeno znakem ";"

- terminální symboly se píší do uvozovek, neterminální ne
- $[, ]$  značí volitelnou část výrazu
- $\{, \}$  indukují možnost opakování
- $(, )$  jsou použity pro shlukování výrazů
- místo  $::=$  se používá  $=$
- terminální a neterminální symboly se oddělují čárkou

**Příklad 44:** Gramatiku z předchozího příkladu můžeme zapsat pomocí EBNF takto:

```
expr = "(" , expr , ")" | expr , op , expr | number;  
op = "+" | "-" | "*" | "/";  
number = [signum] , digit , {digit};  
digit = "0" | ... | "9";  
signum = "+" | "-";
```