

Kapitola VIII.

(Bonus)

LR syntaktická analýza

LR syntaktický analyzátor

- Necht' $G = (N, T, P, S)$ je BKG,
kde $N = \{A_1, A_2, \dots, A_n\}$, $T = \{a_1, a_2, \dots, a_m\}$
- LR-syntaktický analyzátor je rozšířený zásobníkový automat M se stavy $Q = \{q_0, q_1, \dots, q_k\}$, kde q_0 je počáteční stav.
- Činnost M je založena na LR tabulce, která má následující dvě části:
 - 1) **Akční část** (tabulka akcí)
 - 2) **Přechodová část** (tabulka přechodů)

Akční část & přechodová část

Akční část:

α	a_1	...	a_j	...	a_m	\$
q_0						
...						
q_i						
...						
q_k						

$\alpha[q_i, a_j] = 1, 2, 3$ nebo 4
 1) **sq**: **s** = shift, $q \in Q$
 2) **rp**: **r** = redukce, $p \in P$
 3) 😊 : úspěch
 4) **prázdné políčko**: chyba

Přechodová část:

β	A_1	...	A_j	...	A_n
q_0					
...					
q_i					
...					
q_k					

$\beta[q_i, A_j] = 1$ nebo 2
 1) **q**: $q \in Q$
 2) **prázdné políčko**

LR syntaktický analyzátor: Algoritmus

- **Vstup:** LR tabulka pro $G = (N, T, P, S)$; $x \in T^*$
- **Výstup:** Právý rozbor x , pokud $x \in L(G)$, jinak **chyba**

- **Metoda:**

- Vlož $\langle \$, q_0 \rangle$ na zásobník; $stav := q_0$;

- **repeat**

- nechť a = aktuální znak na vstupu

case $\alpha[stav, a]$ **of:**

- sq : push($\langle a, q \rangle$) & přečti další symbol a ze vstupu & $stav := q$;



- rp : zaměň $\langle ?, q \rangle \langle X_1, ? \rangle \langle X_2, ? \rangle \dots \langle X_n, ? \rangle$ za $\langle A, stav \rangle$ na vrcholu zásobníku & zapiš p na výstup, kde $p: A \rightarrow X_1 X_2 \dots X_n \in P$ **and** $stav := \beta[q, A]$;

- ☺: úspěch

- prázdné políčko: chyba

until úspěch **or** chyba

LR syntaktický analyzátor: Příklad 1/2

$G_{expr1} = (N, T, P, E)$, kde $N = \{E, F, T\}$, $T = \{i, +, *, (,)\}$,
 $P = \{$ **1**: $E \rightarrow E+T$, **2**: $E \rightarrow T$, **3**: $T \rightarrow T*F$,
 4: $T \rightarrow F$, **5**: $F \rightarrow (E)$, **6**: $F \rightarrow i$ $\}$

LR-tabulka pro G_{expr1} :

α	i	$+$	$*$	$($	$)$	$\$$
0	s5			s4		
1		s6				☺
2		r2	s7		r2	r2
3		r4	r4		r4	r4
4	s5			s4		
5		r6	r6		r6	r6
6	s5			s4		
7	s5			s4		
8		s6			s11	
9		r1	s7		r1	r1
10		r3	r3		r3	r3
11		r5	r5		r5	r5

Akční část
pro G_{expr1}

Přechodová
část pro G_{expr1}

β	E	T	F
0	1	2	3
1			
2			
3			
4	8	2	3
5			
6		9	3
7			10
8			
9			
10			
11			

LR syntaktický analyzátor: Příklad 2/2

Pravidla: 1: $E \rightarrow E+T$, 2: $E \rightarrow T$, 3: $T \rightarrow T*F$,
 4: $T \rightarrow F$, 5: $F \rightarrow (E)$, 6: $F \rightarrow i$

Vstupní řetězec: $i * i \$$

Zásobník	St.	Vstup	Akce	Pravidlo
$\langle \$, 0 \rangle$	0	$i*i\$$	$\alpha[0, i] = s5$	
$\langle \$, 0 \rangle \langle i, 5 \rangle$	5	$*i\$$	$\alpha[5, *] = r6$ $\beta[0, F] = 3$	6: $F \rightarrow i$
$\langle \$, 0 \rangle \langle F, 3 \rangle$	3	$*i\$$	$\alpha[3, *] = r4$ $\beta[0, T] = 2$	4: $T \rightarrow F$
$\langle \$, 0 \rangle \langle T, 2 \rangle$	2	$*i\$$	$\alpha[2, *] = s7$	
$\langle \$, 0 \rangle \langle T, 2 \rangle \langle *, 7 \rangle$	7	$i\$$	$\alpha[7, i] = s5$	
$\langle \$, 0 \rangle \langle T, 2 \rangle \langle *, 7 \rangle \langle i, 5 \rangle$	5	$\$$	$\alpha[5, \$] = r6$ $\beta[7, F] = 10$	6: $F \rightarrow i$
$\langle \$, 0 \rangle \langle T, 2 \rangle \langle *, 7 \rangle \langle F, 10 \rangle$	10	$\$$	$\alpha[10, \$] = r3$ $\beta[0, T] = 2$	3: $T \rightarrow T*F$
$\langle \$, 0 \rangle \langle T, 2 \rangle$	2	$\$$	$\alpha[2, \$] = r2$ $\beta[0, E] = 1$	2: $E \rightarrow T$
$\langle \$, 0 \rangle \langle E, 1 \rangle$	1	$\$$	$\alpha[1, \$] = \text{☺}$	Úspěch Pravý rozbor: 64632

Konstrukce LR tabulky: Úvod

- Jeden algoritmus pro syntaktickou analýzu, ale spousta algoritmů pro konstrukci LR-tabulky.

Základní algoritmy pro konstrukci LR tabulky:

- 1) **Simple LR (SLR)**: nejslabší, ale jednoduchý a vytvoří málo stavů
 - 2) **Canonical LR**: více silný, ale vytvoří poměrně hodně stavů
 - 3) **Lookahead LR (LALR)**: nejlepší, protože nejsilnější a vytvoří stejný počet stavů jako SLR
-

Rozšířená gramatika s „hloupým“ pravidlem

Myšlenka: Gramatika se speciálním „startovacím pravidlem“

Definice: Necht' $G = (N, T, P, S)$ je BKG, $S' \notin N$.
Rozšířená gramatika pro G je gramatika
 $G' = (N \cup \{S'\}, T, P \cup \{S' \rightarrow S\}, S')$.

Proč hloupé pravidlo? Až je použito pravidlo $S' \rightarrow S$ a vstupní token je ukončovač řetězce, potom je syntaktická analýza úspěšně dokončena.

Příklad:

$K = (N, T, P, S)$, where $N = \{S, A\}$, $T = \{i, o, (,)\}$,
 $P = \{1: S \rightarrow SoA, 2: S \rightarrow A, 3: A \rightarrow i, 4: A \rightarrow (S)\}$

Rozšířená gramatika pro K :

$H = (N, T, P, S')$, kde $N = \{S', S, A\}$, $T = \{i, o, (,)\}$,
 $P = \{0: S' \rightarrow S, 1: S \rightarrow SoA, 2: S \rightarrow A, 3: A \rightarrow i, 4: A \rightarrow (S)\}$

Konstrukce LR tabulky: Položky

Myšlenka: Položka je pravidlo s tečkou • na pravé straně pravidla.

Definice: Necht' $G = (N, T, P, S)$ je BKG, $A \rightarrow x \in P, x = yz$. Potom $A \rightarrow y \bullet z$ je *položka*.

Příklad: Uvažujme $S \rightarrow SoA$

Všechny položky pro pravidlo $S \rightarrow SoA$ jsou:

$S \rightarrow \bullet SoA, S \rightarrow S \bullet oA, S \rightarrow So \bullet A, S \rightarrow SoA \bullet$

Význam: $A \rightarrow y \bullet z$ říká, že pokud y se vyskytuje na zásobníku a prefix zbytku vstupního řetězce se dá postupně zredukovat na z , potom yz ($= x$) může být zredukováno na A užitím pravidla $A \rightarrow x$.

Uzávěr položek: Algoritmus

Pozn.: Uzávěr položky I , $Closure(I)$ je množina položek definována pomocí následujícího algoritmu:

- **Vstup:** $G = (N, T, P, S)$; položka I
 - **Výstup:** $Closure(I)$
-

- **Metoda:**
- $Closure(I) := \{I\};$
- Používej následující pravidlo, dokud bude možné měnit množinu $Closure(I)$:
 - if $A \rightarrow y \bullet Bz \in Closure(I)$ and $B \rightarrow x \in P$
 then přidej položku $B \rightarrow \bullet x$ do $Closure(I)$

Uzávěr položek: Příklad 1/2

$H = (N, T, P, S')$, kde $N = \{S', S, A\}$, $T = \{i, o, (,)\}$,
 $P = \{0: S' \rightarrow S, 1: S \rightarrow SoA, 2: S \rightarrow A, 3: A \rightarrow i, 4: A \rightarrow (S)\}$

Určeme: $Closure(I)$ for $I = S' \rightarrow \bullet S$

$Closure(I) := \{S' \rightarrow \bullet S\}$

1) $S' \rightarrow \bullet S \in Closure(I)$ & $S \rightarrow SoA \in P$:
 přidej $S \rightarrow \bullet SoA$ do $Closure(I)$

$Closure(I) = \{S' \rightarrow \bullet S, S \rightarrow \bullet SoA\}$

2) $S' \rightarrow \bullet S \in Closure(I)$ & $S \rightarrow A \in P$:
 přidej $S \rightarrow \bullet A$ do $Closure(I)$

$Closure(I) = \{S' \rightarrow \bullet S, S \rightarrow \bullet SoA, S \rightarrow \bullet A\}$

Uzávěr položek: Příklad 2/2

$H = (N, T, P, S')$, kde $N = \{S', S, A\}$, $T = \{i, o, (,)\}$,
 $P = \{0: S' \rightarrow S, 1: S \rightarrow SoA, 2: S \rightarrow A, 3: A \rightarrow i, 4: A \rightarrow (S)\}$

3) $S \rightarrow \bullet A \in \text{Closure}(I)$ & $A \rightarrow i \in P$:
 přidej $A \rightarrow \bullet i$ do $\text{Closure}(I)$

$\text{Closure}(I) = \{S' \rightarrow \bullet S, S \rightarrow \bullet SoA, S \rightarrow \bullet A, A \rightarrow \bullet i\}$

4) $S \rightarrow \bullet A \in \text{Closure}(I)$ & $A \rightarrow (S) \in P$:
 přidej $A \rightarrow \bullet (S)$ do $\text{Closure}(I)$

Celkově:

$\text{Closure}(I) = \{S' \rightarrow \bullet S, S \rightarrow \bullet SoA, S \rightarrow \bullet A, A \rightarrow \bullet i, A \rightarrow \bullet (S)\}$

Množina $\Theta_U(I)$ pro G

Myšlenka: Pro symbol U a množinu položek I , $\Theta_U(I)$ značí sjednocení všech uzávěrů tvaru $Closure(A \rightarrow yU \bullet z)$, kde $A \rightarrow y \bullet Uz \in I$.

Definice: Necht' $G = (N, T, P, S)$ je BKG, I je množina položek a $U \in T \cup N$. Potom

$$\Theta_U(I) = \{j: j \in Closure(A \rightarrow yU \bullet z), A \rightarrow y \bullet Uz \in I\}$$

Příklad:

$H = (N, T, P, S')$, kde $N = \{S', S, A\}$, $T = \{i, o, (,)\}$,
 $P = \{0: S' \rightarrow S, 1: S \rightarrow SoA, 2: S \rightarrow A, 3: A \rightarrow i, 4: A \rightarrow (S)\}$,
 $I = \{S \rightarrow So \bullet A, S \rightarrow \bullet A, A \rightarrow \bullet (S)\}$

Určeme: $\Theta_A(I)$

$$Closure(S \rightarrow SoA \bullet) \cup Closure(S \rightarrow A \bullet) = \{S \rightarrow SoA \bullet, S \rightarrow A \bullet\}$$

Určeme: $\Theta_{(}(I)$

$$Closure(A \rightarrow (\bullet S)) = \{A \rightarrow (\bullet S), S \rightarrow \bullet SoA, S \rightarrow \bullet A, A \rightarrow \bullet i, A \rightarrow \bullet (S)\}$$

Množina Θ_G pro gramatiku G

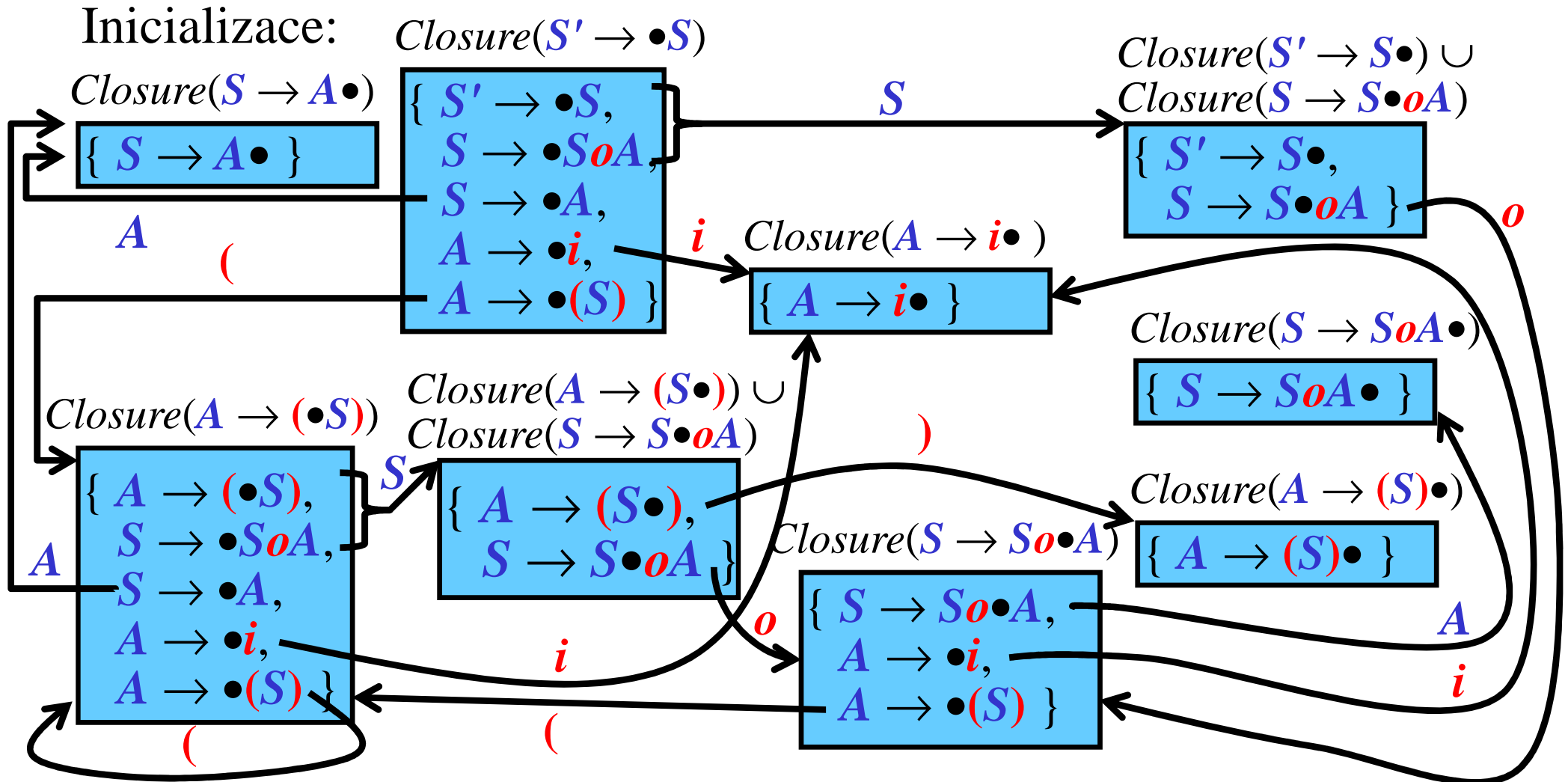
Pozn.: Množina Θ_G pro gramatiku G je množina množin položek definovaných následujícím algoritmem:

- **Vstup:** Rozšířená $G = (N, T, P, S')$
 - **Výstup:** Θ_G pro gramatiku G
-
- **Metoda:**
 - $\Theta_G := \{ \text{Closure}(S' \rightarrow \bullet S) \};$
 - for each $I \in \Theta_G$ and $U \in N \cup T$
 - if $\Theta_U(I) \neq \emptyset$ then přidej $\Theta_U(I)$ do Θ_G

Množina Θ_G : Příklad

$H = (N, T, P, \mathbf{S'})$, kde $N = \{\mathbf{S'}, S, A\}$, $T = \{\mathbf{i}, \mathbf{o}, (,)\}$,

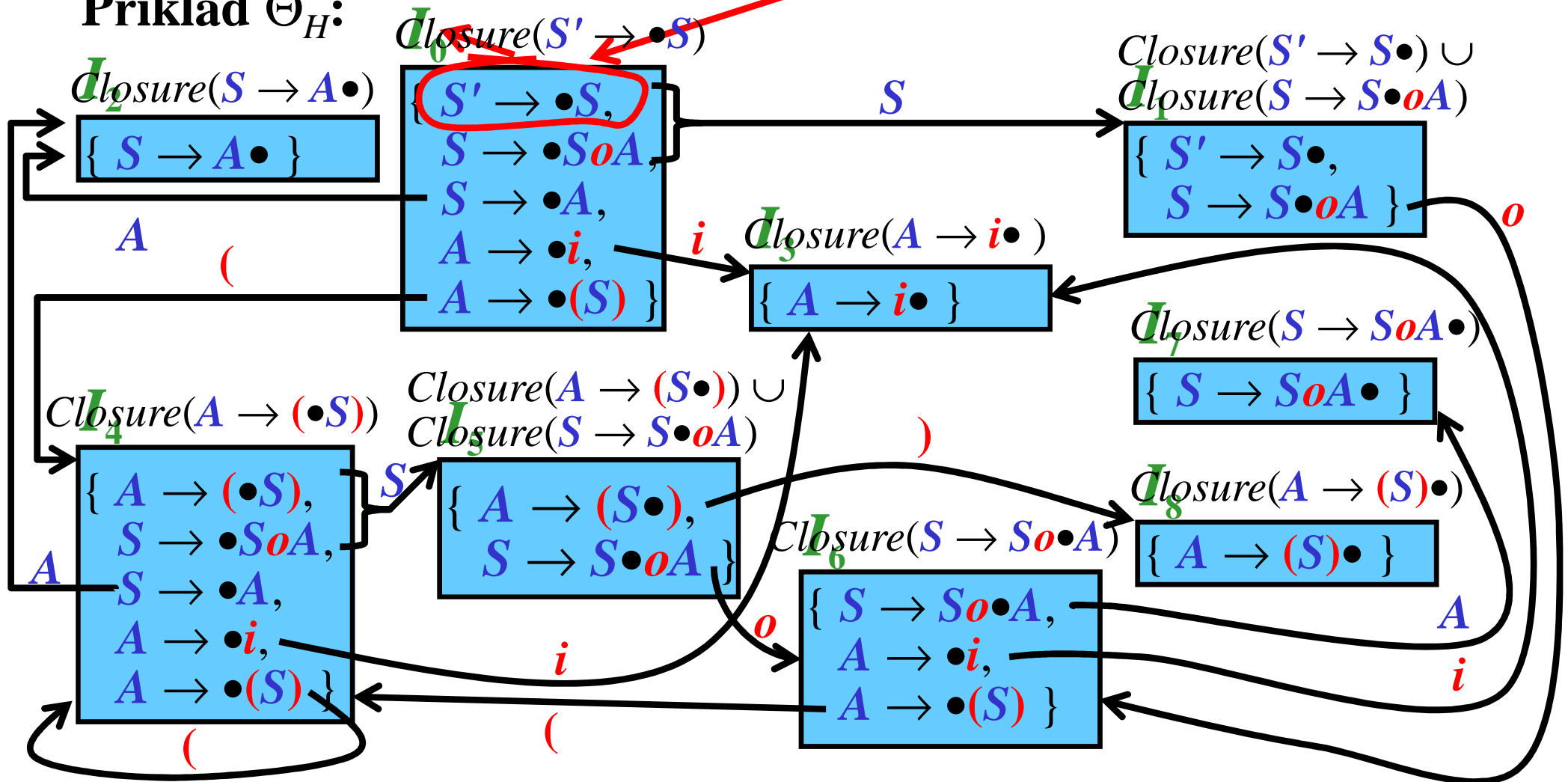
$P = \{\mathbf{0}: \mathbf{S'} \rightarrow S, \mathbf{1}: S \rightarrow S\mathbf{o}A, \mathbf{2}: S \rightarrow A, \mathbf{3}: A \rightarrow \mathbf{i}, \mathbf{4}: A \rightarrow (S)\}$



Pojmenování členů množiny Θ_G

Pojmenujte prvky Θ_G jako I_0 až I_n , kde $n+1$ je počet prvků (množin) v Θ_G . Množinu obsahující $S' \rightarrow \bullet S$ označme I_0 .

Příklad Θ_H :



Konstrukce LR tabulky: SLR Algoritmus

- **Vstup:** Rozšířená gramatika $G = (N, T, P, S')$; Θ_G ;
 $Follow(A)$ pro všechna $A \in N$
- **Výstup:** LR tabulka pro G (α = akční č., β = přechodová č.)

• Metoda:

- $StatesOfTable := \Theta_G$; $StartState := Closure(S' \rightarrow \bullet S)$;
- for each $x \in \Theta_G$ do
- for each $I \in x$ do
 - case I of
 - $I = A \rightarrow y \bullet Xz$, kde $X \in N$:
 $\beta[x, X] := \Theta_X(x)$
 - $I = A \rightarrow y \bullet Xz$, kde $X \in T$:
 $\alpha[x, X] := s \Theta_X(x)$
 - $I = S' \rightarrow S \bullet$: $\alpha[x, \$] := \text{☺}$
 - $I = A \rightarrow y \bullet$ ($A \neq S'$):
for each $a \in Follow(A)$ do $\alpha[x, a] := rp$,
 kde p je návěští pravidla $A \rightarrow y$

Konstrukce LR tabulky: Příklad 1/5

$$\begin{aligned} \Theta_H = \{ & I_0: \{ S' \rightarrow \bullet S, S \rightarrow \bullet S o A, S \rightarrow \bullet A, A \rightarrow \bullet i, A \rightarrow \bullet (S) \}, \\ & I_1: \{ S' \rightarrow S \bullet, S \rightarrow S \bullet o A \}, I_2: \{ S \rightarrow A \bullet \}, I_3: \{ A \rightarrow i \bullet \}, \\ & I_4: \{ A \rightarrow (\bullet S), S \rightarrow \bullet S o A, S \rightarrow \bullet A, A \rightarrow \bullet i, A \rightarrow \bullet (S) \}, \\ & I_5: \{ A \rightarrow (S \bullet), S \rightarrow S \bullet o A \}, I_6: \{ S \rightarrow S o \bullet A, A \rightarrow \bullet i, A \rightarrow \bullet (S) \}, \\ & I_7: \{ S \rightarrow S o A \bullet \}, I_8: \{ A \rightarrow (S) \bullet \} \} \end{aligned}$$

Určeme: LR tabulku pro K

	α					β	
	i	o	$($	$)$	$\$$	S	A
I_0	sI_3		sI_4			I_1	I_2

$S' \rightarrow \bullet S \in I_0, S \in N: \beta[I_0, S] := \Theta_S(I_0) = I_1$

$S \rightarrow \bullet S o A \in I_0, S \in N: \beta[I_0, S] := \Theta_S(I_0) = I_1$

$S \rightarrow \bullet A \in I_0, A \in N: \beta[I_0, A] := \Theta_A(I_0) = I_2$

$A \rightarrow \bullet i \in I_0, i \in T: \alpha[I_0, i] := s\Theta_i(I_0) = sI_3$

$A \rightarrow \bullet (S) \in I_0, (\in T: \alpha[I_0, (] := s\Theta_{(}(I_0) = sI_4$

Konstrukce LR tabulky: Příklad 2/5

$$\begin{aligned} \Theta_H = \{ & I_0: \{ S' \rightarrow \bullet S, S \rightarrow \bullet SoA, S \rightarrow \bullet A, A \rightarrow \bullet i, A \rightarrow \bullet (S) \}, \\ & I_1: \{ S' \rightarrow S \bullet, S \rightarrow S \bullet oA \}, I_2: \{ S \rightarrow A \bullet \}, I_3: \{ A \rightarrow i \bullet \}, \\ & I_4: \{ A \rightarrow (\bullet S), S \rightarrow \bullet SoA, S \rightarrow \bullet A, A \rightarrow \bullet i, A \rightarrow \bullet (S) \}, \\ & I_5: \{ A \rightarrow (S \bullet), S \rightarrow S \bullet oA \}, I_6: \{ S \rightarrow So \bullet A, A \rightarrow \bullet i, A \rightarrow \bullet (S) \}, \\ & I_7: \{ S \rightarrow SoA \bullet \}, I_8: \{ A \rightarrow (S) \bullet \} \} \end{aligned}$$

Určeme: LR tabulku pro K

	α					β	
	i	o	$($	$)$	$\$$	S	A
I_0	sI_3		sI_4			I_1	I_2
I_1		sI_6			☺		

$$S' \rightarrow S \bullet \in I_1: \alpha[I_1, \$] := \text{☺}$$

$$S \rightarrow S \bullet oA \in I_1, o \in T: \alpha[I_1, o] := s\Theta_o(I_1) = sI_6$$

Konstrukce LR tabulky: Příklad 3/5

$$\begin{aligned} \Theta_H = \{ & I_0: \{S' \rightarrow \bullet S, S \rightarrow \bullet SoA, S \rightarrow \bullet A, A \rightarrow \bullet i, A \rightarrow \bullet (S)\}, \\ & I_1: \{S' \rightarrow S\bullet, S \rightarrow S\bullet oA\}, I_2: \{S \rightarrow A\bullet\}, I_3: \{A \rightarrow i\bullet\}, \\ & I_4: \{A \rightarrow (\bullet S), S \rightarrow \bullet SoA, S \rightarrow \bullet A, A \rightarrow \bullet i, A \rightarrow \bullet (S)\}, \\ & I_5: \{A \rightarrow (S\bullet), S \rightarrow S\bullet oA\}, I_6: \{S \rightarrow So\bullet A, A \rightarrow \bullet i, A \rightarrow \bullet (S)\}, \\ & I_7: \{S \rightarrow SoA\bullet\}, I_8: \{A \rightarrow (S)\bullet\} \} \end{aligned}$$

Určeme: LR tabulku pro K

	α					β	
	i	o	$($	$)$	$\$$	S	A
I_0	sI_3		sI_4			I_1	I_2
I_1		sI_6			☺		
I_2		$r2$		$r2$	$r2$		

$S \rightarrow A\bullet \in I_2, Follow(S) = \{o,), \$\}$:
 $\alpha[I_2, o] = \alpha[I_2,)] = \alpha[I_2, \$] := r2$

Konstrukce LR tabulky: Příklad 4/5

$$\begin{aligned} \Theta_H = & \{I_0: \{S' \rightarrow \bullet S, S \rightarrow \bullet SoA, S \rightarrow \bullet A, A \rightarrow \bullet i, A \rightarrow \bullet (S)\}, \\ & I_1: \{S' \rightarrow S\bullet, S \rightarrow S\bullet oA\}, I_2: \{S \rightarrow A\bullet\}, I_3: \{A \rightarrow i\bullet\}, \\ & I_4: \{A \rightarrow (\bullet S), S \rightarrow \bullet SoA, S \rightarrow \bullet A, A \rightarrow \bullet i, A \rightarrow \bullet (S)\}, \\ & I_5: \{A \rightarrow (S\bullet), S \rightarrow S\bullet oA\}, I_6: \{S \rightarrow So\bullet A, A \rightarrow \bullet i, A \rightarrow \bullet (S)\}, \\ & I_7: \{S \rightarrow SoA\bullet\}, I_8: \{A \rightarrow (S)\bullet\}\} \end{aligned}$$

Určeme: LR tabulku pro K

	α					β	
	i	o	$($	$)$	$\$$	S	A
I_0	sI_3		sI_4			I_1	I_2
I_1		sI_6			☺		
I_2		$r2$		$r2$	$r2$		
I_3		$r3$		$r3$	$r3$		

Zbytek tabulky
sestrojte
analogicky.

$A \rightarrow i\bullet \in I_3, \text{Follow}(A) = \{o,), \$\}:$
 $\alpha[I_3, o] = \alpha[I_3,)] = \alpha[I_3, \$] := r3$

Konstrukce LR tabulky: Příklad 5/5

Výsledná LR tabulka pro K

	α					β	
	i	o	$($	$)$	$\$$	S	A
I_0	sI_3		sI_4			I_1	I_2
I_1		sI_6			☺		
I_2		$r2$		$r2$	$r2$		
I_3		$r3$		$r3$	$r3$		
I_4	sI_3		sI_4			I_5	I_2
I_5		sI_6		sI_8			
I_6	sI_3		sI_4				I_7
I_7		$r1$		$r1$	$r1$		
I_8		$r4$		$r4$	$r4$		

Přejmenování stavů

Přejmenovat stavy:

Old	New
I_0	0
I_1	1
I_2	2
I_3	3
I_4	4
I_5	5
I_6	6
I_7	7
I_8	8

LR tabulka pro K s přejmenovanými stavy:

α	i	o	()	\$
0	s3		s4		
1		s6			☺
2		r2		r2	r2
3		r3		r3	r3
4	s3		s4		
5		s6		s8	
6	s3		s4		
7		r1		r1	r1
8		r4		r4	r4

β	S	A
0	1	2
1		
2		
3		
4	5	2
5		
6		7
7		
8		