Chomského NF, Pumping Lemma pro CFL

Marta Vomlelová

MS 303

Bezkontextové gramatiky (CFG) v Chomského normální formě

- Chomského normální forma bezkontextové gramatiky:
  - neobsahuje zbytečné symboly (S není nikdy zbytečný)
  - všechna pravidla tvaru $A \to BC$ nebo $A \to a$, A, B, C jsou neterminály, a terminál,
    nebo $S \to \epsilon$ pro startovní symbol S, který se pak nesmí vyskytovat na pravé straně žádného pravidla.
  - Pro každý bezkontextový jazyk L existuje gramatika v Chomského normálním tvaru, která generuje L.

Postupně provedeme zjednodušení gramatiky, nejdřív:

- Eliminace *zbytečných symbolů*
- eliminace ϵ*–pravidel* $A \to \epsilon$; $A \in V$
- eliminace *jednotkových pravidel* $A \to B$ pro $A, B \in V$.

- To vše potřebujeme k formulaci iteračního lemmatu pro bezkontextové jazyky.
- A ověření, zda zadaný řetězec patří do jazyka gramatiky pomocí Cocke-Younger-Kasami algoritmu.

Eliminace zbytečných symbolů

Definice 1.35: zbytečný, užitečný, generující, dosažitelný symbol

- Startovní symbol S je užitečný vždy (i když negeneruje žádné $w \in T^*$).
- Nestartovní symbol X je **užitečný** v gramatice $G = (V, T, P, S)$ pokud existuje derivace tvaru $S \Rightarrow^* \alpha X \beta \Rightarrow^* w$ kde $w \in T^*, X \in (V \cup T)$, $\alpha, \beta \in (V \cup T)^*$.
- Pokud X není užitečný, říkáme, že je **zbytečný**.
- X je **generující** pokud $X \Rightarrow^* w$ pro nějaké slovo $w \in T^*$. Vždy $w \Rightarrow^* w$ v nula krocích.
- X je **dosažitelný** pokud $S \Rightarrow^* \alpha X \beta$ pro nějaká $\alpha, \beta \in (V \cup T)^*$.

Chceme eliminovat ne–generující a ne–dosažitelné symboly.

Příklad 1.41

Uvažujme gramatiku:
$S \to AB|a$
$A \to b$

Eliminujeme B (ne–generující):
$S \to a$
$A \to b$.

Eliminujeme A (nedosažitelný):
$S \to a$.

Lemma 1.4 (Eliminace zbytečných symbolů)

Nechť $G = (V, T, P, S)$ je CFG. Zkonstruujeme $G_1 = (V_1, T_1, P_1, S)$ následovně:

- Eliminujeme ne–generující symboly (kromě S) a pravidla je obsahující
- poté eliminujeme všechny nedosažitelné symboly

Pak G_1 nemá zbytečné symboly a $L(G_1) = L(G)$.

Algoritmus 1.9: Generující symboly

Základ Každý $a \in T$ je generující.

Indukce Pro každé pravidlo $A \to \alpha$, kde každý symbol v α je generující. Pak i A je generující.
(Včetně $A \to \epsilon$).

Algoritmus 1.10: Dosažitelné symboly

Základ S je dosažitelný.

Indukce Je–li A dosažitelný, pro všechna pravidla $A \to \alpha$ jsou všechny symboly v α dosažitelné.

Lemma 1.5 (generující/dosažitelné symboly)

Výše uvedené algoritmy najdou právě všechny generující / dosažitelné symboly.

Eliminace ϵ pravidel

Definice 1.36: nulovatelný neterminál

Neterminál A je **nulovatelný** pokud $A \Rightarrow^* \epsilon$.

Pro nulovatelné neterminály na pravé straně pravidla $B \to CAD$, vytvoříme dvě verze pravidla – s a bez nulovatelného neterminálu.

Algoritmus 1.11: Nalezení nulovatelných symbolů v G

Základ Pokud $A \to \epsilon$ je pravidlo G, pak A je nulovatelné.

Indukce Pokud $B \to C_1 \dots C_k$, kde jsou všechna C_i nulovatelná, je i B nulovatelné (terminály nejsou nulovatelné nikdy).

Algoritmus 1.12: Konstrukce gramatiky bez ϵ–pravidel z G

- Najdi nulovatelné symboly
- Pro každé pravidlo $A \to X_1 \dots X_k \in P, k \geq 1$, nechť m z X_i je nulovatelných. Nová gramatika G_1 bude mít 2^m verzí tohoto pravidla s/bez každého nulovatelného symbolu kromě ϵ v případě $m = k$.

Algoritmus 1.13: $S \to \epsilon$

Pokud S je startovní symbol a gramatika obsahuje pravidlo $S \to \epsilon$, zavedeme nový počáteční symbol S_1, přidáme $S_1 \to S|\epsilon$, toto pravidlo necháváme a s $S \to \epsilon$ eliminujeme.

Příklad 1.42

Mějme gramatiku $(\{A, B, S\}, \{a, b\}, P, S)$:

$$P = \left\{ \begin{array}{l} S \to AB \\ A \to AA|B|\epsilon \\ B \to bBB|\epsilon \end{array} \right\}.$$

- Pro každé pravidlo přidáme verze s/bez nulovatelných symbolů kromě ϵ.

$$\left\{ \begin{array}{l} S \to AB|A|B \\ A \to AA|A|A|B \\ B \to bBB|bB|bB|b \end{array} \right\}.$$

- Výsledná gramatika:

$$\left\{ \begin{array}{l} S_1 \to S|\epsilon \\ S \to AB|A|B \\ A \to AA|A|B \\ B \to bBB|bB|b \end{array} \right\}.$$

Eliminace jednotkových pravidel

Definice 1.37: jednotkové pravidlo

Jednotkové pravidlo je $A \to B \in P$ kde A, B jsou oba neterminály.

Idea: Rodič iterativně kopíruje pravou stranu dítěte jednotkového pravidla

Příklad 1.43 (Eliminace jednotkových pravidel)

$$\left\{\begin{array}{l} I \to a|b|Ia|Ib|I0|I1 \\ F \to I|(E) \\ T \to F|T*F \\ E \to T|E+T \end{array}\right\}$$

Expanze T v $E \to T$

$E \to F|T*F$

Expanze $E \to F$

$E \to I|(E)$

Expanze $E \to I$

$E \to a|b|Ia|Ib|I0|I1$

Dohromady: $E \to a|b|Ia|Ib|I0|I1|(E)|T*F|E+T$.

Dále se musíme vyhnout možným cyklům.

Definice 1.38: jednotkový pár

Dvojici $A, B \in V$ takovou, že $A \Rightarrow^* B$ pouze jednotkovými pravidly nazýváme **jednotkový pár** (jednotková dvojice).

Algoritmus 1.14: Nalezení jednotkových párů

Základ (A, A) pro každý $A \in V$ je jednotkový pár.

Indukce Je–li (A, B) jednotkový pár a $(B \to C) \in P$, pak (A, C) je jednotkový pár.

Příklad 1.44 (Jednotkové páry z předchozí gramatiky)

$(E, E), (T, T), (F, F), (I, I), (E, T), (E, F), (E, I), (T, F), (T, I), (F, I)$.

Algoritmus 1.15: Eliminace jednotkových pravidel z G

- najdi všechny jednotkové páry v G
- pro každý jednotkový pár (A, B) dáme do nové gramatiky všechna pravidla $A \to \alpha$ kde $B \to \alpha \in P$ a $B \to \alpha$ není jednotkové pravidlo.

Příklad 1.45

$$\left\{\begin{array}{l} I \to a|b|Ia|Ib|I0|I1 \\ F \to I|(E) \\ T \to F|T * F \\ E \to T|E + T \end{array}\right\} \text{ převedeme } \left\{\begin{array}{l} I \to a|b|Ia|Ib|I0|I1 \\ F \to (E)|a|b|Ia|Ib|I0|I1 \\ T \to T * F|(E)|a|b|Ia|Ib|I0|I1 \\ E \to E + T|T * F|(E)|a|b|Ia|Ib|I0|I1 \end{array}\right\}$$

Gramatiky v normálním tvaru

Lemma 1.6 (Gramatika v normálním tvaru, redukovaná)

Mějme bezkontextovou gramatiku G,pak existuje CFG G_1 taková že $L(G_1) = L(G)$ a G_1 neobsahuje ϵ–pravidla kromě $S \to \epsilon$, neobsahuje jednotková pravidla ani zbytečné symboly. Gramatika G_1 se nazývá **redukovaná**.

Redukce bezkontextové gramatiky.

Idea důkazu:

- Začneme eliminací ϵ–pravidel.
- Eliminujeme jednotková pravidla. Tím nepřidáme ϵ–pravidla.
- Eliminujeme zbytečné symboly. Tím nepřidáme žádná pravidla.
  - Nejdříve negenerující (kromě počátečního symbolu S),
  - pak nedosažitelné.

Definice 1.39: Chomského normální tvar

O bezkontextové gramatice $G = (V, T, P, S)$ bez zbytečných symbolů kde jsou všechna pravidla v jednom ze tří tvarů

- $A \to BC$, $A, B, C \in V$,
- $A \to a$, $A \in V$, $a \in T$,
- případně $S \to \epsilon$, pak S není na pravé straně žádného pravidla,

říkáme, že je v **Chomského normálním tvaru (ChNF)**.

Potřebujeme dva další kroky:

- pravé strany délky 2 a více předělat na samé neterminály
- rozdělit pravé strany délky 3 a více neterminálů na více pravidel

Algoritmus 1.16: neterminály

- Pro každý terminál a vytvoříme nový neterminál, řekněme A,
- přidáme pravidlo $A \to a$,
- použijeme A místo a na pravé straně pravidel délky 2 a více.

Algoritmus 1.17: rozdělení pravidel

- Pro pravidlo $A \to B_1 \dots B_k$ zavedeme $k-2$ neterminálů C_i
- Přidáme pravidla $A \to B_1 C_1, C_1 \to B_2 C_2$, $\dots, C_{k-2} \to B_{k-1} B_k$.

Věta 1.17: Chomského normální tvar bezkontextové gramatiky

Mějme bezkontextovou gramatiku G.Pak existuje CFG G_1 v Chomského normálním tvaru taková, že $L(G_1) = L(G)$.

Příklad 1.46

$$\{I \to a|b|Ia|Ib|I0|I1, F \to I|(E), T \to F|T * F, E \to T|E + T\}$$

$$\left\{\begin{array}{l} I \to a|b|IA|IB|IZ|IU \\ F \to LER|a|b|IA|IB|IZ|IU \\ T \to TMF|LER|a|b|IA|IB|IZ|IU \\ E \to EPT|TMF|LER|a|b|IA|IB|IZ|IU \\ A \to a \\ B \to b \\ Z \to 0 \\ U \to 1 \\ P \to + \\ M \to * \\ L \to (\\ R \to) \end{array}\right\} \left\{\begin{array}{l} F \to LC_3|a|b|IA|IB|IZ|IU \\ T \to TC_2|LC_3|a|b|IA|IB|IZ|IU \\ E \to EC_1|TC_2|LC_3|a|b|IA|IB|IZ|IU \\ C_1 \to PT \\ C_2 \to MF \\ C_3 \to ER \\ I, A, B, Z, U, P, M, L, R \text{ jako předchozí} \end{array}\right\}$$

Příprava na (pumping) lemma o vkládání

Lemma (Velikost derivačního stromu gramatiky v CNF)

Mějme derivační strom podle gramatiky $G = (V, T, P, S)$ v Chomského normálním tvaru, který dává slovo w. Je–li délka nejdelší cesty n, pak $|w| \leq 2^{n-1}$.

Proof.

Indukcí podle n,

Základ $|a| = 1 = 2^0$

Indukce $2^{n-2} + 2^{n-2} = 2^{n-1}$.

□

Lemma (Důsledek)

Mějme derivační strom podle gramatiky $G = (V, T, P, S)$ v Chomského normální formě, který dává slovo w, $|w| > p = 2^{n-1}$. Pak ve stromě existuje cesta delší než n.

Lemma o vkládání (pumping) pro bezkontextové jazyky

Věta 1.18: !!Lemma o vkládání (pumping) pro bezkontextové jazyky

Mějme bezkontextový jazyk L. Pak existuje přirozené číslo $n \in \mathbb{N}$ takové, že každé $z \in L, |z| > n$ lze rozložit na $z = uvwxy$ kde:

- $|vwx| \leq n$
- $vx \neq \epsilon$
- $\forall i \geq 0,\ uv^i wx^i y \in L$.

Idea důkazu:

- vezmeme derivační strom pro z
- najdeme nejdelší cestu
- na ní dva stejné neterminály
- tyto neterminály určí dva podstromy
- podstromy definují rozklad slova
- nyní můžeme větší podstrom posunout ($i > 1$)
- nebo nahradit menším podstromem ($i = 0$).

Důkaz: $|z| > n : z = uvwxy, |vwx| \leq n, vx \neq \epsilon, \forall i \geq 0 uv^i wx^i y \in L$

- vezmeme gramatiku v Chomského NF (pro $L = \{\epsilon\}$ a $\emptyset$ zvol $n = 1$).
- Nechť $|V| = k$. Položíme $n = 2^k$.
- Pro $z \in L, |z| > 2^k$, má v derivačním stromu z cestu délky $> k$
- vezmeme cestu maximální délky; terminál kam vede označíme t
- Aspoň dva z posledních $(k+1)$ neterminálů na cestě do t jsou stejné
- vezmeme dvojici A^1, A^2 nejblíže k t (určuje podstromy T^1, T^2)
- cesta z A^1 do t je nejdelší v podstromu T^1 a má délku maximálně $(k+1)$

  tedy slovo dané stromem T^1 není delší než 2^k (tedy $|vwx| \leq n$)
- z A^1 vedou dvě cesty (ChNF), jedna do T^2 druhá do zbytku vx

  ChNF je nevypouštějící, tedy $vx \neq \epsilon$
- derivace slova $(A^1 \Rightarrow^* vA^2x, A^2 \Rightarrow^* w)$

  $S \Rightarrow^* uA^1y \Rightarrow^* uvA^2xy \Rightarrow^* uvwxy$
- posuneme–li A^2 do A^1 $(i = 0)$

  $S \Rightarrow^* uA^2y \Rightarrow^* uwy$
- posuneme–li A^1 do A^2 $(i = 2, 3, \ldots)$

  $S \Rightarrow^* uA^1y \Rightarrow^* uvA^1xy \Rightarrow^* uvvA^2xxy \Rightarrow^* uvvwxxy.$

□

Použití lemma o vkládání

Příklad 1.47 (ne–bezkontextový jazyk)

Následující jazyk není bezkontextový

- $\{0^i 1^i 2^i | i \geq 1\}$

- důkaz sporem: předpokládejme bezkontextovost
- z lemmatu o vkládání máme n
- zvolme $z = |0^n 1^n 2^n| > n$

žádné dělení nesplňuje PL neboť

- pumpovací slovo $|vwx| \leq n$
- tj. vždy lze pumpovat maximálně dva různé symboly
- $vx \neq \epsilon$, iterací se slovo změní
- poruší se rovnost počtu symbolů – SPOR

Příklad 1.48 (ne–bezkontextový jazyk)

Následující jazyk není bezkontextový

- $\{0^i 1^j 2^k | 0 \leq i \leq j \leq k\}$

- důkaz sporem: předpokládejme bezkontextovost
- z lemmatu o vkládání máme n
- zvolme $z = |0^n 1^n 2^n| > n$

žádné dělení nesplňuje PL neboť

- pumpovací slovo $|vwx| \leq n$
- tj. vždy lze pumpovat maximálně dva různé symboly
  - pokud 0 (nebo 1), pumpujeme nahoru – SPOR $i \leq j$ (nebo $j \leq k$)
  - pokud 2 (nebo 1), pumpujeme

Použití lemma o vkládání

Příklad 1.49 (ne–bezkontextový jazyk)

Následující jazyk není bezkontextový

- $\{0^i 1^j 2^i 3^j | i, j \geq 1\}$

- důkaz sporem: předpokládejme bezkontextovost
- z lemmatu o vkládání máme n
- zvolme $z = |0^n 1^n 2^n 3^n| > n$
- pumpovací slovo $|vwx| \leq n$
- tj. vždy lze pumpovat maximálně dva různé sousední symboly
- poruší se rovnost počtu symbolů 0 a 2 nebo 1 a 3 – SPOR.

Příklad 1.50 (ne–bezkontextový jazyk)

Následující jazyk není bezkontextový

- $\{ww | w \in \{0,1\}^*\}$

- důkaz sporem: předpokládejme bezkontextovost
- z lemmatu o vkládání máme n
- zvolme $z = |0^n 1^n 0^n 1^n| > n$
- pumpovací slovo $|vwx| \leq n$
- tj. vždy lze pumpovat maximálně dva různé sousední symboly
- poruší se buď rovnost nul či jedniček
  - rozeberete 4 případy, vx obsahuje znak z prvních nul, prvních jedniček, druhých nul, druhých jedniček.

Kdy lemma o vkládání nezabere

- Lemma o vkládání je pouze implikace!

Příklad 1.51 (pumpovatelný, ne–bezkontextový jazyk)

$L = \{a^i b^j c^k d^l | i = 0 \vee j = k = l\}$ není bezkontextový jazyk, přesto lze pumpovat.

$i = 0 : b^j c^k d^l$ lze pumpovat v libovolném písmenu
$i > 0 : a^i b^n c^n d^n$ lze pumpovat v části obsahující a

Co s tím?

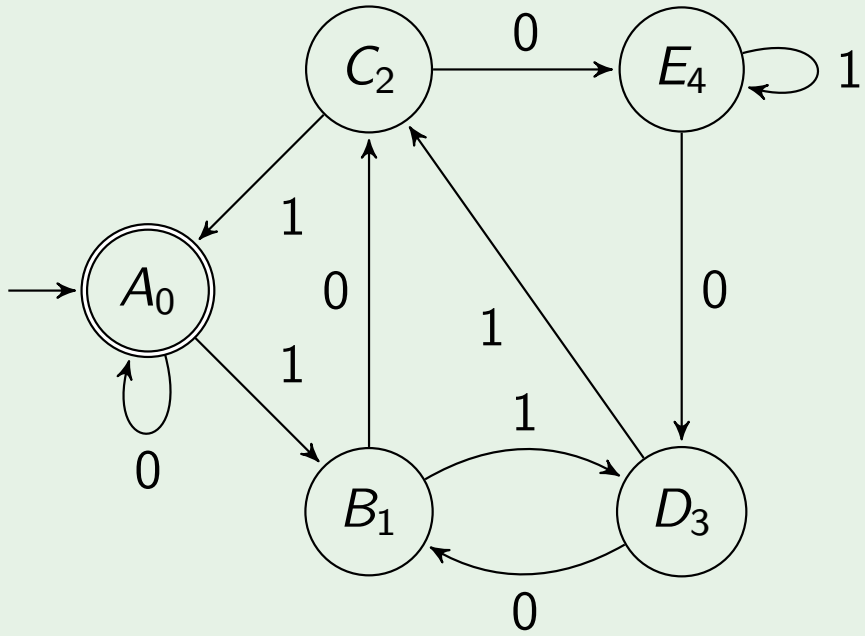
- zobecnění pumping lemmatu (Ogdenovo lemma)
  - pumpování vyznačených symbolů
- uzávěrové vlastnosti.

Příklad převodu FA na gramatiku

Příklad 1.52 (G, FA binární zápis čísla dělitelného 5)

$L = \{w | w \in \{0,1\}^* \,\&\, w$ je binární zápis čísla dělitelného 5$\}$

$$
\begin{aligned}
A &\to 1B|0A|\epsilon \\
B &\to 0C|1D \\
C &\to 0E|1A \\
D &\to 0B|1C \\
E &\to 0D|1E
\end{aligned}
$$



Příklady derivací

$$
\begin{aligned}
&A \Rightarrow 0A \Rightarrow 0 && (0)\\
&A \Rightarrow 1B \Rightarrow 10C \Rightarrow 101A \Rightarrow 101 && (5)\\
&A \Rightarrow 1B \Rightarrow 10C \Rightarrow 101A \Rightarrow 1010A \Rightarrow 1010 && (10)\\
&A \Rightarrow 1B \Rightarrow 11D \Rightarrow 111C \Rightarrow 1111A \Rightarrow 1111 && (15)
\end{aligned}
$$

Převod konečného automatu na gramatiku typu 3

Věta 1.19: $L \in RE \Rightarrow L \in \mathcal{L}_3$

Pro každý jazyk rozpoznávaný konečným automatem existuje gramatika typu 3, která ho generuje.

Důkaz: Převod konečného automatu na gramatiku typu 3

- $L = L(A)$ pro deterministický konečný automat $A = (Q, \Sigma, \delta, q_0, F)$.
- definujme gramatiku $G = (Q, \Sigma, P, q_0)$, kde pravidla P mají tvar
  $p \to aq$, když $\delta(p, a) = q$
  $p \to \epsilon$, když $p \in F$
- je $L(A) = L(G)$?
  - $\epsilon \in L(A) \Leftrightarrow q_0 \in F \Leftrightarrow (q_0 \to \epsilon) \in P \Leftrightarrow \epsilon \in L(G)$
  - $a_1 \dots a_n \in L(A) \Leftrightarrow \exists q_0, \dots, q_n \in Q$ tž. $\delta(q_i, a_{i+1}) = q_{i+1}, q_n \in F$
    $\Leftrightarrow (q_0 \Rightarrow a_1 q_1 \Rightarrow \dots a_1 \dots a_n q_n \Rightarrow a_1 \dots a_n)$ je derivace pro $a_1 \dots a_n$
    $\Leftrightarrow a_1 \dots a_n \in L(G)$

Příprava převodu gramatiky typu 3 na FA

- Opačný směr
  - pravidla $A \to aB$ kódujeme do přechodové funkce
  - pravidla $A \to \epsilon$ určují koncové stavy
  - pravidla $A \to a_1 \dots a_n B, A \to a_1 \dots a_n$ s více neterminály rozepíšeme
    - zavedeme nové neterminály $Y_2, \dots, Y_n, Z_1, \dots, Z_n$
    - vytvoříme pravidla $A \to a_1 Y_2, Y_2 \to a_2 Y_3, \dots, Y_n \to a_n B$
    - resp. $Z \to a_1 Z_1, Z_1 \to a_2 Z_2, \dots, Z_{n-1} \to a_n Z_n, Z_n \to \epsilon$
  - pravidla $A \to B$ odpovídají ϵ přechodům
    - zbavíme se jich tranzitivním uzávěrem
    - nebo musíme tranzitivně uzavřít $S \to B$ pro hledání $S \to \epsilon$.

Lemma

Ke každé gramatice typu 3 existuje gramatika typu 3, která generuje stejný jazyk a obsahuje pouze pravidla ve tvaru: $A \to aB, A \to \epsilon$, $A, B \in V, a \in T$.

Standardizace gramatiky typu 3

Lemma

Ke každé gramatice typu 3 existuje gramatika typu 3, která generuje stejný jazyk a obsahuje pouze pravidla ve tvaru: $A \to aB, A \to \epsilon$, $A, B \in V, a \in T$.

Proof.

Pro gramatiku $G = (V, T, S, P)$ definujeme $G^| = (V^|, T, S, P^|)$, kde pro každé pravidlo zavedeme dostatečný počet nových neterminálů $Y_2, \ldots, Y_n, Z_1, \ldots, Z_n$ a definujeme

P	$P^|$
$A \to aB$ $A \to \epsilon$ $A \to a_1 \ldots a_n B$ $Z \to a_1 \ldots a_n$	$A \to aB$ $A \to \epsilon$ $A \to a_1 Y_2, Y_2 \to a_2 Y_3, \ldots Y_n \to a_n B$ $Z \to a_1 Z_1, Z_1 \to a_2 Z_2, \ldots, Z_{n-1} \to a_n Z_n, Z_n \to \epsilon$
odstraníme i pravidla: $A \to B$	tranzitivní uzávěr $U(A) = \{B \mid B \in V \& A \Rightarrow^* B\}$ $A \to \gamma$ pro všechna $Z \in U(A)$ a $(Z \to \gamma) \in P^|$

□

Pouze pravidla $A \to aB, A \to \epsilon$

Příklad 1.53

P	$P^{\mid}$
$B \to a_1$	$B \to a_1 H_1, H_1 \to \epsilon$
	$U(A) = \{A, B\}$, proto
$A \to B$	$A \to a_1 H_2, H_2 \to \epsilon$
$A \to a_2$	$A \to a_2 H_3, H_3 \to \epsilon$

Převod gramatiky typu 3 na konečný automat

Věta 1.20: ϵ–NFA pro gramatiku typu 3 rozpoznávající stejný jazyk

Pro každý jazyk L generovaný gramatikou typu 3 existuje ϵ–NFA rozpoznávající L.

Důkaz: Převod gramatiky typu 3 na konečný automat

- Vezmeme $G = (V, T, P, S)$ obsahující jen pravidla tvaru $A \to aB$, $A \to \epsilon$, $A, B \in V, a \in T$ generující L (předchozí lemma)
- definujeme nedeterministický ϵ–NFA $A = (V, T, \delta, S, F)$, kde:

  $F = \{A | (A \to \epsilon) \in P\}$

  $\delta(A, a) = \{B | (A \to aB) \in P\}$
- $L(G) = L(A)$
  - $\epsilon \in L(G) \Leftrightarrow (S \to \epsilon) \in P \Leftrightarrow S \in F \Leftrightarrow \epsilon \in L(A)$
  - $a_1 \dots a_n \in L(G) \Leftrightarrow$ existuje derivace $(S \Rightarrow a_1 H_1 \Rightarrow \dots \Rightarrow a_1 \dots a_n H_n \Rightarrow a_1 \dots a_n)$

  $\Leftrightarrow$ $\exists H_0, \dots, H_n \in V$ tak že $H_0 = S, H_n \in F$

  $H_{i+1} \in \delta(H_i, a_k)$ pro krok $a_1 \dots a_{k-1} H_i \Rightarrow a_1 \dots a_{k-1} a_k H_{i+1}$

  $\Leftrightarrow$ $a_1 \dots a_n \in L(A)$

Levé (a pravé) lineání gramatiky

Definice 1.40: Levé (a pravé) lineání gramatiky

Gramatiky typu 3 nazýváme také **pravé lineární** (neterminál je vždy vpravo). Gramatika G je **levá lineání**, jestliže má pouze pravidla tvaru $A \to Bw, A \to w, A, B \in V, w \in T^*$.

Lemma

Jazyky generované levou lineání gramatikou jsou právě regulární jazyky.

Důkaz:

- $\Rightarrow$ 'otočením' pravidel levé lineární gramatiky dostaneme pravou lineární $A \to Bw, A \to w$ převedeme na $A \to w^R B, A \to w^R$
  - získaná gramatika generuje jazyk L^R, najdeme automat
  - víme, že regulární jazyky jsou uzavřené na reverzi, L^R je regulární, tudíž i $L = (L^R)^R$ je regulární
- $\Leftarrow$ takto lze získat všechny regulární jazyky

Lineární gramatiky (a jazyky)

- Levá a pravá lineární pravidla dohromady jsou už silnější.

Definice 1.41: lineární gramatika, jazyk

Gramatika je lineární, jestliže má pouze pravidla tvaru $A \to uBw, A \to w, A, B \in V, u, w \in T^*$ (na pravé straně vždy maximálně jeden neterminál).
Lineární jazyky jsou právě jazyky generované lineáními gramatikami.

- Zřejmě platí: regulární jazyky $\subseteq$ lineární jazyky.
- Jde o vlastní podmnožinu $\subsetneq$.

Příklad 1.54 (lineární, neregulární jazyk)

Jazyk $L = \{0^i 1^i | i \geq 1\}$ není regulární jazyk, ale je lineární, generovaný gramatikou s pravidly $S \to 0S1|01$.

Pozorování:

- lineární pravidla lze rozložit na levě a pravě lineání pravidla: $S \to 0A, A \to S1$.

Ekvivalence gramatik

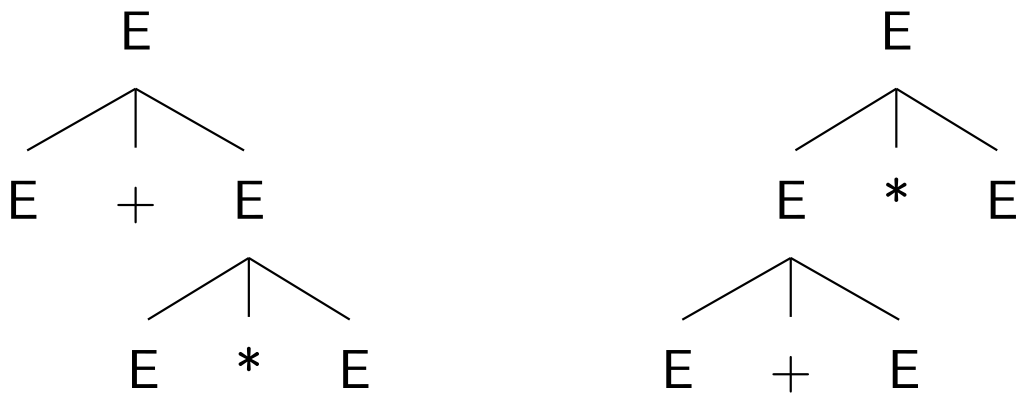
Definice 1.42: ekvivalence gramatik

Gramatiky G_1, G_2 jsou **ekvivalentní**, jestliže $L(G_1) = L(G_2)$, tj. generují stejný jazyk.

Víceznačnost gramatik

Dvě derivace téhož výrazu:

$E \Rightarrow E + E \Rightarrow E + E * E$ $\quad$ $E \Rightarrow E * E \Rightarrow E + E * E$



- Rozdíl je důležitý, vlevo $1 + (2 * 3) = 7$, vpravo $(1 + 2) * 3 = 9$.
- Tato gramatika může být modifikovaná na jednoznačnou.

Příklad 1.55

Různé derivace mohou reprezentovat stejný derivační strom, pak není problém.

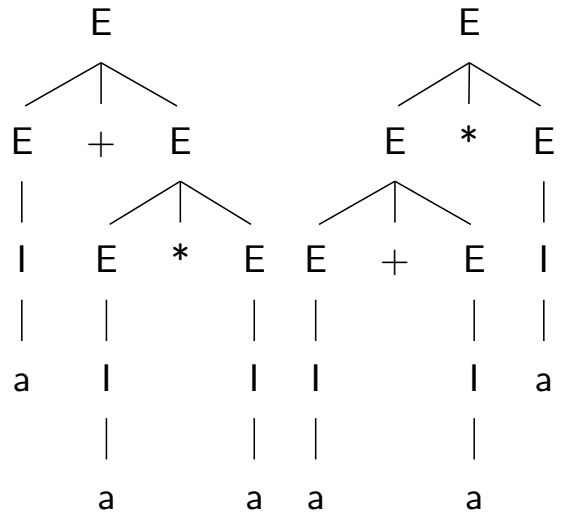
1. $E \Rightarrow E + E \Rightarrow I + E \Rightarrow a + E \Rightarrow a + I \Rightarrow a + b$
2. $E \Rightarrow E + E \Rightarrow E + I \Rightarrow I + I \Rightarrow I + b \Rightarrow a + b$.

Definice 1.43: Jednoznačnost a víceznačnost CFG

- Bezkontextová gramatika $G = (V, T, P, S)$ je **víceznačná** pokud existuje aspoň jeden řetězec $w \in T^*$ pro který můžeme najít dva různé derivační stromy, oba s kořenem S dávající slovo w.
- V opačném případě nazýváme gramatiku **jednoznačnou**.
- Bezkontextový jazyk L je **jednoznačný**, jestliže existuje jednoznačná CFG G tak, že $L = L(G)$.
- **Bezkontextový jazyk L je (podstatně) nejednoznačný**, jestliže každá CFG G taková, že $L = L(G)$, je nejednoznačná. Takovému jazyku říkáme i **víceznačný**.

Příklad 1.56 (nejednoznačnost CFG)

Dva derivační stromy dávající $a + a * a$ ukazující víceznačnost gramatiky.



Příklad víceznačného jazyka

Příklad 1.57 (Víceznačný jazyk)

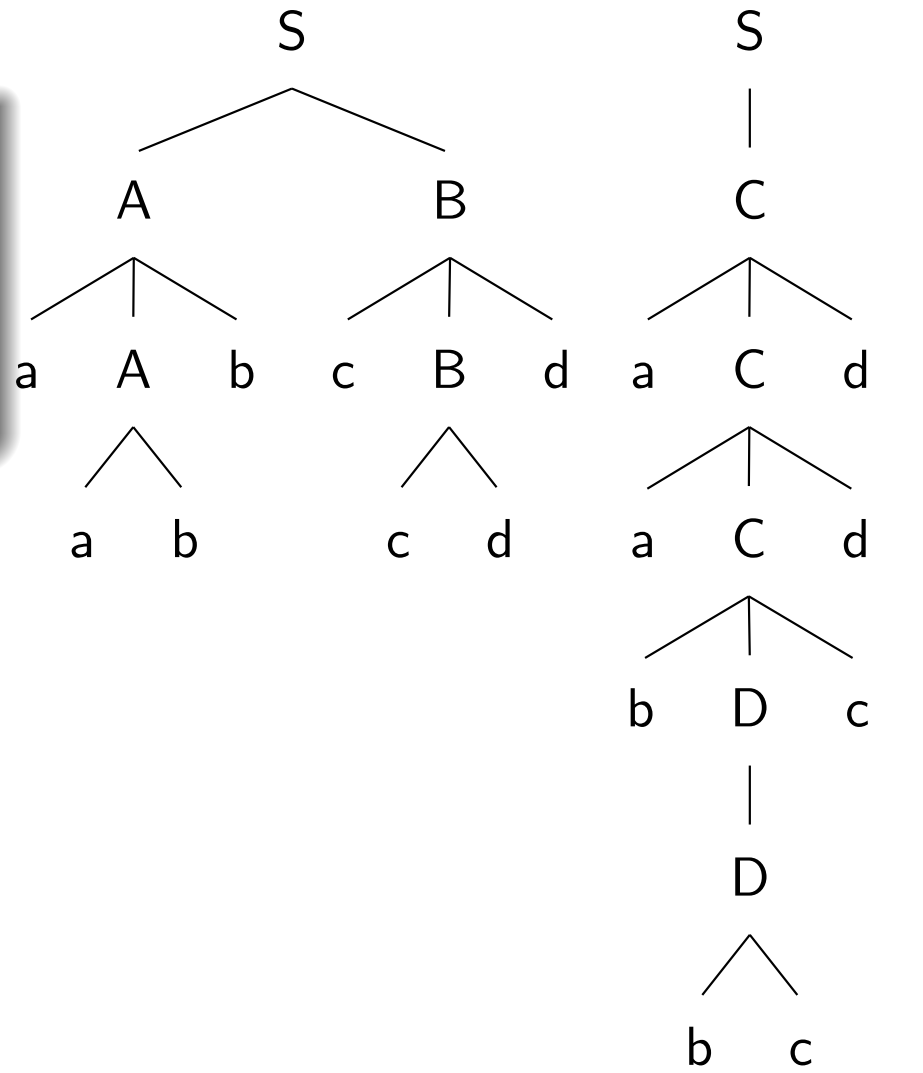
Příklad víceznačného jazyka:

$$L = \{a^n b^n c^m d^m | n \geq 1, m \geq 1\} \cup \cup \{a^n b^m c^m d^n | n \geq 1, m \geq 1\}.$$

1. $S \to AB|C$
2. $A \to aAb|ab$
3. $B \to cBd|cd$
4. $C \to aCd|aDd$
5. $D \to bDc|bc$.

Jakákoli gramatika pro daný jazyk bude generovat pro některá slova typu $a^n b^n c^n d^n$ dva různé derivační stromy.

Dva derivační stromy pro *aabbccdd*.



Odstanění víceznačnosti gramatiky

- Neexistuje algoritmus, který nám řekne, zda je daná gramatika víceznačná.
- Existují bezkontextové jazyky, pro které neexistuje jednoznačná bezkontextová gramatika, pouze víceznačné CFG.
- Existují určitá doporučení pro odstranění víceznačnosti.

Víceznačnost má různé příčiny:

- Není respektovaná priorita operátorů.
- Posloupnost identických operátorů lze shlukovat zleva i zprava.
- $S \rightarrow$ if then S else $S|$ if then $S|\epsilon$

  slovo 'if then if then else' má dva významy

  'if then (if then else)' nebo 'if then (if then) else'

  Řešení:
  - syntaktická chyba (Algol 60)
  - else patří k bližšímu if (preference pořadí pavidel)
  - závorky begin-end,odsazení v Python (asi nejčistší řešení).

Vynucení priority

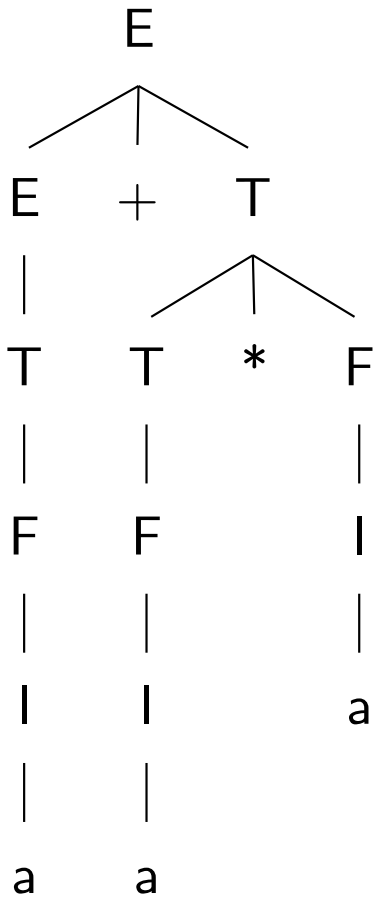
Řešením je zavést více různých proměnných, každou pro jednu úroveň 'priority'.
Konkrétně:

- **Faktor** je výraz který nesmí rozdělit žádný operátor.
  - identifikátory
  - výraz v závorkách
- **Term** je výraz, který nemůže rozdělit operátor $+$.
- **Výraz** může být rozdělen $*$ i $+$.

Jednoznačná gramatika pro výrazy:

1. $I \to a|b|Ia|Ib|I0|I1$
2. $F \to I|(E)$
3. $T \to F|T * F$
4. $E \to T|E + T$.

Jediný derivační strom pro $a + a * a$.



Jednoznačnost a kompilátory

Kompilace výrazu (zásobník na mezivýsledky + dva registry):

(1) $E \to E + T$ ... pop r1; pop r2; add r1,r2; push r2
(2) $E \to T$
(3) $T \to T * F$ ... pop r1; pop r2; mul r1,r2; push r2
(4) $T \to F$
(5) $F \to (E)$
(6) $F \to a$ ... push a

- 'a+a*a' získáme postupnou aplikací pravidel 1,2,4,6,3,4,6,6
- posloupnost obrátíme a vybereme pouze pravidla generující kód

  6,6,3,6,1
- nyní nahradíme pravidla příslušným kódem

  push a; push a; pop r1; pop r2; mul r1,r2; push r2; push a; pop r1; pop r2; add r1,r2; push r2

Shrnutí

- Gramatiky
  - obecné
  - kontextové
  - bezkontextové
  - regulární, pravé lineární
- jazyk gramatiky, derivace, derivace dává slovo, derivační strom (pro bezkontextové gramatiky), ekvivalentní gramatiky
- ne každá lineární gramatika má ekvivalentní pravou lineární
- bezkontextové gramatiky
- jednoznačné a (podstatně) víceznačné gramatiky.