

# Stručný úvod do problematiky jazyků a gramatik

doc. Ing. Roman Šenkeřík, Ph.D.

Univerzita Tomáše Bati ve Zlíně

2021

Teorie formálních jazyků:

- Důležitá oblast teoretické informatiky
- Základy položil Noam Chomský v roce 1956
- Obsahuje formální jazyky a formální modely výpočtu

- 1 Formální jazyky a gramatiky
  - Základní terminologie
  - Gramatiky
  - Hierarchie jazyků
  - Stromy odvození

# Analogie

Pohled na definici jazyka a gramatiky v souvislosti s lidskou řečí:

## 1. analogie

- Základem jazyka jsou písmena abecedy.
- Písmena skládají slova, slova věty, atd.
- Ne všechny kombinace písmen jsou slova.
- Jazyk je určen množinou slov jazyka - pravidly pro tvorbu slov.
- Pravidlům pro tvorbu slov říkáme **gramatika**.

## 2. analogie

- Základem jazyka jsou slova dále nedělitelná.
- Místo z abecedy vycházíme ze slovníku.
- Gramatika určuje pravidla sestavování vět.
- Jazyk tvoří všechny věty, které lze vytvořit.

# Abeceda

## Abeceda[Slovník]

- Libovolná neprázdná konečná množina znaků
- Obvykle se vyžadují min. 2 znaky
- Příklady abeced:
  - $\{A, B, C, \dots, Z\}, \{\alpha, \beta, \gamma, \dots, \omega\}, \{0, 1\}$
  - $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, +, -, *, /\}$
  - $\{begin, end, if, then, else, while, repeat, until, for, read, write, \dots\}$

# Slovo

## Slovo[věta]

- Libovolná konečná, případně i prázdná posloupnost
- Řetězec znaků abecedy
- Prázdné slovo  $\varepsilon$
- Značení:
  - $V$  abeceda
  - $V^*$  množina všech slov
  - $V^+$  množina všech neprázdných slov  $V^+ = V^* \div \{\varepsilon\}$
- Příklady:
  - $V = \{0, 1\}, V^* = \{\varepsilon, 0, 1, 00, 01, 000, 001, 010, \dots\},$   
 $V^+ = \{1, 00, 01, 000, 001, 010, \dots\}$
  - $V = \{a, b, c\}, V^* = \{\varepsilon, a, b, c, aa, ab, ac, ba, \dots\}$

# Jazyk

*Je-li dána abeceda  $V$ , potom libovolná podmnožina množiny  $V^*$  všech slov nad touto abecedou se nazývá **jazyk**.*

$$L \subseteq V^*$$

- Příklady vymezení jazyka

- Množina slov zadané délky. BYTE (abeceda  $V = \{0, 1\}$ , délka 8 - jazyk má 256 slov)
- Množina slov nad abecedou  $V = \{0, 1\}$ , kde počet jedniček je prvočíslo
- Množina všech slov nad libovolnou abecedou, která jsou schodná se slovy vytvořenými opačným pořadím znaků ve slově.
- Množina syntakticky správně vytvořených programů daného programovacího jazyka.

# Gramatiky

## Gramatika

- Systém jak pomocí přepisovacích pravidel vytvořit všechna slova daného jazyka z počátečního symbolu
- Počáteční symbol
  - Počáteční symbol nepatří do abecedy jazyka (terminál)
- Terminální symbol - prvek abecedy jazyka
- Neterminální symbol - proměnná (prvek abecedy proměnných), dále se nahrazuje (za další terminální nebo neterminální symbol)



# Gramatiky

## Příklad analytické gramatiky

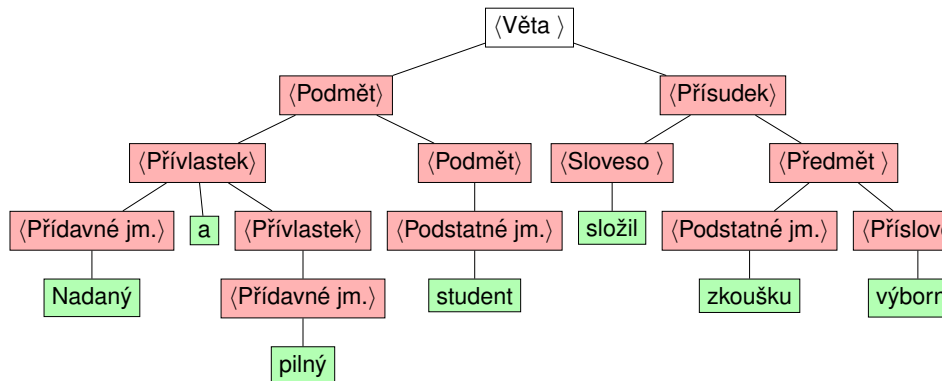
Používá 2. analogii!

- ① ⟨Věta⟩ → ⟨Podmět⟩⟨Přísudek⟩
- ② ⟨Podmět⟩ → ⟨Přívlastek⟩⟨Podmět⟩
- ③ ⟨Přívlastek⟩ → ⟨Přídavné jm.⟩ **a** ⟨Přívlastek⟩
- ④ ⟨Přívlastek⟩ → ⟨Přídavné jm.⟩
- ⑤ ⟨Podmět⟩ → ⟨Podstatné jm.⟩
- ⑥ ⟨Přísudek⟩ → ⟨Sloveso⟩⟨Předmět⟩
- ⑦ ⟨Předmět⟩ → ⟨Podstatné jm.⟩⟨Příslovce⟩
- ⑧ ⟨Přídavné jm.⟩ → **Nadaný**
- ⑨ ⟨Přídavné jm.⟩ → **pilný**
- ⑩ ⟨Podstatné jm.⟩ → **student**
- ⑪ ⟨Podstatné jm.⟩ → **zkoušku**
- ⑫ ⟨Sloveso⟩ → **složil**
- ⑬ ⟨Příslovce⟩ → **výborně**

# Gramatiky

## Příklad analytické gramatiky

Používá 2. analogii!



# Gramatiky - formální zápis

Generativní gramatika je uspořádaná čtveřice  $G=(V_n, V_t, P, S)$

- $V_n$  - neprázdná konečná množina neterminálních znaků
- $V_t$  - neprázdná konečná množina terminálních znaků - abeceda
- $P$  - neprázdná konečná množina přepisovacích pravidel
- $S$  - vybraný počáteční symbol

Gramatiky lze rozdělovat podle typu - **Chomského hierarchie gramatik**

# Gramatika typu 0

- Obsahuje všechny gramatiky a jimi generované jazyky  $L_0$
- Ne každý jazyk lze generovat nějakou gramatikou - existují jazyky které nejsou ani typu 0

# Gramatika typu 1

- **Kontextová gramatika**
- Levá strana produkčních pravidel obsahuje definice v kontextu
- Každé produkční slovo musí být typu  $\alpha X \beta \rightarrow \alpha \gamma \beta$ 
  - $\alpha, \beta \in (V_n \cup V_t)^*$
  - $X$  - jeden neterminální symbol
  - $\gamma \in (V_n \cup V_t)^+$  - neprázdný řetězec
- Příklad:
  - $G = (\{A, S\}, \{0, 1\}, P, S)$
  - $S \rightarrow 0A | \varepsilon$
  - $0A \rightarrow 00A1 (\alpha = 0, \beta = \varepsilon, X = A, \gamma = 0A1)$
  - $A \rightarrow 1$

# Gramatika typu 2

- **Bezkontextová gramatika**

- Na levé straně se nacházejí pouze neterminální symboly

- Každé produkční slovo musí být typu  $X \rightarrow \gamma$

- $X$  - jeden neterminální symbol

- $\gamma \in (V_n \cup V_t)^*$

- Příklad

- $G = (\{S\}, \{0, 1\}, P, S)$

- $S \rightarrow 0S1 | \varepsilon$

# Gramatika typu 3

- **Regulární gramatika**

- Každé produkční slovo musí být typu  $A \rightarrow aB$  nebo  $A \rightarrow a$

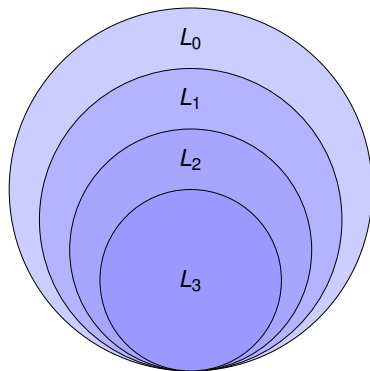
- $A, B$  - neterminální symboly
- $a$  - terminální symbol

- **Příklad**

- $G = (\{A, B\}, \{a, b, c\}, P, S)$
- $A \rightarrow aaB \mid ccB$
- $B \rightarrow bB \mid \varepsilon$

# Hierarchie jazyků podle Chomského

- $L_0$  - všechny jazyky
- $L_1$  - kontextové
- $L_2$  - bezkontextové
- $L_3$  - regulární





# Gramatika

## Příklad na jinou formu zápisu gramatiky

Bude použita **Backus - Naurova forma** (BNF). Neterminální symboly se zapisují do špičatých závorek. „Svislítko“ znamená „nebo“ a šipka v přepisovacím pravidle je nahrazena trojznakem „::=“.

Pravidla pro zápis čísel typu REAL

- $V_t = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, +, -, ., E\}$
- $V_n = \{\langle \text{Číslo} \rangle, \langle \text{Číslo bez znaménka} \rangle, \langle \text{Číslice} \rangle, \dots\}$
- $P = \{\langle \text{Neterminální symbol} \rangle ::= XX | YY \dots\}$
- $S \rightarrow \langle \text{Číslo} \rangle$

# Stromy odvození

Přepisovací pravidla:

- 1  $\langle \text{Číslo} \rangle ::= \langle \text{Číslo bez znaménka} \rangle \mid + \langle \text{Číslo bez znaménka} \rangle \mid - \langle \text{Číslo bez znaménka} \rangle$
- 2  $\langle \text{Číslo bez znaménka} \rangle ::= \langle \text{Desetinné číslo} \rangle \mid \langle \text{Exponentová část} \rangle \mid \langle \text{Desetinné číslo} \rangle \langle \text{Exponentová část} \rangle$
- 3  $\langle \text{Desetinné číslo} \rangle ::= \langle \text{Celé číslo bez znaménka} \rangle \mid \langle \text{Desetinná část} \rangle \mid \langle \text{Celé číslo bez znaménka} \rangle \langle \text{Desetinná část} \rangle$
- 4  $\langle \text{Exponentová část} \rangle ::= E \langle \text{Celé číslo} \rangle$
- 5  $\langle \text{Desetinná část} \rangle ::= . \langle \text{Celé číslo bez znaménka} \rangle$
- 6  $\langle \text{Celé číslo} \rangle ::= \langle \text{Celé číslo bez znaménka} \rangle \mid + \langle \text{Celé číslo bez znaménka} \rangle \mid - \langle \text{Celé číslo bez znaménka} \rangle$
- 7  $\langle \text{Celé číslo bez znaménka} \rangle ::= \langle \text{Číslice} \rangle \mid \langle \text{Celé číslo bez znaménka} \rangle \langle \text{Číslice} \rangle$
- 8  $\langle \text{Číslice} \rangle ::= \langle 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9 \rangle$

# Gramatika

