AUTOMATY A GRAMATIKY

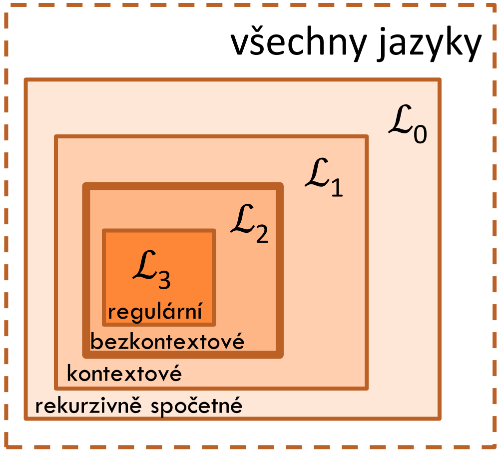
8

Normální formy bezkontextových gramatik
Pumping lemma
Zásobníkové automaty

Pavel Surynek

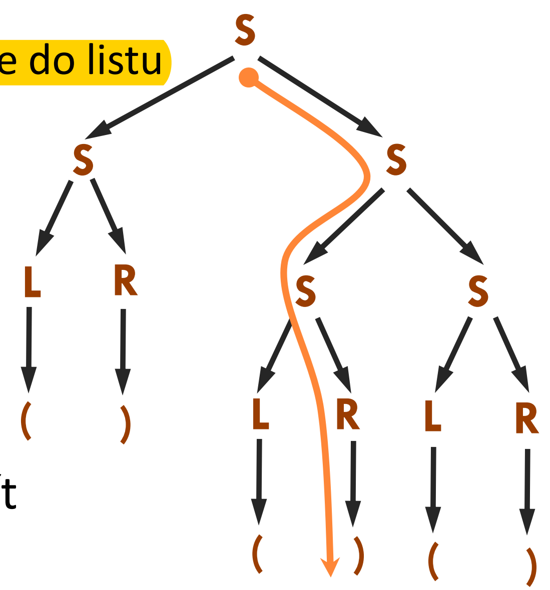
Univerzita Karlova v Praze

Matematicko-fyzikální fakulta
Katedra teoretické informatiky a matematické logiky



Chomského normální forma

- **větvící faktor** u derivačního stromu
  - **počet následníků** vnitřních vrcholů
    - pokud bychom znali **horní** odhad na větvící faktor, můžeme **zdola** odhadnout výšku derivačního stromu pro dané slovo
      - výška odpovídá délce nejdelší orientované cesty z kořene do listu
  - bezkontextová gramatika G = (V_N, V_T, S, P) je v **Chomského normálním** tvaru, jestliže pravidla mají tvar:
    - $X \rightarrow AB$ pro $X, A, B \in V_N$, nebo $X \rightarrow x$ pro $X \in V_N$ a $x \in V_T$
  - větvící faktor je nejvýše 2
    - derivační strom, který přináší slovo délky > 2^{h-1} musí mít výšku aspoň h+2 (k listové hladině se nevětví)
  - **dost vysoký** derivační strom má opakující se neterminály na cestě z kořene do listu



Převod na Chomského formu

- ke každé bezkontextové gramatice $G = (V_N, V_T, S, P)$ existuje bezkontextová gramatika $G' = (V_N', V_T, S, P')$ **v Chomského normální formě**, že $L(G') = L(G)-\{\lambda\}$
  - **předzpracování**
    - eliminace pravidel $X \rightarrow Y$ pro $X, Y \in V_N$
      - stejně jako u regulárních gramatik
    - eliminace pravidel $X \rightarrow \lambda$ pro $X \in V_N$ (zde může dojít ke ztrátě λ)
      - stejně jako u převodu bezkontextové gramatiky na kontextovou
  - zbývají pravidla tvaru $X \rightarrow x$ pro $X \in V_N$ a $x \in V_T$ a pravidla $\mathbf{X \rightarrow Y_1Y_2...Y_n}$, kde $Y_i \in (V_N \cup V_T)$ pro $i = 1,2,...,n$ s $n \geq 2$
    - pro každý terminál $x \in V_T$ zavedeme nový neterminál X_x a pravidlo $\mathbf{X_x \rightarrow x}$
    - $X \rightarrow Y_1Y_2...Y_n$ nahradíme pravidlem $X \rightarrow Z_1Z_2...Z_n$, kde
      - $Z_i = \begin{cases} Y_i & \text{když } Y_i \in V_N \\ X_{Y_i} & \text{když } Y_i \in V_T \end{cases}$ pro $i = 1,2,...,n$
    - jestliže $n>2$ vzniklé pravidlo $X \rightarrow Z_1Z_2...Z_n$ dále nahradíme pravidly
      - $X \rightarrow Z_1B_1$, $B_1 \rightarrow Z_2B_2$, ..., $B_{n-2} \rightarrow Z_{n-1}Z_n$, kde B_j pro $j=1,2,...,n-2$ jsou vždy nové neterminály

Př.: $A \rightarrow BcDe$

nové neterminály
X_c, X_e
a pravidla
$X_c \rightarrow c$, $X_e \rightarrow e$

(i) $A \rightarrow BX_cDX_e$
(ii) $A \rightarrow BB_1$
$B_1 \rightarrow X_cB_2$
$B_2 \rightarrow DX_e$

Poznámky k převodu

- použití původního pravidla $X \rightarrow Y_1Y_2...Y_n$ je **atomické**
  - tomu musí odpovídat **atomicita** (nedělitelnost) použití sekvence náhradních pravidel
    - sekvence nových pravidel se použije buď <u>celá</u> nebo <u>vůbec</u>
      - při použití nového pravidla $X \rightarrow Z_1B_1$ je nutno postupně použít všechna nová pravidla $B_j \rightarrow Z_{j+1}B_{j+1}$
        - jinak nelze B_j postupně přepsat neterminály
        - po použití pravidla je jednoznačně určené, které následující má být použito

Př.: bezkontextová gramatika

$G = (V_N, V_T, A, P)$, kde

$V_N = \{ A, B, C, D, E \}$

$V_T = \{ 0, 1 \}$

$P = \{ A \rightarrow B \mid C \quad B \rightarrow 0B1 \mid 01$

$C \rightarrow D \mid E \quad D \rightarrow 1D0 \mid 1$

$E \rightarrow 0E \mid 0 \}$

Př.: ekvivalentní bezkontextová v Chomského tvaru

$G' = (V_N', V_T, A, P')$, kde

$V_N = \{ A, A_1, A_2, B, B_1, D, D_1, E, X_0, X_1\}$

$V_T = \{ 0, 1 \}$

$P = \{ A \rightarrow X_0A_1 \mid X_0X_1 \mid X_1A_2 \mid 1 \mid X_0E \mid 0 \quad A_1 \rightarrow BX_1$

$A_2 \rightarrow DX_0 \quad B \rightarrow X_0B_1 \mid X_0X_1 \quad B_1 \rightarrow BX_1$

$D \rightarrow X_1D_1 \mid 1 \quad D_1 \rightarrow DX_0 \quad E \rightarrow X_0E \mid 0 \quad X_0 \rightarrow 0 \quad X_1 \rightarrow 1\}$

Greibachové normální forma

- **Greibachové normální** forma bezkontextové gramatiky $G = (V_N, V_T, S, P)$
  - pravidla jsou tvaru $X \rightarrow xY_1Y_2\ldots Y_n$ pro $X, Y_1, Y_2,\ldots,Y_n \in V_N$ a $x \in V_T$, kde $n \in \mathbb{N}_0$
  - důležité v implementaci analyzátoru
    - předstupeň k LL(1) analyzátoru
      - nutno ještě přidat jednoznačnost

Př.: $G = (V_N, V_T, B, P)$, kde
$V_N = \{ R, B \}$
$V_T = \{ (,) \}$
P= { **B → (RB   R →) | (RR** }

- ke každé bezkontextové gramatice $G = (V_N, V_T, S, P)$ existuje bezkontextová gramatika $G'' = (V_N'', V_T, S, P'')$ v **Greibachové normální formě**, že $L(G') = L(G)-\{\lambda\}$
  - provedeme eliminaci pravidel $X \rightarrow \lambda$
    - zjednoduší se rozbor případů v dalších krocích
  - postupně
    - odstraníme přímou **levou rekurzi**, tj. pravidla tvaru $X \rightarrow Xu$ pro $X \in V_N$ a $u \in (V_N \cup V_T)^+$ nahradíme jinými
    - nahradíme **pravidla tvaru $X \rightarrow uYv$** pro $X,Y \in V_N$ a $u,v \in (V_N \cup V_T)^*$
  - u výsledné gramatiky dokončíme úpravu na Greibachové formu

Pavel Surynek, 2015

Převod na **Greibachové** formu (1)

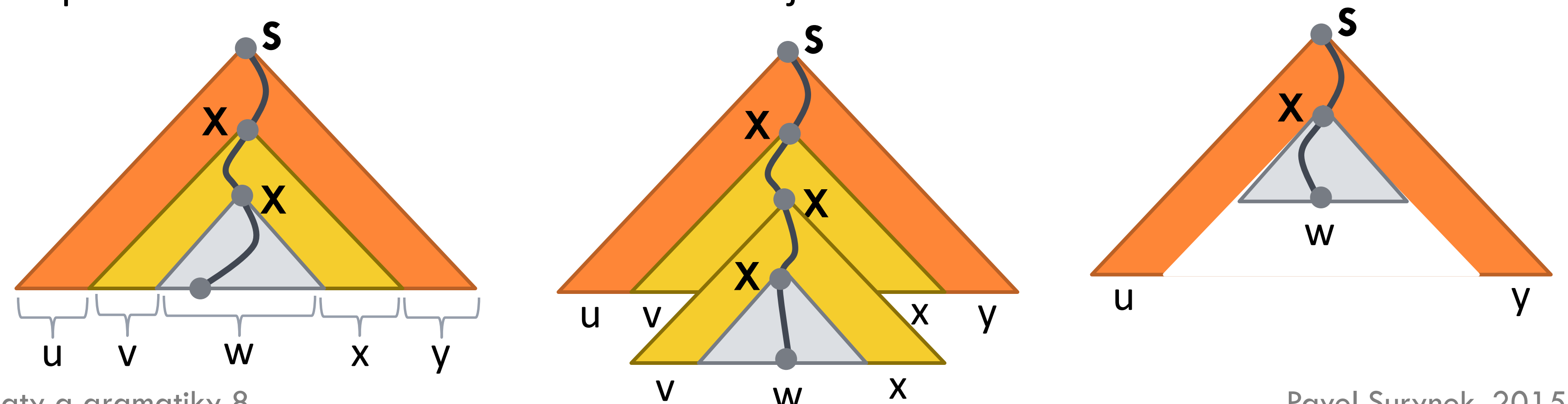
- **náhrada levě rekurzivních** pravidel $X \rightarrow Xu$
  - nechť $X \rightarrow Xu_1$, $X \rightarrow Xu_2$, ... $X \rightarrow Xu_n$ s $u_i \in (V_N \cup V_T)^+$ pro $i \in \{1,2,...,n\}$ jsou všechna levě rekurzivní pravidla pro $X \in V_N$, kde
  - a $X \rightarrow v_1$, $X \rightarrow v_2$, ... $X \rightarrow v_k$ s $v_j \in (V_N \cup V_T)^+$ pro $j \in \{1,2,...,k\}$ jsou všechna zbývající pravidla pro X
    - nahrazením všech těchto pravidel pravidly $X \rightarrow v_j Z | v_j$ a $Z \rightarrow u_i Z | u_i$ pro všechna $i=1,2,...,n$ a $j=1,2,...,k$ dostaneme ekvivalentní gramatiku
- **náhrada** pravidel **tvaru $X \rightarrow uYv$**
  - nechť $Y \rightarrow w_1$, $Y \rightarrow w_2$, ..., $Y \rightarrow w_m$ s $w_i \in (V_N \cup V_T)^*$ pro $i \in \{1,2,...,m\}$ jsou všechna pravidla pro $Y \in V_N$
    - nahrazením pravidla $X \rightarrow uYv$ pravidly $X \rightarrow uw_1v$, $X \rightarrow uw_2v$, ..., $X \rightarrow uw_mv$ dostaneme ekvivalentní gramatiku
- chceme vytvořit gramatiku s pravidly tvaru
  - $X \rightarrow xw$, kde $X \in V_N$, $x \in V_T$, $w \in (V_N \cup V_T)^*$ nebo
  - $X \rightarrow Yw$, kde $X,Y \in V_N$, $w \in (V_N \cup V_T)^*$ a platí, že
    - X předchází Y v nějakém očíslování neterminálů, kde S je první

Převod na **Greibachové** formu (2)

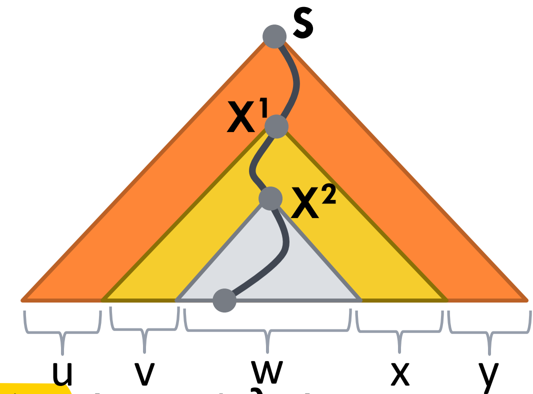
- očíslujeme neterminály
  - $V_N = \{S = X_1, X_2, \ldots, X_n\}$
- postupně zpracujeme pravidla pro neterminály $\mathbf{X_1, X_2, \ldots, X_n}$
  - když zpracovávané pravidlo je tvaru $X_i \to X_i u$ pro $u \in (V_N \cup V_T)^*$
    - provedeme nahrazení levé rekurze
  - když zpracovávané pravidlo je tvaru $X_i \to X_j u$ pro $u \in (V_N \cup V_T)^*$, kde $j < i$
    - provedeme s $X_i \to X_j u$ náhradu pravidla tvaru $X \to uYv$
      - pravidla pro X_j již byla zpracována, na začátku pravé strany se nemůže objevit neterminál s vyšším pořadovým číslem
- finální úprava na Greibachové formu
  - postupně zpracujeme pravidla pro neterminály $\mathbf{X_n, X_{n-1}, \ldots, X_1}$
    - pomocné neterminály z náhrad nikdy nestojí na začátku pravé strany
    - pravidla k dalšímu ošetření jsou tedy tvaru $X_i \to X_j u$ s $u \in (V_N \cup V_T)^*$ a $i<j$
      - provedeme s $X_i \to X_j u$ náhradu pravidla tvaru $X \to uYv$
        - pravidla pro X_j již byla zpracována a pravá strana pravidel pro X_n začínala terminály, na začátku pravé strany se tedy objeví terminál (indukce)
  - obdržíme pravidla tvaru $X \to xY_1Y_2\ldots Y_n$ pro $X \in V_N$, $Y_1, Y_2,\ldots,Y_n \in (V_N \cup V_T)$ a $x \in V_T$, kde $n \in \mathbb{N}_0$
    - stačí terminály uvnitř pravých stran nahradit novými neterminály a přidat pravidla

(bezkontextové) **pumping lemma**

- vezmeme-li dostatečně dlouhé slovo z jazyka lze v něm tandemově na dvou místech pumpovat
  - je-li L bezkontextový jazyk nad abecedou X, pak existují $n,m \in \mathbb{N}$, že
    $$(\forall z \in L)[\,|z| \geq n \Rightarrow (\exists u,v,w,x,y \in X^*)(\; z=u.v.w.x.y \;\wedge\; |vwx| \leq m \;\wedge\; vx \neq \lambda \;\wedge\; (\forall i \in \mathbb{N}_0)\, uv^iwx^iy \in L\,)\,]$$
  - důkaz je založen na opakování neterminálu X na cestě z kořene do listu v **derivačním stromu**
    - pokud taková cesta a neterminál existují:



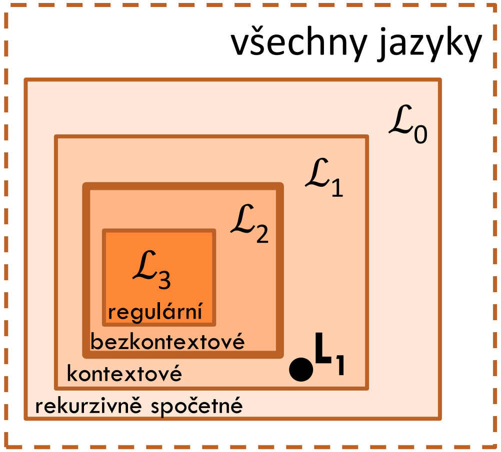
Důkaz pumping lemmatu



- L bezkontextový, pak existují $n,m \in \mathbb{N}$
  - $(\forall z \in L)[\,|z| \geq n \Rightarrow (\exists u,v,w,x,y \in X^*)(z = u.v.w.x.y \wedge |vwx| \leq m \wedge vx \neq \lambda \wedge (\forall i \in \mathbb{N}_0)\, uv^iwx^iy \in L)\,]$
  - nechť $G = (V_N, V_T, S, P)$ je bezkontextová gramatika v Chomského normální formě, že $L(G) = L - \{\lambda\}$
    - větvící faktor je nejvýše 2
    - $h = |V_N|$, $n = 2^{h-1}+1$, $m = 2^h$
      - zvolme slovo $z \in L$, že $|z| \geq 2^{h-1}+1$
      - každý derivační strom vzhledem ke G přinášející <u>z</u> obsahuje cestu z kořene do listu <u>s</u> délky aspoň h+2 (poslední je terminál, tedy aspoň h+1 neterminálů)
      - nějaký neterminál se na této cestě opakuje, nechť je to $X \in V_N$
        - nechť X^1 a X^2 jsou jeho dva různé výskyty nejblíže listu <u>s</u>
      - délka cesty z X^1 do listu <u>s</u> je nejvýše h+2, tedy $|vwx| \leq 2^h = m$
      - X^1 má dva následníky, aspoň z jednoho vede cesta do listů v rámci <u>v</u> nebo <u>x</u>
        - navíc pravidla tvaru $X \rightarrow \lambda$ nemáme, tedy $vx \neq \lambda$
    - $S \Rightarrow_G^* uX^1y \Rightarrow_G^* uvX^2xy \Rightarrow_G^* uvwxy$, pak také
      - $S \Rightarrow_G^* uX^1y \Rightarrow_G^* uvX^1xy \Rightarrow_G^* uvvX^2xxy \Rightarrow_G^* u\underline{vv}w\underline{xx}y$
      - $S \Rightarrow_G^* uX^1y \Rightarrow_G^* uwy$

Ukázky použití pumping lemmatu

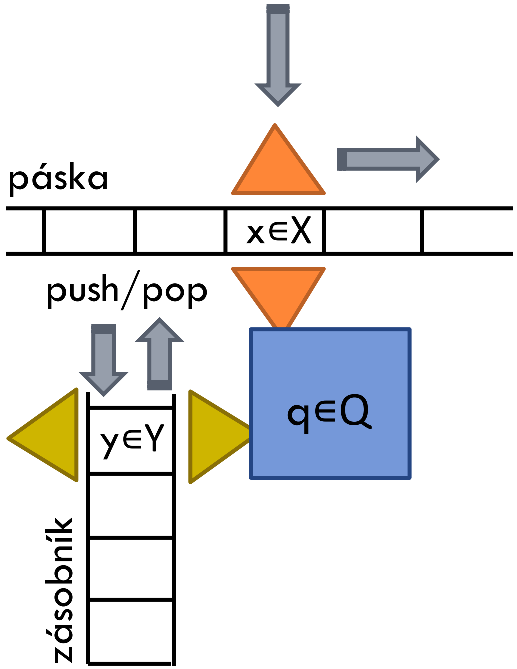
- L bezkontextový, pak existují $n,m \in \mathbb{N}$
  - $(\forall z \in L)[|z| \geq n \Rightarrow (\exists u,v,w,x,y \in X^*)(z=u.v.w.x.y \land |vwx| \leq m \land vx \neq \lambda \land (\forall i \in \mathbb{N}_0) uv^iwx^iy \in L)]$
    - platí-li negace výroku pro každé $n,m \in \mathbb{N}$, tj. pro každé $n,m \in \mathbb{N}$ $(\exists z \in L)[|z| \geq n \land (\forall u,v,w,x,y \in X^*)(z=u.v.w.x.y \land |vwx| \leq m \land vx \neq \lambda \Rightarrow (\exists i \in \mathbb{N}_0) uv^iwx^iy \notin L)]$, jazyk **není bezkontextový**
  - **$L_1 = \{a^ib^ic^i \mid i = 0,1,2,\ldots\}$**
    - $n,m \in \mathbb{N}$ od nepřítele
    - $k = \max(n,m)$, zvolíme $z = a^kb^kc^k$, evidentně $|z| \geq n$
    - zkoumejme všechny možné rozklady $z=u.v.w.x.y$, kde $|vwx| \leq m \land vx \neq \lambda$
      - vwx nemůže obsahovat a a c zároveň
      - při pumpování nedojde k přidání a a c zároveň



  - **$L_2 = \{a^ib^jc^k \mid i,j,k = 0,1,2,\ldots \land i \leq j \leq k\}$**
    - $n,m \in \mathbb{N}$ od nepřítele, opět $k = \max(n,m)$, zvolíme $z = a^kb^kc^k$, evidentně $|z| \geq n$
    - rozklad $z=u.v.w.x.y$ nemůže v části vwx obsahovat a a c zároveň
    - při pumpování buď přidáváme a, nebo ubíráme c (b lze přidávat i ubírat)

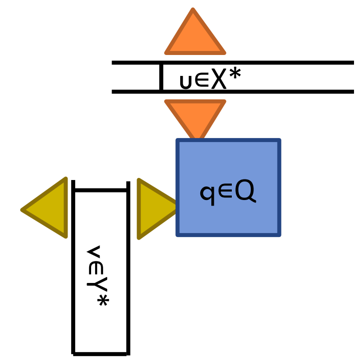
Zásobníkový automat (ZA)

- **zásobníkový** automat (*push-down automaton*)
  - $Z = (Q, X, Y, \delta, q_0, z_0, F)$
    - Q - konečná neprázdná množina stavů
    - X - konečná neprázdná **vstupní** abeceda
    - Y - konečná neprázdná **zásobníková** abeceda
    - $\delta: Q \times (X \cup \{\lambda\}) \times Y \rightarrow 2_{FIN}^{Q \times Y^*}$
      - přechodová funkce trojici stavu, čtenému symbolu pásky (případně **nečtení**) a čtenému symbolu ze zásobníku přiřazuje **konečnou** množinu dvojic stav a zásobníkové slovo
      - $Q \times Y^*$ je nekonečná; nelze tedy použít $2^{Q \times Y^*}$, protože by nešlo konečně reprezentovat
    - $q_0 \in Q$ - počáteční stav
    - $z_0 \in Y$ - počáteční zásobníkový symbol
      - přechodová funkce vždy ze zásobníku čte, tj. provede operaci **pop**
      - na začátku zásobník obsahuje z_0
    - $F \subseteq Q$ – množina přijímajících stavů
  - vyprázdnění zásobníku ukončí výpočet
- deterministický zásobníkový automat
  - $|\delta(q,x,y)| \leq 1$ pro $q \in Q$, $x \in X \cup \{\lambda\}$ a $y \in Y$
  - když $\delta(q,\lambda,y) \neq \emptyset$ pro $q \in Q$ a $y \in Y$, pak $\delta(q,x,y) = \emptyset$ pro všechna $x \in X$



Výpočet zásobníkového automatu

- u ZA hovoříme o **instrukcích**
  - instrukce $(p,x,y) \rightarrow (q,w)$ odpovídá $(q,w)\in\delta(p,x,y)$ pro $p,q\in Q$, $x\in X$, $y\in Y$, $w\in Y^*$
    - je **použitelná**, když
      - stav řídící jednotky je p, na pásce je čten symbol x, na vrcholu zásobníku se nachází symbol y
    - jejím **použitím** dojde ke
      - změně stavu řídící jednotky na q, čtení pokračuje na následující buňce
      - ze zásobníku je odebrán symbol y (**pop(y)**) a na zásobník je vloženo slovo w
        - pro $w = y_1y_2...y_n$ vložení w odpovídá push(y_n), ..., push(y_1)
    - instrukce $(p,\lambda,y) \rightarrow (q,w)$
      - čtený symbol není kontrolován, <u>nedojde</u> k posunu čtení na další buňku



- **konfigurace** ZA
  - trojice (q,u,v), kde $q\in Q$, $u\in X^*$ a $v\in Y^*$
    - u - část slova na pásce, kterou zbývá přečíst (včetně právě čteného symbolu)
    - v - obsah zásobníku (směrem shora)
  - konfigurace $K_1=(p,xu,yv)$ **vede přímo** na konfiguraci $K_2=(q,u,wv)$, kde $q\in Q$, $u\in X^*$, $x\in X$, $v,w\in Y^*$ a $y\in Y$, jestliže $(q,w)\in\delta(p,x,y)$, píšeme $K_1 \vdash_Z K_2$
    - případně $K_1=(p,xu,yv)$ vede na $K_2=(q,xu,wv)$, jestliže $(q,w)\in\delta(p,\lambda,y)$
  - K **vede** na K'; $K \vdash_Z^* K'$ jestliže existují $K_1, K_2, ..., K_n$, kde $n\in\mathbb{N}_0$, že $K \vdash_Z K_1 \vdash_Z ... \vdash_Z K_n \vdash_Z K'$

Jazyky **přijímané** ZA (1)

- $Z = (Q, X, Y, \delta, q_0, z_0, F)$
  - jazyk přijímaný <u>přijímajícím stavem</u> L(Z)
    - $L(Z) = \{ w \mid w \in X^* \wedge (\exists v,f)[v \in Y^* \wedge f \in F \wedge (q_0,w,z_0) \vdash_Z^* (f, \lambda, v)]\}$
  - jazyk přijímaný <u>prázdným zásobníkem</u> N(Z)
    - $N(Z) = \{ w \mid w \in X^* \wedge (\exists q \in Q)(q_0,w,z_0) \vdash_Z^* (q, \lambda, \lambda)\}$

Př.: $L = \{0^i 1^i \mid i=0,1,2,\ldots\}$, ZA Z, že L(Z)=L

$Z = (\{p,q,r\}, \{0, 1\}, \{a, z\}, \delta, p, z, \{r\})$, kde

$\delta(p,0,z) = \{(p, az)\}$
$\delta(p,0,a) = \{(p, aa)\}$
$\delta(p,1,a) = \{(q, \lambda)\}$
$\delta(q,1,a) = \{(q, \lambda)\}$
$\delta(p,\lambda,z) = \{(r, \lambda)\}$
$\delta(q,\lambda,z) = \{(r, \lambda)\}$

Př.: $L = \{0^i 1^i \mid i=0,1,2,\ldots\}$, ZA Z, že N(Z)=L

$Z = (\{p,q\}, \{0, 1\}, \{a, z\}, \delta, p, z, \emptyset)$, kde

$\delta(p,0,z) = \{(p, a)\}$
$\delta(p,0,a) = \{(p, aa)\}$
$\delta(p,1,a) = \{(q, \lambda)\}$
$\delta(q,1,a) = \{(q, \lambda)\}$
$\delta(p, \lambda,z) = \{(p, \lambda)\}$

Jazyky **přijímané** ZA (2)

- $Z = (Q, X, Y, \delta, q_0, z_0, F)$
  - pro zásobníkový automat Z existuje zásobníkový automat Z_1, že $L(Z)=N(Z_1)$
  - $Z_1 = (Q\cup\{q_0',q_f\}, X, Y\cup\{z_0'\}, \delta', q_0', z_0', \emptyset)$
    - na zásobník pomocný symbol z_0'
      - původně vyprázdnění znamenalo konec, nyní by znamenalo přijmutí
    - simulujeme Z
    - v přijímajícím stavu odebereme pomocný z_0'
      - z_0', q_0', q_f jsou nové
      - $\delta'(q_0', \lambda, z_0') = \{(q_0, z_0'z_0)\}$
      - $\delta'(q, x, y) = \delta(q, x, y)$ pro všechna $q\in Q$, $x\in X$, $y\in Y$
      - $\delta'(q, \lambda, y) = \delta(q, \lambda, y)$ pro všechna $q\in Q-F$, $y\in Y$
      - $\delta'(q, \lambda, y) = \delta(q, \lambda, y)\cup\{(q_f,\lambda)\}$ pro $q\in F$, $y\in Y\cup\{z_0'\}$
        - zde rušíme determinismus
      - $\delta'(q_f, \lambda, y)=\{(q_f,\lambda)\}$ pro $y\in Y\cup\{z_0'\}$
  - $w\in L(Z) \Leftrightarrow (q_0,w, z_0)\vdash_Z^* (f, \lambda, v)$ pro $f\in F$ a $v\in Y^*$ $\Leftrightarrow$ $(q_0',w, z_0') \vdash_{Z_1} (q_0,w, z_0z_0') \vdash_{Z_1}^* (f, \lambda, vz_0') \vdash_{Z_1} (q_f, \lambda, vz_0') \vdash_{Z_1}^* (q_f, \lambda, \lambda)$ $\Leftrightarrow w\in N(Z_1)$

Pavel Surynek, 2015

Jazyky **přijímané** ZA (3)

- $Z = (Q, X, Y, \delta, q_0, z_0, F)$
  - pro zásobníkový automat Z existuje zásobníkový automat Z_2, že $N(Z)=L(Z_2)$
    - navíc, když Z je deterministický, je i Z_2 deterministický
  - $Z_2 = (Q \cup \{q_0', q_f\}, X, Y \cup \{z_0'\}, \delta', q_0', z_0', \{q_f\})$
    - na zásobník pomocný symbol z_0'
    - simulujeme Z
    - je-li na zásobníku vidět pomocný z_0', přijímáme, neboť to odpovídá původnímu vyprázdnění
      - z_0', q_0', q_f jsou nové
      - $\delta'(q_0', \lambda, z_0') = \{(q_0, z_0' z_0)\}$
      - $\delta'(q, x, y) = \delta(q, x, y)$ pro všechna $q \in Q$, $x \in X$, $y \in Y$
      - $\delta'(q, \lambda, y) = \delta(q, \lambda, y)$ pro všechna $q \in Q$, $y \in Y$
      - $\delta'(q, \lambda, z_0') = \{(q_f, \lambda)\}$ pro všechna $q \in Q$
  - $w \in N(Z) \Leftrightarrow (q_0, w, z_0) \vdash_Z^* (q, \lambda, \lambda)$ pro $q \in Q \Leftrightarrow$
    $(q_0', w, z_0') \vdash_{Z_2} (q_0, w, z_0 z_0') \vdash_{Z_2}^* (q, \lambda, z_0') \vdash_{Z_2} (q_f, \lambda, \lambda)$
    $\Leftrightarrow w \in L(Z_2)$