

# 11. Синтаксический анализ. Часть 3

## Разделы:

- Метод «перенос-свертка»
- LR(0)-автоматы
- Общий алгоритм LR(1)-анализа
- Таблицы SLR-анализа
- Активные префиксы

# ПС-анализ

$$E \rightarrow E + T \mid T$$

$$T \rightarrow T * F \mid F$$

$$F \rightarrow (E) \mid \mathbf{id}$$

$\mathbf{id} \ * \ \mathbf{id}$

$F \ * \ \mathbf{id}$   
|  
 $\mathbf{id}$

$T \ * \ \mathbf{id}$   
|  
 $F$   
|  
 $\mathbf{id}$

$T \ * \ F$   
|       |  
 $F$       $\mathbf{id}$   
|  
 $\mathbf{id}$

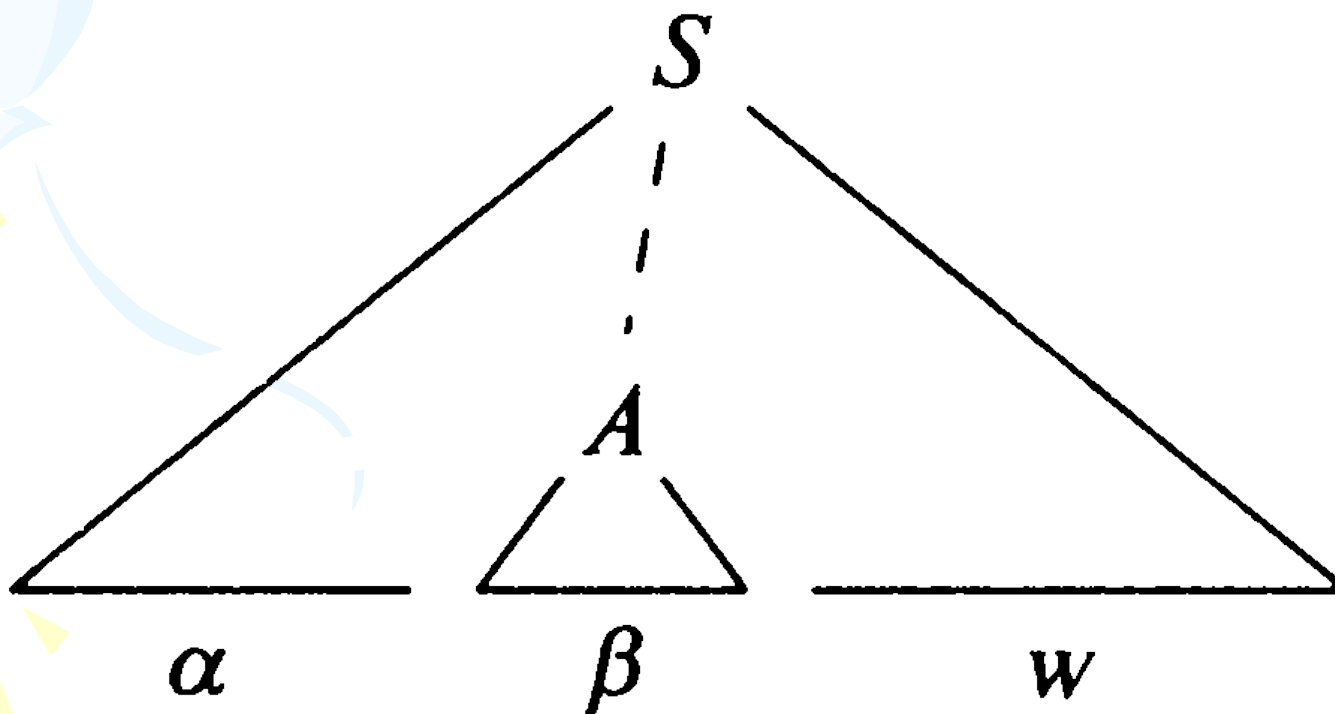
$T$   
/ | \  
 $T \ * \ F$   
|       |  
 $F$       $\mathbf{id}$   
|  
 $\mathbf{id}$

$E$   
|  
 $T$   
/ | \  
 $T \ * \ F$   
|       |  
 $F$       $\mathbf{id}$   
|  
 $\mathbf{id}$

$$E \Rightarrow_r T \Rightarrow_r T * F \Rightarrow_r T * \mathbf{id} \Rightarrow_r F * \mathbf{id} \Rightarrow_r \mathbf{id} * \mathbf{id}$$

# ПС-анализ

ПРАВАЯ СЕНТЕНЦИАЛЬНАЯ ФОРМА	ОСНОВА	СВОРАЧИВАЮЩАЯ ПРОДУКЦИЯ
$\mathbf{id_1 * id_2}$	$\mathbf{id_1}$	$F \rightarrow \mathbf{id}$
$F * \mathbf{id_2}$	$F$	$T \rightarrow F$
$T * \mathbf{id_2}$	$\mathbf{id_2}$	$F \rightarrow \mathbf{id}$
$T * F$	$T * F$	$E \rightarrow T * F$



# ПС-анализ

- Начинаем с анализируемой строки терминалов
- Если  $w$  – строка грамматики, то пусть  $w = \gamma_n$ , где  $\gamma_n$  –  $n$ -я правосентенциальная форма некоторого неизвестного правого порождения
$$S \Rightarrow_r \gamma_0 \Rightarrow_r \gamma_1 \Rightarrow_r \gamma_2 \Rightarrow_r^* \dots \Rightarrow_r \gamma_{n-1} \Rightarrow_r \gamma_n \Rightarrow_r w$$
- Для его воссоздания мы находим основу  $\beta_n$  в  $\gamma_n$  и заменяем ее LHS продукции  $A_n \rightarrow \beta_n$  для получения  $\gamma_{n-1}$
- Затем мы повторяем описанный процесс, т.е. находим в  $\gamma_{n-1}$  основу  $\beta_{n-1}$  и свертываем ее для получения  $\gamma_{n-2}$
- Если после очередного шага СФ содержит только  $S$ , то мы прекращаем процесс и сообщаем об успешном завершении анализа
- Обратная последовательность продукций, использованных в свертках, представляет собой правое порождение входной строки

# ПС-анализ

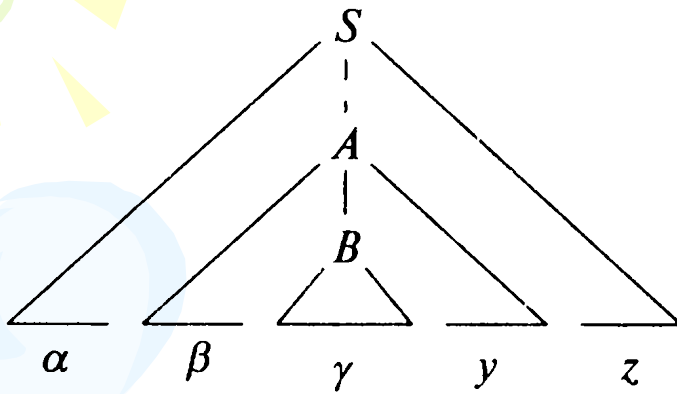
Стек	Вход	Действие
\$	$id_1 * id_2 \$$	Перенос
$\$id_1$	$* id_2 \$$	Свертка по $F \rightarrow id$
$\$F$	$* id_2 \$$	Свертка по $T \rightarrow F$
$\$T$	$* id_2 \$$	Перенос
$\$T*$	$id_2 \$$	Перенос
$\$T*id_2$	\$	Свертка по $F \rightarrow id$
$\$T*F$	\$	Свертка по $T \rightarrow T*F$
$\$T$	\$	Свертка по $E \rightarrow T$
$\$E$	\$	Принятие

- **Перенос** (*shift*) – перемещение очередного входного символа на вершину стека
- **Свертка** (*reduce*) - следующая операция
  - На вершине стека должна располагаться RHS сворачиваемой строки
  - Определяется левый конец строки в стеке и принимается решение о том, каким нетерминалом будет заменена строка.
- **Принятие** (*accept*) - объявление об успешном завершении ВСА
- Обнаружение синтаксической **ошибки** (*error*) и вероятный вызов подпрограммы восстановления после ошибки

# ПС-анализ

1.  $S \Rightarrow_r \alpha Az \Rightarrow_r \alpha \beta Byz \Rightarrow_r \alpha \beta \gamma yz$ .

2.  $S \Rightarrow_r \alpha BxAz \Rightarrow_r \alpha Bxyz \Rightarrow_r \alpha \gamma xyz$ .



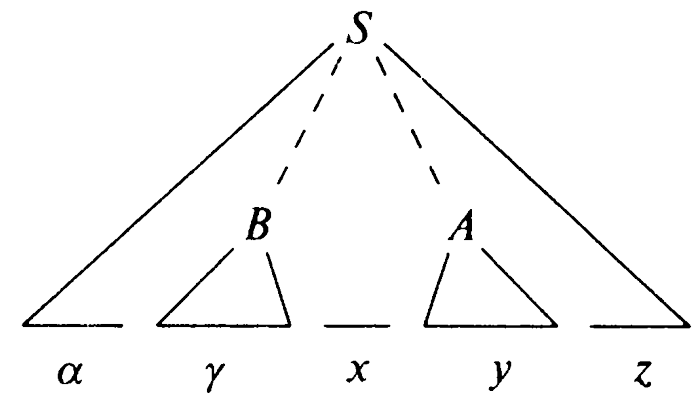
Случай (1)

Стек

...  
 $\$ \alpha \beta \gamma$   
 $\$ \alpha \beta B$   
 $\$ \alpha \beta Byz$   
 $\$ \alpha A$   
 ...

Вход

...  
 $yz\$$   
 $yz\$$   
 $z\$$   
 $z\$$



Случай (2)

Стек

...  
 $\$ \alpha \gamma$   
 $\$ \alpha B$   
 $\$ \alpha Bxy$   
 ...

Вход

...  
 $xyz\$$   
 $xyz\$$   
 $z\$$

# ПС-анализ

- Любой ПС-анализатор может достичь конфигурации, в которой синтаксический анализатор, обладая информацией о содержимом стека и очередных  $k$  входных символах, не может принять решение о том, что следует выполнить:
  - перенос или свертку (это **конфликт «перенос/свертка»**)
  - либо какое именно из нескольких сверток должно быть выполнено (**конфликт «свертка/свертка»**)

# ПС-анализ

*stmt* → **if** *expr* **then** *stmt*  
          | **if** *expr* **then** *stmt* **else** *stmt*  
          | **other**

Стек

**...if** *expr* **then** *stmt*

Вход

*else...\$*



# ПС-анализ

*stmt* → **id** ( *parameter\_list* )

*stmt* → *expr* := *expr*

*parameter\_list* → *parameter\_list* , *parameter*

*parameter\_list* → *parameter*

*parameter* → **id**

*expr* → **id** ( *expr\_list* )

*expr* → **id**

*expr\_list* → *expr\_list* , *expr*

*expr\_list* → *expr*

Стек  
...**id** ( **id**

Вход  
, *id* )...

Стек  
... **procid** ( **id**

Вход  
, *id* )...

# LR(0)-автоматы

- **LR( $k$ )-анализ**
- $L$  – просмотр входной строки слева направо,  $R$  – построение обратного правого порождения, а  $k$  – количество предпросматриваемых символов входного потока, необходимое для принятия решения
- Неформально КСГ является LR, если синтаксический анализатор, работающий слева направо методом переноса-свертки, способен распознавать основы правых СФ при их появлении на вершине стека
- Причины популярности LR-анализа:
  - Можно создавать анализаторы для распознавания основных конструкций языков программирования
  - Это наиболее общий метод ПС-анализа без возврата, который не уступает другим, более примитивным методам
  - LR-анализатор может обнаруживать ошибки на первом же некорректном входном символе
  - Класс LR-грамматик представляет собой истинное подмножество грамматик, анализируемых LL-методом

# LR(0)-автоматы

Стек	Вход	Действие
\$	$id_1 * id_2 \$$	Перенос
$\$id_1$	$* id_2 \$$	Свертка по $F \rightarrow id$
$\$F$	$* id_2 \$$	Свертка по $T \rightarrow F$
$\$T$	$* id_2 \$$	Перенос
$\$T*$	$id_2 \$$	Перенос
$\$T*id_2$	\$	Свертка по $F \rightarrow id$
$\$T*F$	\$	Свертка по $T \rightarrow T*F$
$\$T$	\$	Свертка по $E \rightarrow T$
$\$E$	\$	Принятие

## • Множество LR(0)-ситуаций

$$A \rightarrow \cdot XYZ \quad A \rightarrow \varepsilon$$

$$A \rightarrow X \cdot YZ \quad A \rightarrow \cdot$$

$$A \rightarrow XY \cdot Z$$

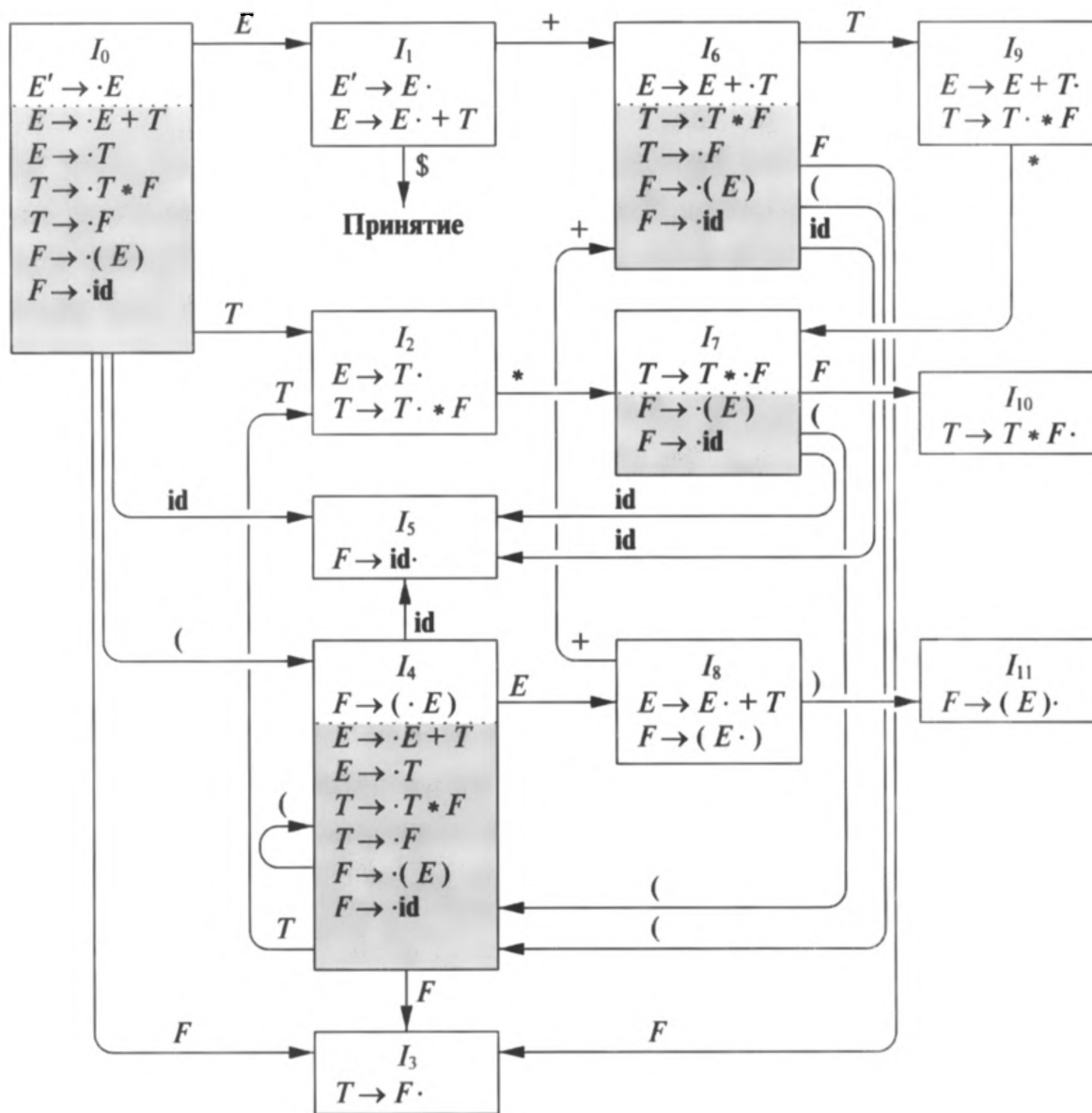
$$A \rightarrow XYZ \cdot$$



# LR(0)-автоматы

- Один набор множеств LR(0)-ситуаций - **канонический набор LR(0)**
- Обеспечивает основу для построения ДКА, который используется для принятия решений в процессе СА
- Такой автомат называется **LR(0)-автоматом**
- Каждое состояние LR(0)-автомата представляет множество ситуаций в каноническом наборе LR(0)

# LR(0)-автоматы



# LR(0)-автоматы

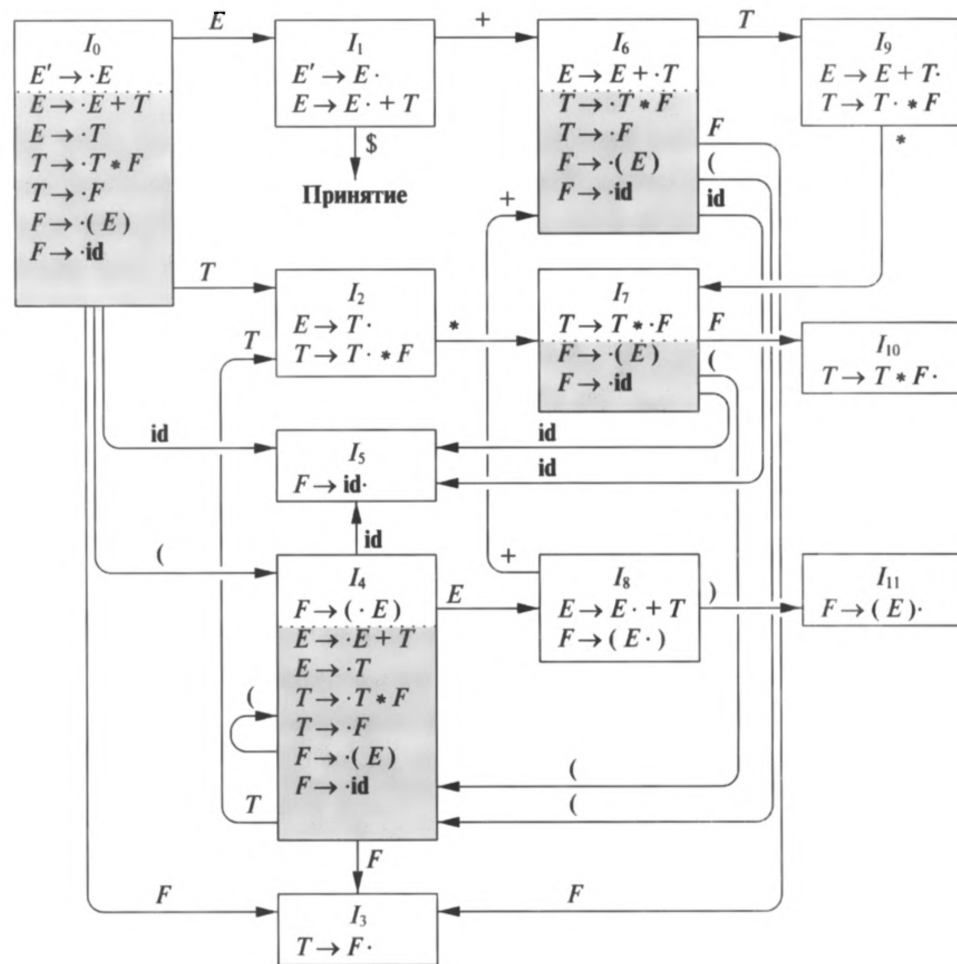
- Для построения канонического LR(0)-набора мы определяем пополненную грамматику и две функции, *CLOSURE* и *GOTO*
- Если  $G$  – грамматика с аксиомой  $S$ , то пополненная грамматика  $G'$  представляет собой  $G$  с новой аксиомой  $S'$  и продукцией  $S' \rightarrow S$
- Назначение новой аксиомы – указать синтаксическому анализатору, когда следует прекратить анализ и сообщить о принятии входной строки
  - принятие осуществляется тогда и только тогда, когда синтаксический анализатор выполняет свертку с использованием продукции  $S' \rightarrow S$

# LR(0)-автоматы

- Если  $I$  – множество ситуаций грамматики  $G$ , то  $CLOSURE(I)$  – множество ситуаций, построенное из  $I$  по двум правилам
  1. Изначально в  $CLOSURE(I)$  добавляются все элементы  $I$
  2. Если  $A \rightarrow a \bullet B \beta$  входит в  $CLOSURE(I)$ , а  $B \rightarrow \gamma$  является продукцией, то в  $CLOSURE(I)$  добавляется ситуация  $B \rightarrow \bullet \gamma$ , если ее там еще нет
    - Это правило применяется до тех пор, пока не останется ситуаций для добавления в  $CLOSURE(I)$ .
- Неформально,  $A \rightarrow a \bullet B \beta$  в  $CLOSURE(I)$  указывает, что в некоторой точке процесса синтаксического анализа мы полагаем, что далее во входной строке мы можем встретить подстроку, порождаемую из  $B \beta$
- Эта подстрока имеет префикс, порождаемый из  $B$  путем применения одной из  $B$ -продукций
- Мы добавляем ситуации для всех  $B$ -продукций
  - если  $B \rightarrow \gamma$  является продукцией, то мы включаем  $B \rightarrow \bullet \gamma$  в  $CLOSURE(I)$

# LR(0)-автоматы

$E' \rightarrow E$   
 $E \rightarrow E + T \mid T$   
 $T \rightarrow T * F \mid F$   
 $F \rightarrow (E) \mid \text{id}$





# LR(0)-автоматы

SetOfItems CLOSURE(I)

```
{
  J = I;
  repeat
    foreach (пункт  $A \rightarrow a \cdot B\beta$  из J)
      foreach (продукция  $B \rightarrow \gamma$  из G)
        if ( $B \rightarrow \cdot \gamma$  не входит в J)
          Добавить  $B \rightarrow \cdot \gamma$  в J;
  until больше нет пунктов для добавления в J за один проход;
  return J;
}
```

- Удобный способ реализации функции *closure* состоит в поддержании булева массива *added*, индексированного нетерминалами грамматики *G*
- *added[B]* устанавливается равным *true*, если (и когда) мы добавляем ситуацию  $B \rightarrow \cdot \gamma$  для каждой *B*-продукции  $B \rightarrow \gamma$

# LR(0)-автоматы

- Если одна  $B$ -продукция добавляется в замыкание  $I$  с точкой на левом конце, то в замыкание будут аналогичным образом добавлены все  $B$ -продукции
- Значит, при некоторых условиях нет необходимости в перечислении ситуаций  $B \rightarrow \bullet \gamma$ , добавленных в  $I$  функцией  $CLOSURE$
- Достаточно списка нетерминалов  $B$ , продукции которых были добавлены таким способом
- Множество ситуаций разделим на два класса
  1. **Базисные ситуации**, или **ситуации ядра** (*kernel items*): начальная ситуация  $S' \rightarrow \bullet S$  и все ситуации, у которых точки расположены не у левого края
  2. **Небазисные** (*nonkernel items*) **ситуации**, у которых точки расположены слева, за исключением  $S' \rightarrow \bullet S$

# LR(0)-автоматы

- Вторая функция -  $GOTO(I, X)$ , где  $I$  – множество ситуаций, а  $X$  – символ грамматики
- $GOTO(I, X)$  определяется как замыкание множества всех ситуаций  $[A \rightarrow aX \bullet \beta]$ , таких, что  $[A \rightarrow aX \bullet \beta]$  находится в  $I$
- Неформально, функция  $GOTO$  используется для определения переходов в LR(0)-автомате грамматики
- Состояния автомата соответствуют множествам ситуаций, и  $GOTO(I, X)$  указывает переход из состояния  $I$  при входном символе  $X$

# LR(0)-автоматы

- Если  $I$  – множества из двух ситуаций  $\{[E' \rightarrow E \bullet], [E \rightarrow E \bullet + T]\}$ , то  $GOTO(I, +)$  содержит ситуации:

$$E \rightarrow E + \cdot T$$

$$T \rightarrow \cdot T * F$$

$$T \rightarrow \cdot F$$

$$F \rightarrow \cdot (E)$$

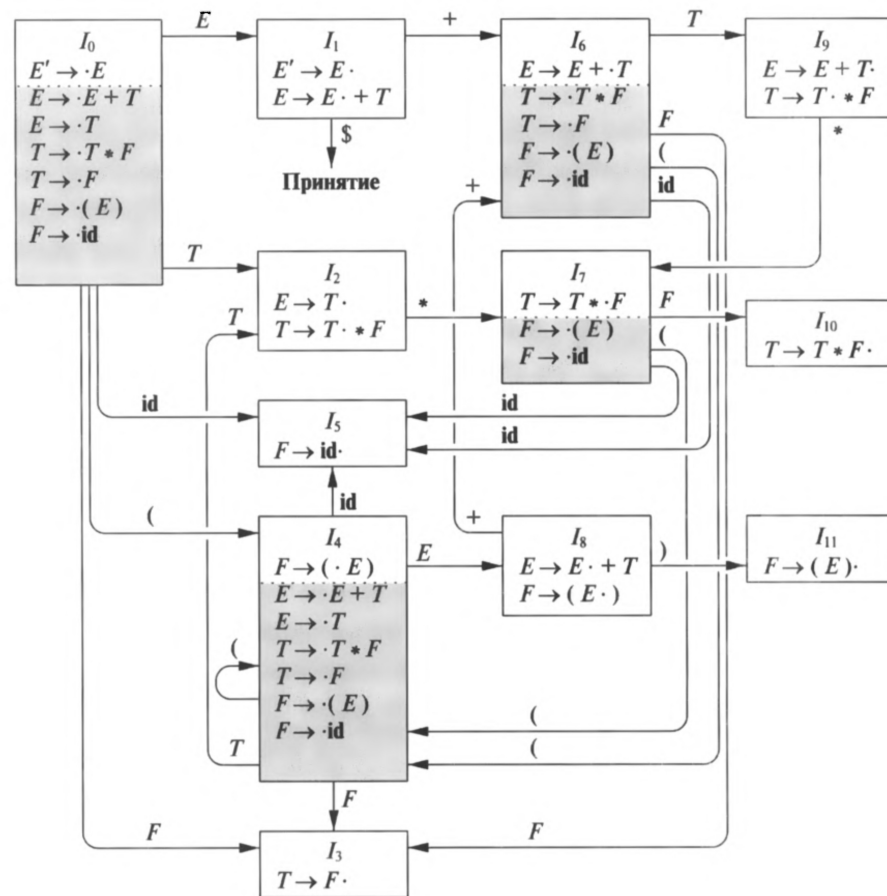
$$F \rightarrow \cdot \mathbf{id}$$

# LR(0)-автоматы

```
void items(G')
{
    C = {closure({[S' -> • S]})};
    repeat
        foreach (множество ситуаций I в C )
            foreach (грамматический символ X)
                if (множество GOTO (I, X) не пустое и не входит в C )
                    Добавить GOTO(I, X) в C;
    until нет новых множеств ситуаций для добавления в C за один проход;
}
```

# LR(0)-автоматы

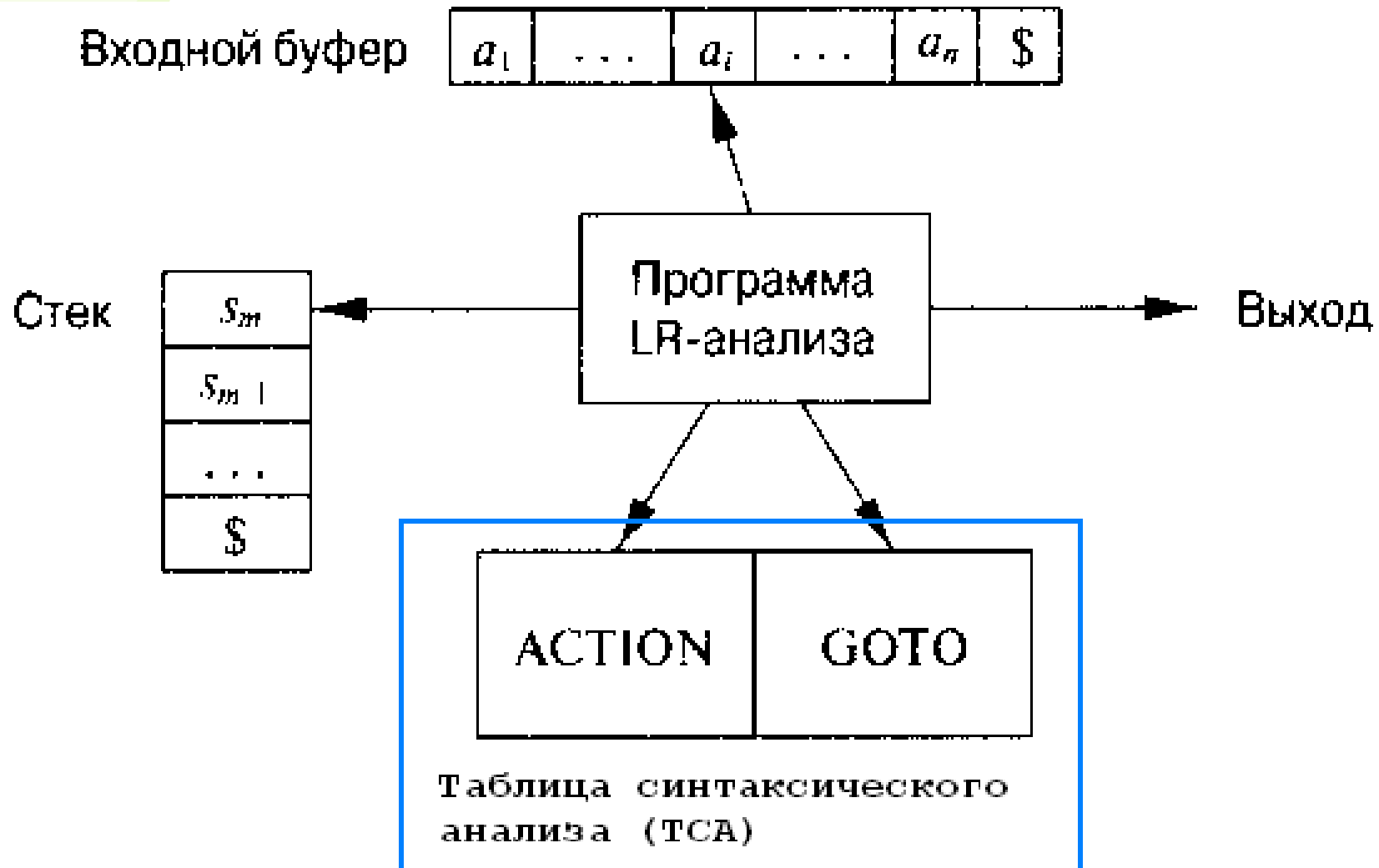
- Стартовое состояние LR(0)-автомата –  $CLOSURE(\{[S' \rightarrow \bullet S]\})$ 
  - $S'$  – новая аксиома
- Все состояния являются принимающими
- Под  $j$  мы понимаем состояние, соответствующее множеству ситуаций  $I_j$
- Пусть строка  $\gamma$  из символов грамматики переводит LR(0)-автомат из состояния 0 в некоторое состояние  $j$
- Тогда выполним **перенос** символа  $a$ , если состояние  $j$  имеет для него переход
- В противном случае выбирается **свертка**
  - Ситуация в состоянии  $j$  говорит нам, какую продукцию следует для этого использовать



# LR(0)-автоматы

СТРОКА	СТЕК	СИМВОЛЫ	ВХОД	ДЕЙСТВИЕ
(1)	0	\$	<b>id * id</b> \$	Перенос в 5
(2)	0 5	<b>\$id</b>	<b>*id</b> \$	Свертка по $F \rightarrow \mathbf{id}$
(3)	0 3	<b>\$F</b>	<b>*id</b> \$	Свертка по $T \rightarrow F$
(4)	0 2	<b>\$T</b>	<b>*id</b> \$	Перенос в 7
(5)	0 2 7	<b>\$T*</b>	<b>id</b> \$	Перенос в 5
(6)	0 2 7 5	<b>\$T * id</b>	\$	Свертка по $F \rightarrow \mathbf{id}$
(7)	0 2 7 10	<b>\$T * F</b>	\$	Свертка по $T \rightarrow T * F$
(8)	0 2	<b>\$T</b>	\$	Свертка по $E \rightarrow T$
(9)	0 1	<b>\$E</b>	\$	Принятие

# LR-анализ





# LR-анализ

- ТСА состоит из двух частей
  - 1) Функция *ACTION* принимает в качестве аргумента состояние  $i$  и терминал  $a$  (или маркер конца входной строки)
    - Значение  $ACTION[i, a]$  может быть одного из следующих видов:
      - 1а) **Перенос** в состояние  $j$
      - 1б) **Свертка** согласно продукции  $A \rightarrow \beta$
      - 1в) **Принятие**
      - 1г) **Ошибка**
  - 2) Функция *GOTO*, определенная на множествах ситуаций, распространяется на состояния
    - если  $GOTO[Ii, A] = Ij$ , то *GOTO* отображает также состояние  $i$  и нетерминал  $A$  на состояние  $j$

# LR-анализ

- Конфигурация LR-анализатора - пара  $(s_0s_1\dots s_m, a_ia_{i+1}\dots a_n\$)$
- Эта конфигурация представляет собой правую СФ  $X_1X_2 \dots X_ma_ia_{i+1}\dots a_n$
- Стартовое состояние синтаксического анализатора  $s_0$  не представляет символ грамматики, а служит маркером дна стека
- Очередной шаг анализатора из конфигурации определяется текущим входным символом  $a_i$  и состоянием на вершине стека  $s_m$  путем обращения к записи  $ACTION[s_m, a_i]$  в таблице действий

# LR-анализ

- В результате обращения к  $ACTION[s_m, a_i]$  получаются конфигурации (всего – 4):
  1. Если  $ACTION[...]$  = перенос  $s$ , то анализатор выполняет перенос в стек состояния  $s$  и очередной конфигурацией становится  $(s_0s_1...s_ms, a_{i+1}... a_n\$)$

# LR-анализ

- В результате обращения к  $ACTION[s_m, a_i]$  получаются конфигурации (всего – 4):
  2. Если  $ACTION[...]$  = свертка  $A \rightarrow \beta$ , то анализатор выполняет свертку и его конфигурацией становится  $(s_0 s_1 \dots s_{m-r} s, a_i a_{i+1} \dots a_n \$)$ 
    - $r = |\beta|$ , а  $s = GOTO[s_{m-r}, A]$
    - Анализатор вначале снимает  $r$  символов состояний с вершины стека, что переносит на вершину стека состояние  $s_{m-r}$ , после чего на вершину стека помещается  $s$ , запись из  $GOTO[s_{m-r}, A]$
    - Текущий входной символ не изменяется
    - В LR-анализаторах последовательность символов  $X_{m-r+1} \dots X_m$  всегда соответствует  $\beta$ , RHS продукции свертки

# LR-анализ

- В результате обращения к  $ACTION[s_m, a_i]$  получаются конфигурации(всего – 4):
  3. Если  $ACTION[...]$  = принятие, то синтаксический анализ завершается
  4. Если  $ACTION[...]$  = ошибка, то анализатор обнаруживает ошибку и вызывает подпрограмму восстановления после ошибки

# LR-анализ

- Алгоритм LR-анализа
  - ВХОД: входная строка  $w$  и таблица LR-анализа с функциями *ACTION* и *GOTO* для грамматики  $G$
  - ВЫХОД: если  $w$  принадлежит  $L(G)$ , то шаги свертки восходящего синтаксического анализа  $w$ ; в противном случае – указание на ошибку
  - МЕТОДИКА: изначально в стеке анализатора находится начальное состояние  $s_0$ , а во входном буфере –  $w\$$
  - Далее выполняется программа (псевдокод – на следующем слайде)

# LR-анализ

Пусть  $a$  — первый символ  $w\$$

```
while (!0) /* Бесконечный цикл */
```

```
{
```

Пусть  $s$  — состояние на вершине стека.

```
if (ACTION[s, a] = перенос t)
```

```
{
```

Внести  $t$  в стек.

Присвоить  $a$  очередной входной символ.

```
}
```

```
else if (ACTION[s, a] = свертка  $A \rightarrow \beta$ )
```

```
{
```

Снять  $|\beta|$  символов со стека.

Теперь на вершине стека находится состояние  $t$ .

Внести в стек GOTO  $[t, A]$ .

Вывести продукцию  $A \rightarrow \beta$ .

```
}
```

```
else if (ACTION[s, a] = принятие)
```

```
{
```

```
break; /* Синтаксический анализ завершен */
```

```
}
```

```
else
```

```
{
```

Вызов подпрограммы восстановления после ошибки.

```
}
```

```
}
```

# LR-анализ

JFLAP : (SLR1-1.jfl)

File Input Test Convert Help

Editor Build SLR(1) Parse

Do Selected Do Step Do All Next Parse

Parse table complete. Press "parse" to use it.

Parse table:

	FIRST	FOLLOW
E	{ (, i }	{ \$, +, ) }
F	{ (, i }	{ \$, *, +, ) }
T	{ (, i }	{ \$, *, +, ) }

LR(0) items and transitions:

```

graph LR
    q0((q0  
F->(E)  
F->i  
T->F  
E->E+T  
T->T*F  
E->E))
    q1((q1  
F->(E)  
F->(E)  
F->i  
T->F  
E->E+T  
T->T*F))
    q2((q2  
E->E+T  
E->E))
    q3((q3  
T->F))
    q4((q4  
E->T  
T->T*F))
    q5((q5  
F->i))
    q6((q6  
E->E+T  
F->(E)))
    q7((q7  
F->(E)  
F->i  
T->F  
E->E+T  
T->T*F))
    q8((q8  
F->(E)  
F->i  
T->T*F))
    q9((q9  
F->i))
    q10((q10  
E->E+T  
T->T*F))
    q11((q11  
T->T*F))

    q0 -- "(" --> q1
    q0 -- "i" --> q5
    q0 -- "+" --> q2
    q0 -- ")" --> q7
    q0 -- "*" --> q4
    q1 -- "(" --> q1
    q1 -- "i" --> q5
    q1 -- "+" --> q2
    q1 -- ")" --> q7
    q1 -- "*" --> q4
    q1 -- "E" --> q6
    q2 -- "E" --> q2
    q3 -- "F" --> q3
    q4 -- "E" --> q2
    q4 -- "T" --> q4
    q5 -- "F" --> q5
    q6 -- "(" --> q1
    q6 -- "i" --> q5
    q6 -- "+" --> q2
    q6 -- ")" --> q7
    q6 -- "*" --> q4
    q6 -- "E" --> q6
    q7 -- "(" --> q1
    q7 -- "i" --> q5
    q7 -- "+" --> q2
    q7 -- ")" --> q7
    q7 -- "*" --> q4
    q7 -- "E" --> q10
    q8 -- "(" --> q1
    q8 -- "i" --> q5
    q8 -- "+" --> q2
    q8 -- ")" --> q7
    q8 -- "*" --> q4
    q8 -- "E" --> q8
    q9 -- "F" --> q9
    q10 -- "E" --> q2
    q10 -- "T" --> q10
    q11 -- "T" --> q11
  
```

LR(0) automaton diagram showing states q0 through q11 and transitions for terminals (, i, +, ), \*, and non-terminals E, F, T.

LR(0) item sets and transitions:

	(	)	*	+	i	\$	E	F	T
0	s1				s5		2	3	4
1	s1				s5		6	3	4
2				s7		acc			
3		r4	r4	r4		r4			
4		r2	s8	r2		r2			
5		r6	r6	r6		r6			
6		s9		s7					
7	s1				s5			3	10
8	s1				s5			11	
9		r5	r5	r5		r5			
10		r1	s8	r1		r1			
11		r3	r3	r3		r3			



# LR-анализ

**JFLAP : (SLR1-1.jff)**

File Input Test Convert Help

Editor Build SLR(1) Parse **SLR(1) Parsing**

Table Text Size

	(	)	*	+	i	\$	E	F	T
0	s1				s5		2	3	4
1	s1				s5		6	3	4
2				s7		acc			
3		r4	r4	r4		r4			
4		r2	s8	r2		r2			

Start Step Derivation Table

Input  $i*i+i$

Input Remaining \$

Stack E0

LHS		RHS	Production	Derivation
E'	→	E		$i*i+i$
E	→	E+T	$F \rightarrow i$	$F*i+i$
E	→	T	$T \rightarrow F$	$T*i+i$
T	→	T*F	$F \rightarrow i$	$T*iF+i$
T	→	F	$T \rightarrow T*F$	$T+i$
F	→	(E)	$E \rightarrow T$	$E+i$
F	→	i	$F \rightarrow i$	$E+F$
			$T \rightarrow F$	$E+T$
			$E \rightarrow E+T$	E

String accepted

# LR-анализ

- ТСА, построенные методом  $SLR(1)$  - **SLR-таблицы**
- LR-анализатор, использующий  $SLR$ -таблицы, – **SLR-анализатор**
- $SLR$ -метод начинается с  $LR(0)$ -ситуаций и  $LR(0)$ -автомата
  - Для данной грамматики  $G$  мы строим ее расширение  $G'$  с новой аксиомой  $S'$
  - Для  $G'$  строится канонический набор  $C$  множеств ситуаций  $G'$  вместе с функцией  $GOTO$
  - Затем строятся записи  $ACTION$  и  $GOTO$  в ТСА с использованием алгоритма, который требует знания  $FOLLOW(A)$  для каждого нетерминала  $A$  грамматики

# LR-анализ

- ВХОД: расширенная грамматика  $G'$
- ВЫХОД: функции таблицы SLR-анализа *ACTION* и *GOTO* для  $G'$
- *Методика*: выполняются действия (всего – 5):
  1. Строится  $C = \{I_0, I_1, \dots, I_n\}$  – набор множеств LR(0)-ситуаций для  $G'$

# LR-анализ

- ВХОД: расширенная грамматика  $G'$
- ВЫХОД: функции таблицы SLR-анализа  $ACTION$  и  $GOTO$  для  $G'$
- Методика: выполняются действия (всего – 5):
  2. Из  $Ii$  строится состояние  $i$ , а действие  $CA$  для состояния  $i$  определяем так:
    - а) Если  $[A \rightarrow a \bullet a\beta]$  принадлежит  $Ii$  и  $GOTO(Ii, a) = Ij$ , то устанавливаем  $ACTION[i, a] = \text{«перенос } j\text{»}$
    - б) Если  $[A \rightarrow a \bullet]$  принадлежит  $Ii$ , то устанавливаем  $ACTION[i, a] = \text{«свертка } A \rightarrow a\text{»}$  для всех  $a$  из  $FOLLOW(A)$
    - в) Если  $[S' \rightarrow S \bullet]$  принадлежит  $Ii$ , то устанавливаем  $ACTION[i, \$] = \text{«принятие»}$

# LR-анализ

- ВХОД: расширенная грамматика  $G'$
- ВЫХОД: функции таблицы SLR-анализа  $ACTION$  и  $GOTO$  для  $G'$
- Методика: выполняются действия (всего – 5):
  3. Переходы для состояния  $i$  строим для всех нетерминалов  $A$  с использованием следующего правила: если  $GOTO(Ii, A) = Ij$ , то  $GOTO[i, A] = j$
  4. Все записи, не определенные правилами 2 и 3, получают значение «ошибка»
  5. Начальное состояние синтаксического анализатора строится из множества ситуаций, содержащего  $[S' \rightarrow \bullet S]$

# LR-анализ

JFLAP : (SLR1-1.jff)

File Input Test Convert Help

Editor Build SLR(1) Parse

Do Selected Do Step Do All Next Parse

Parse table complete. Press "parse" to use it.

	FIRST	FOLLOW
E	{ (, ) }	{ \$, *, + }
F	{ (, ) }	{ \$, *, + }
T	{ (, ) }	{ \$, *, + }

$E' \rightarrow E$   
 $E \rightarrow E+T$   
 $E \rightarrow T$   
 $T \rightarrow T*F$   
 $T \rightarrow F$   
 $F \rightarrow (E)$   
 $F \rightarrow i$

	(	)	*	+	
0	s1				s7
1	s1				r4
2					r4
3					r2
4					r6
5					r6
6					s7
7	s1				
8	s1				
9			r5	r5	r5
10			r1	s8	r1
11			r3	r3	r3

JFLAP : (SLR1-1.jff)

File Input Test Convert Help

Editor Build SLR(1) Parse SLR(1) Parsing

Table Text Size

Start Step Derivation Table

	(	)	*	+	i	\$	E	F	T
0	s1				s5		2	3	4
1	s1				s5		6	3	4
2					s7	acc			
3					r4				
4					r2				

Input:  $i*i+i$

Input Remaining:  $$$

Stack:  $E0$

LHS	RHS	Production	Derivation
$E'$	$\rightarrow E$		$i*i+i$
$E$	$\rightarrow E+T$	$F \rightarrow i$	$F*i+i$
$E$	$\rightarrow T$	$T \rightarrow F$	$T*i+i$
$T$	$\rightarrow T*F$	$F \rightarrow i$	$T*i*F+i$
$T$	$\rightarrow F$	$T \rightarrow T*F$	$T+i$
$F$	$\rightarrow (E)$	$E \rightarrow T$	$E+i$
$F$	$\rightarrow i$	$F \rightarrow i$	$E+F$
		$T \rightarrow F$	$E+T$
		$E \rightarrow E+T$	$E$

String accepted

# Построение таблиц SLR-анализа

3FLAP : (SLR1-2.jff)

File Input Test Convert Help

Editor Build SLR(1) Parse

Do Selected Do Step Do All Next Parse

Parse table complete. Press "parse" to use it.

	FIRST	FOLLOW
L	{*, i}	{\$, =}
R	{*, i}	{\$, =}
S	{*, i}	{\$}

$S' \rightarrow S$   
 $S \rightarrow L=R$   
 $S \rightarrow R$   
 $L \rightarrow *R$   
 $L \rightarrow i$   
 $R \rightarrow L$

Diagram illustrating the SLR(1) item sets and transitions:

- q0**:  $L \rightarrow *R$ ,  $S \rightarrow L=R$ ,  $S' \rightarrow S$ ,  $S \rightarrow R$ ,  $R \rightarrow L$ ,  $L \rightarrow i$
- q1**:  $L \rightarrow *R$ ,  $R \rightarrow L$ ,  $L \rightarrow *R$ ,  $L \rightarrow i$
- q2**:  $R \rightarrow L$ ,  $S \rightarrow L=R$
- q3**:  $S \rightarrow R$
- q4**:  $S' \rightarrow S$
- q5**:  $L \rightarrow i$
- q6**:  $R \rightarrow L$
- q7**:  $L \rightarrow *R$
- q8**:  $S \rightarrow L=R$ ,  $L \rightarrow *R$ ,  $R \rightarrow L$ ,  $L \rightarrow i$
- q9**:  $S \rightarrow L=R$

	*	=	i	\$	L	R	S
0	s1		s5		2	3	4
1	s1		s5		6	7	
2		r5		r5			
3				r2			
4				acc			
5		r4		r4			
6		r5		r5			
7		r3		r3			
8	s1		s5		6	9	

# Активные префиксы

- Содержимое стека должно быть префиксом правосентенциальной формы
- Если в стеке хранится  $a$ , а оставшаяся часть входной строки –  $x$ , то последовательность сверток должна привести  $ax$  в  $S$ 
  - В терминах порождений  $S \Rightarrow^* ax$
- Предположим, например,  
$$E \Rightarrow^* F * i \Rightarrow (E) * i$$
- В процессе СА в стеке хранятся  $\{ (, (E$  и  $(E) \}$ , но в нем не должно быть  $(E)^*$ , т.к.  $(E)$  – основа, которую анализатор должен свернуть в  $F$  до переноса  $*$



# Активные префиксы

- Префиксы правосентенциальных форм, которые могут находиться в стеке ПС-анализатора - **активные префиксы**
- Активный префикс является префиксом правосентенциальной формы, не выходящим за пределы правого конца крайней справа основы СФ
- К концу активного префикса всегда можно добавить терминальные символы для получения правосентенциальной формы
- SLR-анализ основан на том факте, что LR(0)-автомат распознает активные префиксы

# Активные префиксы

- Мы говорим, что ситуация  $A \rightarrow \beta_1 \bullet \beta_2$  **допустима** для активного префикса  $a\beta_1$ , если существует правое порождение  $S' \Rightarrow^* aAw \Rightarrow a\beta_1\beta_2w$
- Тот факт, что  $A \rightarrow \beta_1 \bullet \beta_2$  допустима для  $a\beta_1$ , многое говорит о том, что следует выбрать – перенос или свертку – при обнаружении  $a\beta_1$  на вершине стека
- Если  $\beta_2 \neq \varepsilon$ , то основа еще не полностью перенесена в стек и очередное действие анализатора – перенос
- Если  $\beta_2 = \varepsilon$ , то  $A \rightarrow \beta_1$  – основа, и анализатор должен выполнить свертку

# Активные префиксы

$$\begin{array}{lll}
 E' & \xRightarrow{rm} & E \\
 & \xRightarrow{rm} & E + T \\
 & \xRightarrow{rm} & E + T * F \\
 \\
 E' & \xRightarrow{rm} & E \\
 & \xRightarrow{rm} & E + T \\
 & \xRightarrow{rm} & E + T * F \\
 & \xRightarrow{rm} & E + T * (E) \\
 \\
 E' & \xRightarrow{rm} & E \\
 & \xRightarrow{rm} & E + T \\
 & \xRightarrow{rm} & E + T * F \\
 & \xRightarrow{rm} & E + T * id
 \end{array}$$

- НКА  $N$  для распознавания активных префиксов может быть построен путем рассмотрения ситуаций в качестве состояний
- Существует переход из  $A \rightarrow a \bullet X\beta$  в  $A \rightarrow aX \bullet \beta$ , помеченный  $X$ , и переход из  $A \rightarrow a \bullet B\beta$  в  $B \rightarrow \bullet \gamma$ , помеченный  $\varepsilon$

# Активные префиксы

- $CLOSURE(I)$  для множества ситуаций (или состояний  $N$ )  $I$  в точности представляет собой  $\varepsilon$ -замыкание множества состояний НКА
- Таким образом,  $GOTO(I, X)$  дает переход из  $I$  для символа  $X$  в ДКА, построенном из  $N$  при помощи метода конструирования подмножеств

# Дополнительные источники

- LR(0) - [http://ru.wikipedia.org/wiki/LR\(0\)](http://ru.wikipedia.org/wiki/LR(0))
- SLR(1) - [http://ru.wikipedia.org/wiki/SLR\(1\)](http://ru.wikipedia.org/wiki/SLR(1))
- Ахо, А. Компиляторы: принципы, технологии и инструментарий, 2 издание / А. Ахо, М.Лам, Р. Сети, Дж. Ульман. – М.: Издательский дом «Вильямс», 2008. – 1184 с.
- LR-анализатор - <http://ru.wikipedia.org/wiki/LR>
- The Lex & Yacc Page - <http://dinosaur.compilertools.net/>