Syntax Analysis
Part II

Chapter 4

Bottom-Up Parsing

- LR methods (Left-to-right, Rightmost derivation)
  - SLR, Canonical LR, LALR
- Other special cases:
  - Shift-reduce parsing
  - Operator-precedence parsing

Operator-Precedence Parsing

- Special case of shift-reduce parsing
- We will not further discuss (you can skip textbook section 4.6)

Shift-Reduce Parsing

Grammar:

$S \rightarrow \mathbf{a}\ A\ B\ \mathbf{e}$

$A \rightarrow A\ \mathbf{b}\ \mathbf{c} \mid \mathbf{b}$

$B \rightarrow \mathbf{d}$

Reducing a sentence:

$\mathbf{a}\ \underline{\mathbf{b}}\ \mathbf{b}\ \mathbf{c}\ \mathbf{d}\ \mathbf{e}$

$\mathbf{a}\ \underline{A\ \mathbf{b}\ \mathbf{c}}\ \mathbf{d}\ \mathbf{e}$

$\mathbf{a}\ A\ \underline{\mathbf{d}}\ \mathbf{e}$

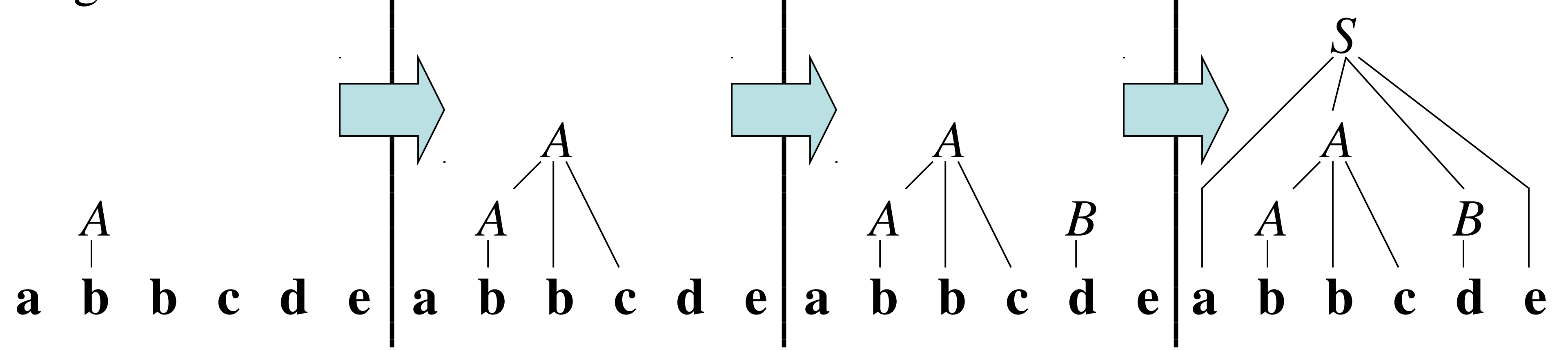
$\underline{\mathbf{a}\ A\ B\ \mathbf{e}}$

S

These match production's right-hand sides

Shift-reduce corresponds to a rightmost derivation:

$$
\begin{aligned}
S &\Rightarrow_{rm} \mathbf{a}\ A\ B\ \mathbf{e} \\
&\Rightarrow_{rm} \mathbf{a}\ A\ \mathbf{d}\ \mathbf{e} \\
&\Rightarrow_{rm} \mathbf{a}\ A\ \mathbf{b}\ \mathbf{c}\ \mathbf{d}\ \mathbf{e} \\
&\Rightarrow_{rm} \mathbf{a}\ \mathbf{b}\ \mathbf{b}\ \mathbf{c}\ \mathbf{d}\ \mathbf{e}
\end{aligned}
$$



Handles

A *handle* is a substring of grammar symbols in a *right-sentential form* that matches a right-hand side of a production

Grammar:

$S \rightarrow \mathbf{a}\ A\ B\ \mathbf{e}$

$A \rightarrow A\ \mathbf{b}\ \mathbf{c} \mid \mathbf{b}$

$B \rightarrow \mathbf{d}$

$\mathbf{a}\ \underline{\mathbf{b}}\ \mathbf{b}\ \mathbf{c}\ \mathbf{d}\ \mathbf{e}$

$\mathbf{a}\ \underline{A\ \mathbf{b}\ \mathbf{c}}\ \mathbf{d}\ \mathbf{e}$

$\mathbf{a}\ A\ \underline{\mathbf{d}}\ \mathbf{e}$

$\underline{\mathbf{a}\ A\ B\ \mathbf{e}}$

S

Handle

$\mathbf{a}\ \underline{\mathbf{b}}\ \mathbf{b}\ \mathbf{c}\ \mathbf{d}\ \mathbf{e}$

$\mathbf{a}\ A\ \underline{\mathbf{b}\ \mathbf{c}}\ \mathbf{d}\ \mathbf{e}$

$\mathbf{a}\ A\ A\ \mathbf{e}$

… ?

NOT a handle, because further reductions will fail (result is not a sentential form)

Stack Implementation of Shift-Reduce Parsing

Grammar:

$E \rightarrow E \mathbf{+} E$

$E \rightarrow E \mathbf{*} E$

$E \rightarrow \mathbf{(}\ E\ \mathbf{)}$

$E \rightarrow \mathbf{id}$

Stack	Input	Action
$\$$	**id+id*id$**	shift
$\$\underline{\mathbf{id}}$	**+id*id$**	reduce $E \rightarrow \mathbf{id}$
$\$E$	**+id*id$**	shift
$\$E\mathbf{+}$	**id*id$**	shift
$\$E\mathbf{+}\underline{\mathbf{id}}$	***id$**	reduce $E \rightarrow \mathbf{id}$
$\$E\mathbf{+}E$	***id$**	shift (or reduce?)
$\$E\mathbf{+}E\mathbf{*}$	**id$**	shift
$\$E\mathbf{+}E\mathbf{*}\underline{\mathbf{id}}$	**$**	reduce $E \rightarrow \mathbf{id}$
$\$E\mathbf{+}\underline{E\mathbf{*}E}$	**$**	reduce $E \rightarrow E \mathbf{*} E$
$\$\underline{E\mathbf{+}E}$	**$**	reduce $E \rightarrow E \mathbf{+} E$
$\$E$	**$**	accept

How to resolve conflicts?

Find handles to reduce

Conflicts

- *Shift-reduce* and *reduce-reduce* conflicts are caused by
  - The limitations of the LR parsing method (even when the grammar is unambiguous)
  - Ambiguity of the grammar

Shift-Reduce Parsing: Shift-Reduce Conflicts

Stack	Input	Action
\$…	…**\$**	…
\$…**if** E **then** S	**else**…**\$**	shift or reduce?

Ambiguous grammar:
$S \rightarrow$ **if** E **then** S
    | **if** E **then** S **else** S
    | **other**

Resolve in favor of shift, so **else** matches closest **if**

Shift-Reduce Parsing: Reduce-Reduce Conflicts

Stack	Input	Action
\$	**aa\$**	shift
\$<u>**a**</u>	**a\$**	reduce $A \to$ **a** <u>or</u> $B \to$ **a** ?

Grammar:

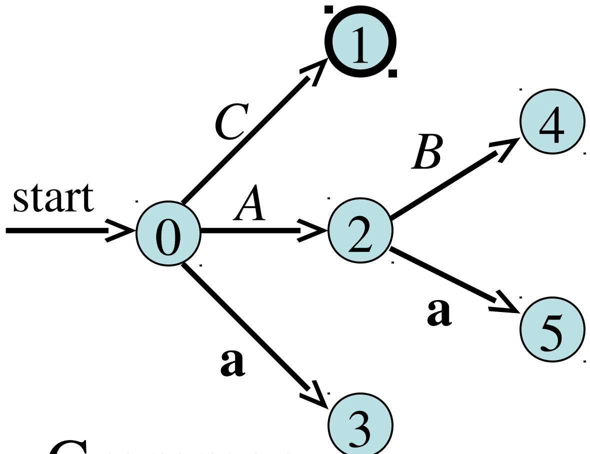
$C \to A\ B$

$A \to$ **a**

$B \to$ **a**

Resolve in favor of reduce $A \to$ **a**, otherwise we're stuck!

LR(k) Parsers: Use a DFA for Shift/Reduce Decisions



Grammar:

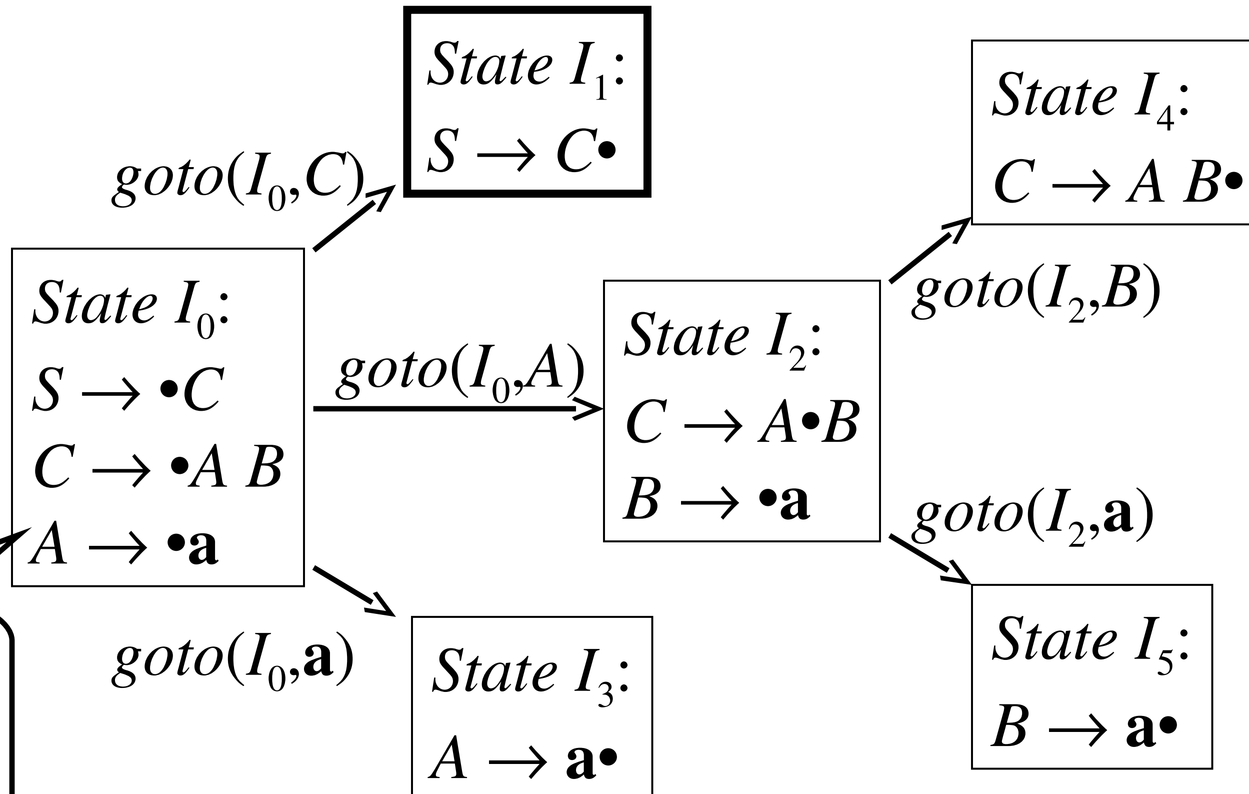
$S \rightarrow C$

$C \rightarrow A\ B$

$A \rightarrow \mathbf{a}$

$B \rightarrow \mathbf{a}$

Can only reduce $A \rightarrow \mathbf{a}$ (not $B \rightarrow \mathbf{a}$)



DFA for Shift/Reduce Decisions

The states of the DFA are used to determine if a handle is on top of the stack

Grammar:
$S \rightarrow C$
$C \rightarrow A\ B$
$A \rightarrow \mathbf{a}$
$B \rightarrow \mathbf{a}$

State I_0:
$S \rightarrow \bullet C$
$C \rightarrow \bullet A\ B$
$A \rightarrow \bullet \mathbf{a}$

$goto(I_0, \mathbf{a})$ $\longrightarrow$

State I_3:
$A \rightarrow \mathbf{a}\bullet$

Stack	Input	Action
\$ 0	**aa\$**	start in state 0
\$ $\underline{0}$	$\underline{\mathbf{a}}\mathbf{a\$}$	shift (and goto state 3)
\$ 0 **a** 3	**a\$**	reduce $A \rightarrow \mathbf{a}$ (goto 2)
\$ 0 A 2	**a\$**	shift (goto 5)
\$ 0 A 2 **a** 5	**\$**	reduce $B \rightarrow \mathbf{a}$ (goto 4)
\$ 0 A 2 B 4	**\$**	reduce $C \rightarrow AB$ (goto 1)
\$ 0 C 1	**\$**	accept ($S \rightarrow C$)

DFA for Shift/Reduce Decisions

The states of the DFA are used to determine if a handle is on top of the stack

Grammar:

$S \to C$

$C \to A\ B$

$A \to \mathbf{a}$

$B \to \mathbf{a}$

State I_0:

$S \to \bullet C$

$C \to \bullet A\ B$

$A \to \bullet\mathbf{a}$

$goto(I_0,A)$ →

State I_2:

$C \to A \bullet B$

$B \to \bullet\mathbf{a}$

Stack	Input	Action
$ 0	**aa$**	start in state 0
$ 0	**aa$**	shift (and goto state 3)
$ $\underline{0}$ $\underline{\mathbf{a}}$ 3	**a$**	reduce $A \to \mathbf{a}$ (goto 2)
$ 0 A 2	**a$**	shift (goto 5)
$ 0 A 2 **a** 5	**$**	reduce $B \to \mathbf{a}$ (goto 4)
$ 0 A 2 B 4	**$**	reduce $C \to AB$ (goto 1)
$ 0 C 1	**$**	accept ($S \to C$)

DFA for Shift/Reduce Decisions

The states of the DFA are used to determine if a handle is on top of the stack

Grammar:

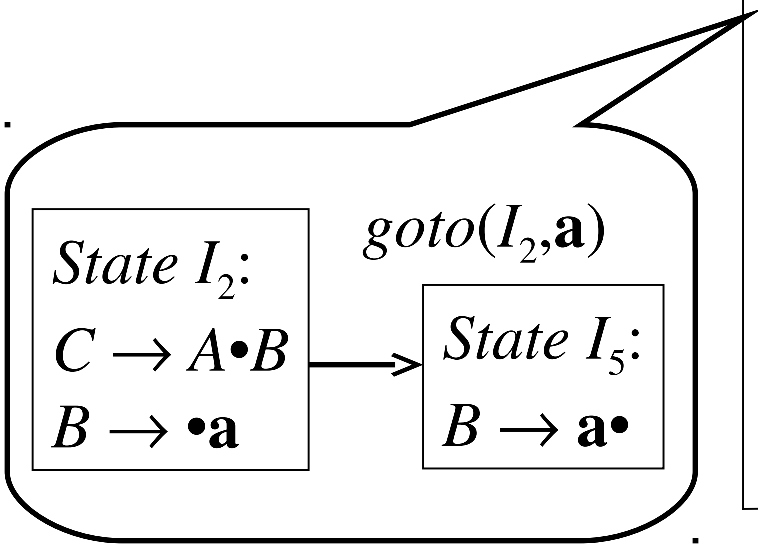
$S \rightarrow C$

$C \rightarrow A\ B$

$A \rightarrow \mathbf{a}$

$B \rightarrow \mathbf{a}$

Stack	Input	Action
\$ 0	**aa\$**	start in state 0
\$ 0	**aa\$**	shift (and goto state 3)
\$ 0 **a** 3	**a\$**	reduce $A \rightarrow \mathbf{a}$ (goto 2)
\$ 0 A $\underline{2}$	$\underline{\mathbf{a}}\mathbf{\$}$	shift (goto 5)
\$ 0 A 2 **a** 5	**\$**	reduce $B \rightarrow \mathbf{a}$ (goto 4)
\$ 0 A 2 B 4	**\$**	reduce $C \rightarrow AB$ (goto 1)
\$ 0 C 1	**\$**	accept ($S \rightarrow C$)



DFA for Shift/Reduce Decisions

The states of the DFA are used to determine if a handle is on top of the stack

Grammar:

$S \rightarrow C$

$C \rightarrow A\ B$

$A \rightarrow \mathbf{a}$

$B \rightarrow \mathbf{a}$

Stack	Input	Action
$\$\ 0$	**aa$**	start in state 0
$\$\ 0$	**aa$**	shift (and goto state 3)
$\$\ 0\ \mathbf{a}\ 3$	**a$**	reduce $A \rightarrow \mathbf{a}$ (goto 2)
$\$\ 0\ A\ 2$	**a$**	shift (goto 5)
$\$\ 0\ A\ \underline{2}\ \underline{\mathbf{a}}\ 5$	**$**	reduce $B \rightarrow \mathbf{a}$ (goto 4)
$\$\ 0\ A\ 2\ B\ 4$	**$**	reduce $C \rightarrow AB$ (goto 1)
$\$\ 0\ C\ 1$	**$**	accept ($S \rightarrow C$)

State I_2:

$C \rightarrow A \bullet B$

$B \rightarrow \bullet \mathbf{a}$

$\xrightarrow{goto(I_2,B)}$

State I_4:

$C \rightarrow A\ B \bullet$

DFA for Shift/Reduce Decisions

The states of the DFA are used to determine if a handle is on top of the stack

Grammar:

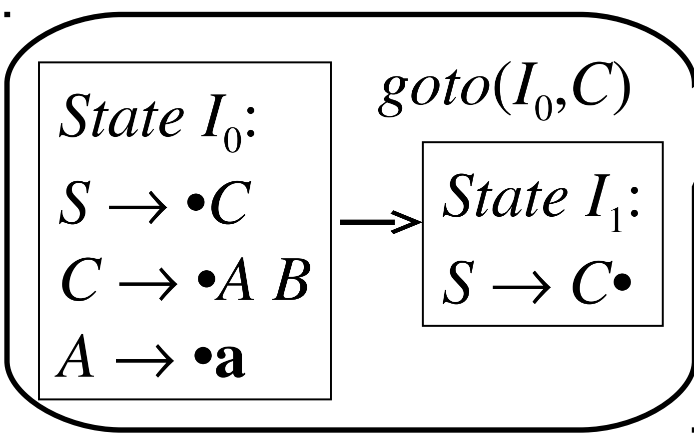
$S \to C$

$C \to A\ B$

$A \to \mathbf{a}$

$B \to \mathbf{a}$

Stack	Input	Action
$ 0	**aa$**	start in state 0
$ 0	**aa$**	shift (and goto state 3)
$ 0 **a** 3	**a$**	reduce $A \to \mathbf{a}$ (goto 2)
$ 0 A 2	**a$**	shift (goto 5)
$ 0 A 2 **a** 5	**$**	reduce $B \to \mathbf{a}$ (goto 4)
$ $\underline{0}$ $\underline{A}$ 2 $\underline{B}$ 4	**$**	reduce $C \to AB$ (goto 1)
$ 0 C 1	**$**	accept ($S \to C$)



DFA for Shift/Reduce Decisions

The states of the DFA are used to determine if a handle is on top of the stack

Grammar:

$S \rightarrow C$

$C \rightarrow A\ B$

$A \rightarrow \mathbf{a}$

$B \rightarrow \mathbf{a}$

Stack	Input	Action
$\$\ 0$	**aa$**	start in state 0
$\$\ 0$	**aa$**	shift (and goto state 3)
$\$\ 0\ \mathbf{a}\ 3$	**a$**	reduce $A \rightarrow \mathbf{a}$ (goto 2)
$\$\ 0\ A\ 2$	**a$**	shift (goto 5)
$\$\ 0\ A\ 2\ \mathbf{a}\ 5$	**$**	reduce $B \rightarrow \mathbf{a}$ (goto 4)
$\$\ 0\ A\ 2\ B\ 4$	**$**	reduce $C \rightarrow AB$ (goto 1)
$\$\ 0\ C\ \underline{1}$	$\underline{\$}$	accept ($S \rightarrow C$)

State I_0:

$S \rightarrow \bullet C$

$C \rightarrow \bullet A\ B$

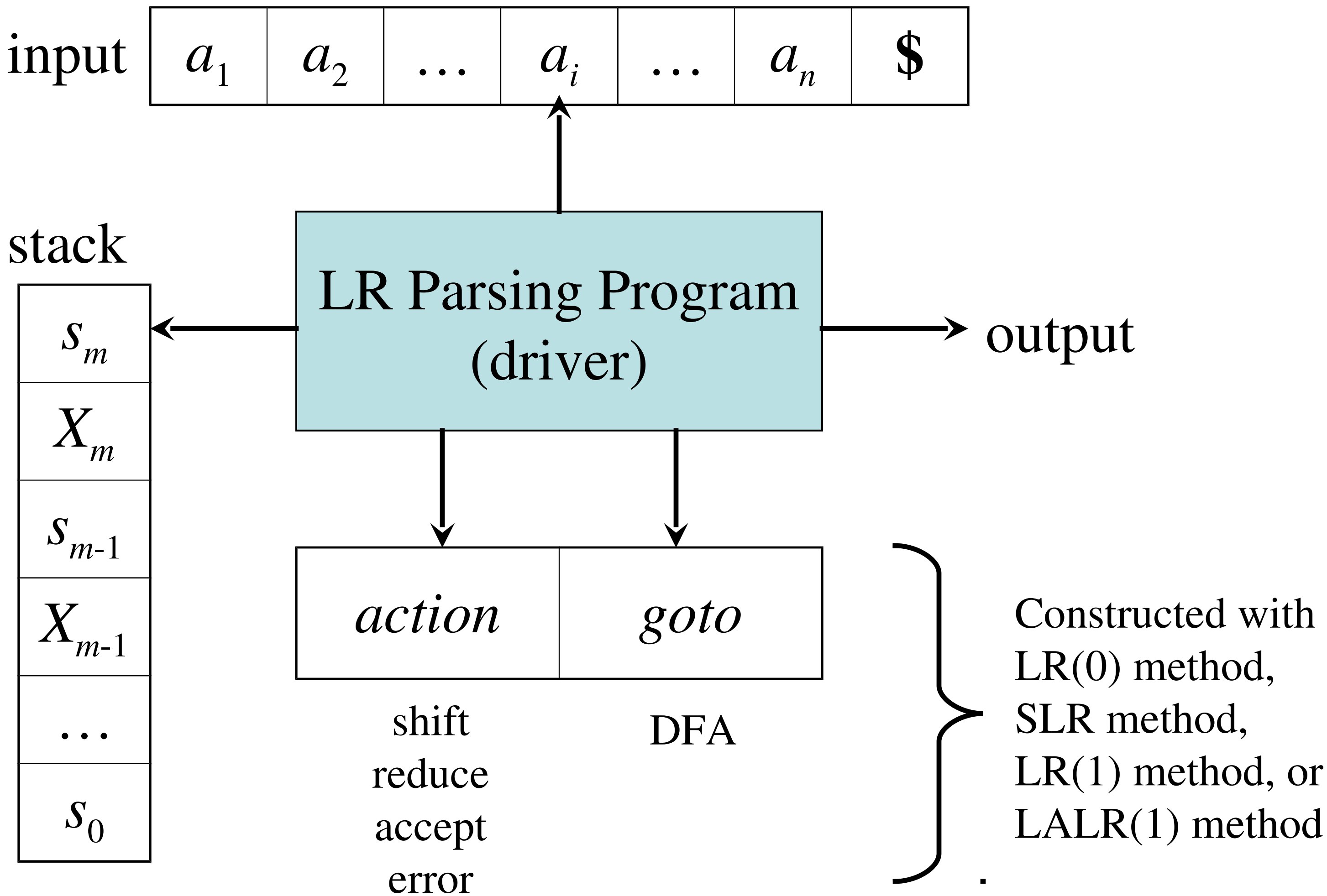
$A \rightarrow \bullet\mathbf{a}$

$\xrightarrow{goto(I_0,C)}$

State I_1:

$S \rightarrow C\bullet$

Model of an LR Parser



LR Parsing (Driver)

Configuration (= LR parser state):

$$\underbrace{(s_0\ X_1\ s_1\ X_2\ s_2\ \ldots\ X_m\ s_m,}_{stack}\quad \underbrace{a_i\ a_{i+1}\ \ldots\ a_n\ \$)}_{input}$$

If *action*$[s_m,a_i]$ = shift s **then** push a_i, push s, and advance input:

$$(s_0\ X_1\ s_1\ X_2\ s_2\ \ldots\ X_m\ s_m\ a_i\ s,\quad a_{i+1}\ \ldots\ a_n\ \$)$$

If *action*$[s_m,a_i]$ = reduce A $\rightarrow$ β and *goto*$[s_{m-r},A] = s$ with r=|β| **then** pop $2r$ symbols, push A, and push s:

$$(s_0\ X_1\ s_1\ X_2\ s_2\ \ldots\ X_{m-r}\ s_{m-r}\ A\ s,\quad a_i\ a_{i+1}\ \ldots\ a_n\ \$)$$

If *action*$[s_m,a_i]$ = accept **then** stop

If *action*$[s_m,a_i]$ = error **then** attempt recovery

Example LR Parse Table

Grammar:

1. $E \rightarrow E + T$
2. $E \rightarrow T$
3. $T \rightarrow T * F$
4. $T \rightarrow F$
5. $F \rightarrow (E)$
6. $F \rightarrow$ **id**

	action						*goto*		
state	**id**	**+**	*****	**(**	**)**	**$**	*E*	*T*	*F*
0	s5			s4			1	2	3
1		s6				acc			
2		r2	s7		r2	r2			
3		r4	r4		r4	r4			
4	s5			s4			8	2	3
5		r6	r6		r6	r6			
6	s5			s4				9	3
7	s5			s4					10
8		s6			s11				
9		r1	s7		r1	r1			
10		r3	r3		r3	r3			
11		r5	r5		r5	r5			

Shift & goto 5

Reduce by production #1

Example LR Parsing

Grammar:

1. $E \rightarrow E + T$
2. $E \rightarrow T$
3. $T \rightarrow T * F$
4. $T \rightarrow F$
5. $F \rightarrow (E)$
6. $F \rightarrow \textbf{id}$

Stack	Input	Action
$ 0	**id*id+id$**	shift 5
$ 0 **id** 5	***id+id$**	reduce 6 goto 3
$ 0 *F* 3	***id+id$**	reduce 4 goto 2
$ 0 *T* 2	***id+id$**	shift 7
$ 0 *T* 2 * 7	**id+id$**	shift 5
$ 0 *T* 2 * 7 **id** 5	**+id$**	reduce 6 goto 10
$ 0 *T* 2 * 7 *F* 10	**+id$**	reduce 3 goto 2
$ 0 *T* 2	**+id$**	reduce 2 goto 1
$ 0 *E* 1	**+id$**	shift 6
$ 0 *E* 1 + 6	**id$**	shift 5
$ 0 *E* 1 + 6 **id** 5	**$**	reduce 6 goto 3
$ 0 *E* 1 + 6 *F* 3	**$**	reduce 4 goto 9
$ 0 *E* 1 + 6 *T* 9	**$**	reduce 1 goto 1
$ 0 *E* 1	**$**	accept

SLR Grammars

- SLR (Simple LR): a simple extension of LR(0) shift-reduce parsing
- SLR eliminates some conflicts by populating the parsing table with reductions $A\rightarrow\alpha$ on symbols in FOLLOW(A)

$S \rightarrow E$
$E \rightarrow \mathbf{id} + E$
$E \rightarrow \mathbf{id}$

State I_0:
$S \rightarrow \bullet E$
$E \rightarrow \bullet\mathbf{id} + E$
$E \rightarrow \bullet\mathbf{id}$

$\xrightarrow{goto(I_0,\mathbf{id})}$

State I_2:
$E \rightarrow \mathbf{id}\bullet + E$
$E \rightarrow \mathbf{id}\bullet$

$\xrightarrow{goto(I_3,+)}$

Shift on **+**

FOLLOW(E)={**$**}
thus reduce on **$**

SLR Parsing Table

- Reductions do not fill entire rows
- Otherwise the same as LR(0)

1. $S \rightarrow E$
2. $E \rightarrow$ **id +** E
3. $E \rightarrow$ **id**

	id	**+**	**$**	*E*
0	s2			1
1			acc	
2		s3	r3	
3	s2			4
4			r2	

Shift on **+**

FOLLOW(*E*)={**$**}
thus reduce on **$**

SLR Parsing

- An LR(0) state is a set of LR(0) items
- An LR(0) item is a production with a • (dot) in the right-hand side
- Build the LR(0) DFA by
  - *Closure operation* to construct LR(0) items
  - *Goto operation* to determine transitions
- Construct the SLR parsing table from the DFA
- LR parser program uses the SLR parsing table to determine shift/reduce operations

Constructing SLR Parsing Tables

1. Augment the grammar with $S' \to S$
2. Construct the set $C=\{I_0,I_1,\ldots,I_n\}$ of $LR(0)$ *items*
3. If $[A\to\alpha\bullet a\beta] \in I_i$ and $goto(I_i,a)=I_j$ then set $action[i,a]$=shift j
4. If $[A\to\alpha\bullet] \in I_i$ then set $action[i,a]$=reduce $\mathrm{A}\to\alpha$ for all $a \in$ FOLLOW(A) (apply only if $A\neq S'$)
5. If $[S'\to S\bullet]$ is in I_i then set $action[i,\$]$=accept
6. If $goto(I_i,A)=I_j$ then set $goto[i,A]=j$
7. Repeat 3-6 until no more entries added
8. The initial state i is the I_i holding item $[S'\to\bullet S]$

LR(0) Items of a Grammar

- An *LR*(0) *item* of a grammar G is a production of G with a • at some position of the right-hand side
- Thus, a production

  $A \rightarrow X\,Y\,Z$

  has four items:

  $[A \rightarrow \bullet\, X\,Y\,Z]$
  $[A \rightarrow X \bullet Y\,Z]$
  $[A \rightarrow X\,Y \bullet Z]$
  $[A \rightarrow X\,Y\,Z\, \bullet]$
- Note that production $A \rightarrow \varepsilon$ has one item $[A \rightarrow \bullet]$

Constructing the set of LR(0) Items of a Grammar

1. The grammar is augmented with a new start symbol S' and production $S' \rightarrow S$
2. Initially, set $C = closure(\{[S' \rightarrow \bullet S]\})$ (this is the start state of the DFA)
3. For each set of items $I \in C$ and each grammar symbol $X \in (N \cup T)$ such that $goto(I,X) \notin C$ and $goto(I,X) \neq \emptyset$, add the set of items $goto(I,X)$ to C
4. Repeat 3 until no more sets can be added to C

The Closure Operation for LR(0) Items

1. Start with $closure(I) = I$
2. If $[A \rightarrow \alpha \bullet B \beta] \in closure(I)$ then for each production $B \rightarrow \gamma$ in the grammar, add the item $[B \rightarrow \bullet \gamma]$ to I if not already in I
3. Repeat 2 until no new items can be added

The Closure Operation (Example)

$closure(\{[E' \to \bullet E]\}) =$

$\{ [E' \to \bullet E] \}$ ⇒ $\{ [E' \to \bullet E]$, $[E \to \bullet E + T]$, $[E \to \bullet T] \}$ ⇒ $\{ [E' \to \bullet E]$, $[E \to \bullet E + T]$, $[E \to \bullet T]$, $[T \to \bullet T * F]$, $[T \to \bullet F] \}$ ⇒ $\{ [E' \to \bullet E]$, $[E \to \bullet E + T]$, $[E \to \bullet T]$, $[T \to \bullet T * F]$, $[T \to \bullet F]$, $[F \to \bullet (E)]$, $[F \to \bullet \mathbf{id}] \}$

Add $[E\to\bullet\gamma]$

Add $[T\to\bullet\gamma]$

Add $[F\to\bullet\gamma]$

Grammar:

$E \to E + T \mid T$

$T \to T * F \mid F$

$F \to (E)$

$F \to \mathbf{id}$

The Goto Operation for LR(0) Items

1. For each item $[A \to \alpha \bullet X \beta] \in I$, add the set of items $closure(\{[A \to \alpha X \bullet \beta]\})$ to $goto(I,X)$ if not already there
2. Repeat step 1 until no more items can be added to $goto(I,X)$
3. Intuitively, $goto(I,X)$ is the set of items that are valid for the viable prefix γX when I is the set of items that are valid for γ

The Goto Operation (Example 1)

Suppose $I = \{\ [E' \to \bullet\ E]$
$[E \to \bullet\ E + T]$
$[E \to \bullet\ T]$
$[T \to \bullet\ T * F]$
$[T \to \bullet\ F]$
$[F \to \bullet\ (\ E\ )]$
$[F \to \bullet\ \mathbf{id}]\ \}$

Then $goto(I,E)$
$= closure(\{[E' \to E\ \bullet,\ E \to E\ \bullet + T]\})$
$= \{\ [E' \to E\ \bullet]$
$[E \to E\ \bullet + T]\ \}$

Grammar:
$E \to E + T \mid T$
$T \to T * F \mid F$
$F \to (\ E\ )$
$F \to \mathbf{id}$

The Goto Operation (Example 2)

Suppose $I = \{\ [E' \to E\ \bullet],\ [E \to E \bullet \mathbf{+}\ T]\ \}$

Then $goto(I,\mathbf{+}) = closure(\{[E \to E\ \mathbf{+} \bullet T]\}) = \{\ [E \to E\ \mathbf{+} \bullet T]$
$[T \to \bullet\ T\ \mathbf{*}\ F]$
$[T \to \bullet\ F]$
$[F \to \bullet\ \mathbf{(}\ E\ \mathbf{)}]$
$[F \to \bullet\ \mathbf{id}]\ \}$

Grammar:
$E \to E\ \mathbf{+}\ T \mid T$
$T \to T\ \mathbf{*}\ F \mid F$
$F \to \mathbf{(}\ E\ \mathbf{)}$
$F \to \mathbf{id}$

Example SLR Grammar and LR(0) Items

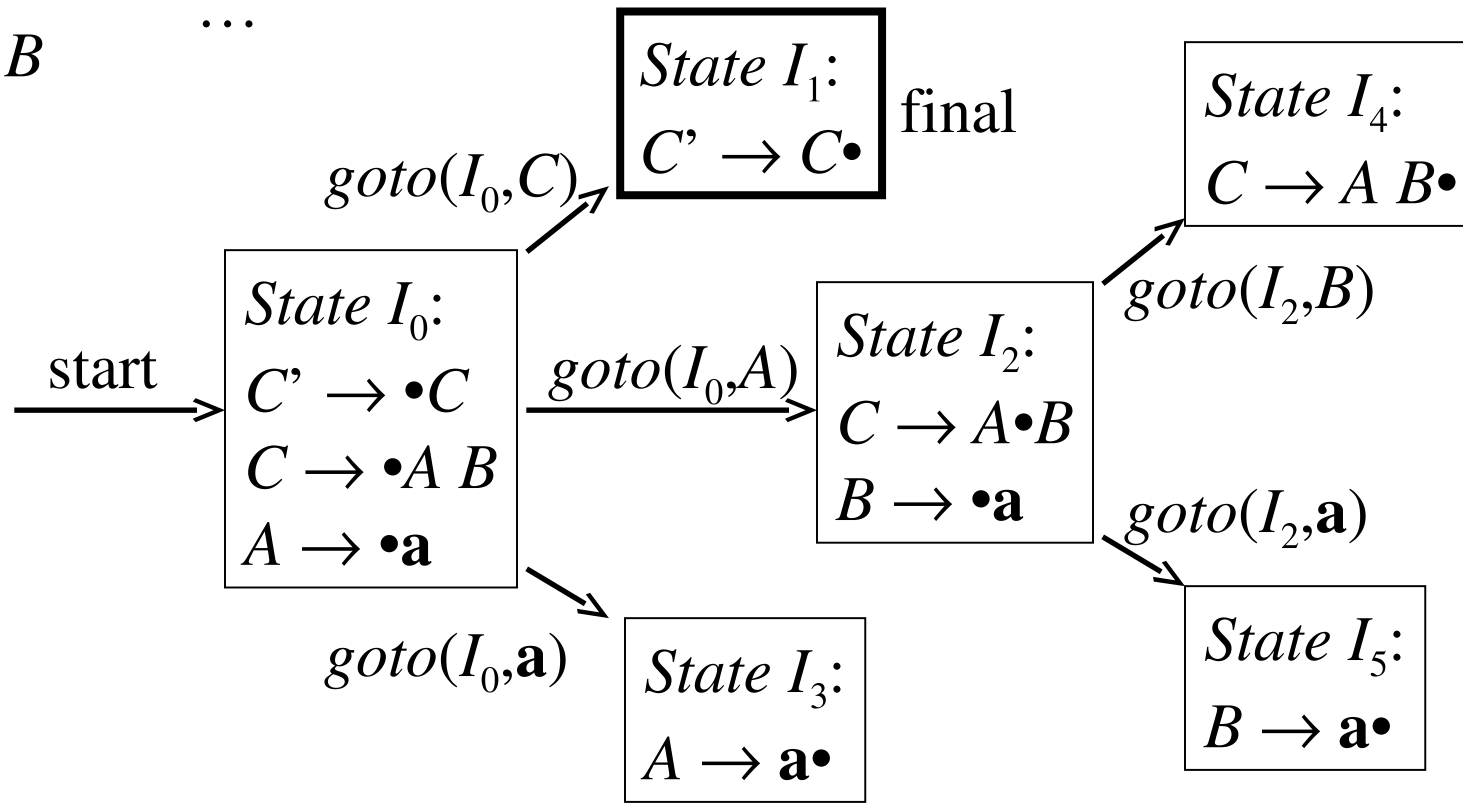
Augmented grammar:

1. $C' \rightarrow C$
2. $C \rightarrow A\ B$
3. $A \rightarrow \mathbf{a}$
4. $B \rightarrow \mathbf{a}$

$I_0 = closure(\{[C' \rightarrow \bullet C]\})$

$I_1 = \text{goto}(I_0, C) = closure(\{[C' \rightarrow C\bullet]\})$

…



Example SLR Parsing Table

State I_0:
$C' \to \bullet C$
$C \to \bullet A\ B$
$A \to \bullet \mathbf{a}$

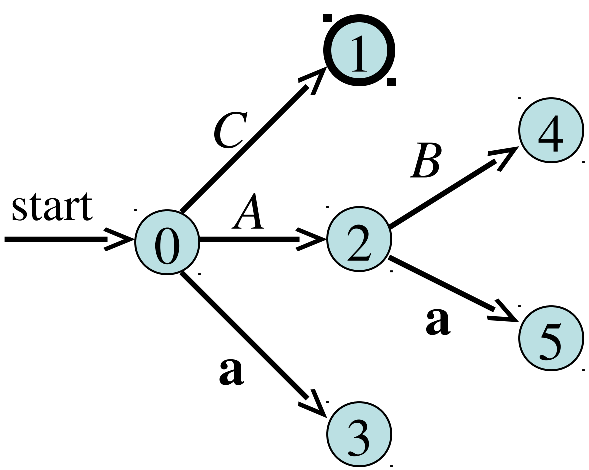
State I_1:
$C' \to C\bullet$

State I_2:
$C \to A \bullet B$
$B \to \bullet \mathbf{a}$

State I_3:
$A \to \mathbf{a}\bullet$

State I_4:
$C \to A\ B\bullet$

State I_5:
$B \to \mathbf{a}\bullet$



	a	**$**	*C*	*A*	*B*
0	s3		1	2	
1		acc			
2	s5				4
3	r3				
4		r2			
5		r4			

Grammar:
1. $C' \to C$
2. $C \to A\ B$
3. $A \to \mathbf{a}$
4. $B \to \mathbf{a}$

SLR and Ambiguity

- Every SLR grammar is unambiguous, but **not** every unambiguous grammar is SLR
- Consider for example the unambiguous grammar

$S \rightarrow L = R \mid R$
$L \rightarrow * R \mid \mathbf{id}$
$R \rightarrow L$

I_0:	I_1:	I_2:	I_3:	I_4:	I_5:	I_6:	I_7:
$S' \rightarrow \bullet S$	$S' \rightarrow S\bullet$	$S \rightarrow L\bullet{=}R$	$S \rightarrow R\bullet$	$L \rightarrow *\bullet R$	$L \rightarrow \mathbf{id}\bullet$	$S \rightarrow L{=}\bullet R$	$L \rightarrow *R\bullet$
$S \rightarrow \bullet L{=}R$		$R \rightarrow L\bullet$		$R \rightarrow \bullet L$		$R \rightarrow \bullet L$	
$S \rightarrow \bullet R$				$L \rightarrow \bullet *R$		$L \rightarrow \bullet *R$	
$L \rightarrow \bullet *R$				$L \rightarrow \bullet \mathbf{id}$		$L \rightarrow \bullet \mathbf{id}$	
$L \rightarrow \bullet \mathbf{id}$							
$R \rightarrow \bullet L$							

I_8:
$R \rightarrow L\bullet$

I_9:
$S \rightarrow L{=}R\bullet$

action[2,**=**]=s6
action[2,**=**]=r5

no

Has no SLR parsing table

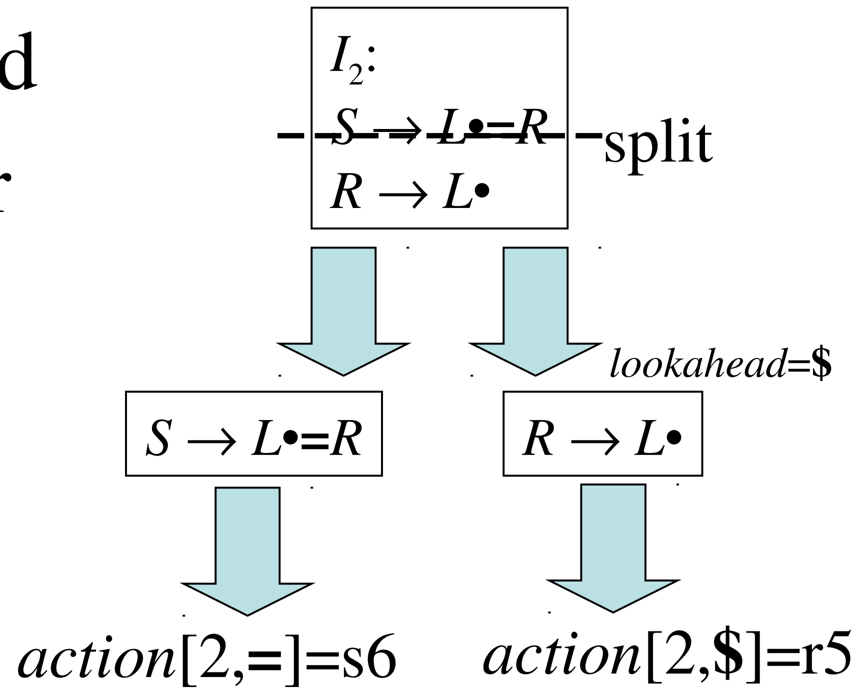
LR(1) Grammars

- SLR too simple
- LR(1) parsing uses lookahead to avoid unnecessary conflicts in parsing table
- LR(1) item = LR(0) item + lookahead

LR(0) item:	LR(1) item:
$[A \to \alpha \bullet \beta]$	$[A \to \alpha \bullet \beta,\ a]$

SLR Versus LR(1)

- Split the SLR states by adding LR(1) lookahead
- Unambiguous grammar
  1. $S \rightarrow L = R$
  2. $\quad\quad | \; R$
  3. $L \rightarrow * \; R$
  4. $\quad\quad | \; \mathbf{id}$
  5. $R \rightarrow L$



Should not reduce on **=**, because no right-sentential form begins with *R***=**

LR(1) Items

- An *LR*(1) *item*

  $[A \to \alpha \bullet \beta, a]$

  contains a *lookahead* terminal a, meaning α already on top of the stack, expect to see βa
- For items of the form

  $[A \to \alpha \bullet, a]$

  the lookahead a is used to reduce $A \to \alpha$ only if the next input is a
- For items of the form

  $[A \to \alpha \bullet \beta, a]$

  with $\beta \neq \varepsilon$ the lookahead has no effect

The Closure Operation for LR(1) Items

1. Start with *closure*(I) = I
2. If $[A \to \alpha \bullet B\beta, a] \in$ *closure*(I) then for each production $B \to \gamma$ in the grammar and each terminal $b \in$ FIRST(βa), add the item $[B \to \bullet\gamma, b]$ to I if not already in I
3. Repeat 2 until no new items can be added

The Goto Operation for LR(1) Items

1. For each item $[A\rightarrow\alpha\bullet X\beta,\ a] \in I$, add the set of items *closure*$(\{[A\rightarrow\alpha X\bullet\beta,\ a]\})$ to *goto*(I,X) if not already there
2. Repeat step 1 until no more items can be added to *goto*(I,X)

Constructing the set of LR(1) Items of a Grammar

1. Augment the grammar with a new start symbol S' and production $S' \to S$
2. Initially, set $C = closure(\{[S' \to \bullet S, \mathbf{\$}]\})$ (this is the start state of the DFA)
3. For each set of items $I \in C$ and each grammar symbol $X \in (N \cup T)$ such that $goto(I,X) \notin C$ and $goto(I,X) \neq \emptyset$, add the set of items $goto(I,X)$ to C
4. Repeat 3 until no more sets can be added to C

Example Grammar and LR(1) Items

- Unambiguous LR(1) grammar:

  $S \rightarrow L \mathbf{=} R$
  $\quad\quad \mathbf{|}\ R$
  $L \rightarrow \mathbf{*}\ R$
  $\quad\quad \mathbf{|}\ \mathbf{id}$
  $R \rightarrow L$

- Augment with $S' \rightarrow S$
- LR(1) items (next slide)

I_0: $[S' \rightarrow \bullet S,$ $\mathbf{\$}]$ goto$(I_0,S)=I_1$
$[S \rightarrow \bullet L\mathbf{=}R,$ $\mathbf{\$}]$ goto$(I_0,L)=I_2$
$[S \rightarrow \bullet R,$ $\mathbf{\$}]$ goto$(I_0,R)=I_3$
$[L \rightarrow \bullet \mathbf{*}R,$ $\mathbf{=/\$}]$ goto$(I_0,\mathbf{*})=I_4$
$[L \rightarrow \bullet \mathbf{id},$ $\mathbf{=/\$}]$ goto$(I_0,\mathbf{id})=I_5$
$[R \rightarrow \bullet L,$ $\mathbf{\$}]$ goto$(I_0,L)=I_2$

I_1:
$[S' \rightarrow S\bullet,$ $\mathbf{\$}]$

I_2:
$[S \rightarrow L\bullet\mathbf{=}R,$ $\mathbf{\$}]$ goto$(I_0,\mathbf{=})=I_6$
$[R \rightarrow L\bullet,$ $\mathbf{\$}]$

I_3:
$[S \rightarrow R\bullet,$ $\mathbf{\$}]$

I_4:
$[L \rightarrow \mathbf{*}\bullet R,$ $\mathbf{=/\$}]$ goto$(I_4,R)=I_7$
$[R \rightarrow \bullet L,$ $\mathbf{=/\$}]$ goto$(I_4,L)=I_8$
$[L \rightarrow \bullet \mathbf{*}R,$ $\mathbf{=/\$}]$ goto$(I_4,\mathbf{*})=I_4$
$[L \rightarrow \bullet \mathbf{id},$ $\mathbf{=/\$}]$ goto$(I_4,\mathbf{id})=I_5$

I_5:
$[L \rightarrow \mathbf{id}\bullet,$ $\mathbf{=/\$}]$

I_6: $[S \rightarrow L\mathbf{=}\bullet R,$ $\mathbf{\$}]$ goto$(I_6,R)=I_9$
$[R \rightarrow \bullet L,$ $\mathbf{\$}]$ goto$(I_6,L$
$[L \rightarrow \bullet \mathbf{*}R,$ $\mathbf{\$}]$ goto$(I_6,\mathbf{*})=I_{11}$
$[L \rightarrow \bullet \mathbf{id},$ $\mathbf{\$}]$ goto$(I_6,\mathbf{id})=I_{12}$

I_7:
$[L \rightarrow \mathbf{*}R\bullet,$ $\mathbf{=/\$}]$

I_8:
$[R \rightarrow L\bullet,$ $\mathbf{=/\$}]$

I_9:
$[S \rightarrow L\mathbf{=}R\bullet,$ $\mathbf{\$}]$

I_{10}:
$[R \rightarrow L\bullet,$ $\mathbf{\$}]$

I_{11}:
$[L \rightarrow \mathbf{*}\bullet R,$ $\mathbf{\$}]$ goto$(I_{11},R)=I_{13}$
$[R \rightarrow \bullet L,$ $\mathbf{\$}]$ goto$(I_{11},L$
$[L \rightarrow \bullet \mathbf{*}R,$ $\mathbf{\$}]$ goto$(I_{11},\mathbf{*})=I_{11}$
$[L \rightarrow \bullet \mathbf{id},$ $\mathbf{\$}]$ goto$(I_{11},\mathbf{id})=I_{12}$

I_{12}:
$[L \rightarrow \mathbf{id}\bullet,$ $\mathbf{\$}]$

I_{13}:

Constructing Canonical LR(1) Parsing Tables

1. Augment the grammar with $S' \rightarrow S$
2. Construct the set $C=\{I_0,I_1,\ldots,I_n\}$ of LR(1) items
3. If $[A\rightarrow\alpha\bullet a\beta, b] \in I_i$ and $goto(I_i,a)=I_j$ then set $action[i,a]$=shift j
4. If $[A\rightarrow\alpha\bullet, a] \in I_i$ then set $action[i,a]$=reduce $A\rightarrow\alpha$ (apply only if $A\neq S'$)
5. If $[S'\rightarrow S\bullet, \$]$ is in I_i then set $action[i,\$]$=accept
6. If $goto(I_i,A)=I_j$ then set $goto[i,A]=j$
7. Repeat 3-6 until no more entries added
8. The initial state i is the I_i holding item $[S'\rightarrow\bullet S,\$]$

Example LR(1) Parsing Table

Grammar:

1. $S' \to S$
2. $S \to L$ **=** R
3. $S \to R$
4. $L \to$ __*__ R
5. $L \to$ **id**
6. $R \to L$

	id	__*__	**=**	**$**	*S*	*L*	*R*
0	s5	s4			1	2	3
1				acc			
2			s6	r6			
3				r3			
4	s5	s4				8	7
5			r5	r5			
6	s12	s11				10	4
7			r4	r4			
8			r6	r6			
9				r2			
10				r6			
11	s12	s11				10	13
12				r5			
13				r4			

LALR(1) Grammars

- LR(1) parsing tables have many states
- LALR(1) parsing (Look-Ahead LR) combines LR(1) states to reduce table size
- Less powerful than LR(1)
  - Will not introduce shift-reduce conflicts, because shifts do not use lookaheads
  - May introduce reduce-reduce conflicts, but seldom do so for grammars of programming languages

Constructing LALR(1) Parsing Tables

1. Construct sets of LR(1) items
2. Combine LR(1) sets with sets of items that share the same first part

I_4: $[L \rightarrow *\bullet R,\ \mathbf{=}]$
$[R \rightarrow \bullet L,\ \mathbf{=}]$
$[L \rightarrow \bullet *R,\ \mathbf{=}]$
$[L \rightarrow \bullet \mathbf{id},\ \mathbf{=}]$

I_{11}: $[L \rightarrow *\bullet R,\ \mathbf{\$}]$
$[R \rightarrow \bullet L,\ \mathbf{\$}]$
$[L \rightarrow \bullet *R,\ \mathbf{\$}]$
$[L \rightarrow \bullet \mathbf{id},\ \mathbf{\$}]$

$[L \rightarrow *\bullet R,\ \mathbf{=/\$}]$
$[R \rightarrow \bullet L,\ \mathbf{=/\$}]$
$[L \rightarrow \bullet *R,\ \mathbf{=/\$}]$
$[L \rightarrow \bullet \mathbf{id},\ \mathbf{=/\$}]$

Shorthand for two items in the same set

Example LALR(1) Grammar

- Unambiguous LR(1) grammar:

$$
\begin{aligned}
S &\rightarrow L \mathbf{=} R \\
&\quad \mid R \\
L &\rightarrow \mathbf{*}\, R \\
&\quad \mid \mathbf{id} \\
R &\rightarrow L
\end{aligned}
$$

- Augment with $S' \rightarrow S$
- LALR(1) items (next slide)

I_0: $[S' \rightarrow \bullet S,$ $\quad$ **$**$]$ $\text{goto}(I_0,S)=I_1$

$[S \rightarrow \bullet L\mathbf{=}R,$ $\quad$ **$**$]$ $\text{goto}(I_0,L)=I_2$

$[S \rightarrow \bullet R,$ $\quad$ **$**$]$ $\text{goto}(I_0,R)=I_3$

$[L \rightarrow \bullet{*}R,$ $\quad$ **=/$**$]$ $\text{goto}(I_0,{*})=I_4$

$[L \rightarrow \bullet\mathbf{id},$ $\quad$ **=/$**$]$ $\text{goto}(I_0,\mathbf{id})=I_5$

$[R \rightarrow \bullet L,$ $\quad$ **$**$]$ $\text{goto}(I_0,L)=I_2$

I_1:

$[S' \rightarrow S\bullet,$ $\quad$ **$**$]$

I_2:

$[S \rightarrow L\bullet\mathbf{=}R,$ $\quad$ **$**$]$ $\text{goto}(I_0,\mathbf{=})=I_6$

$[R \rightarrow L\bullet,$ $\quad$ **$**$]$

I_3:

$[S \rightarrow R\bullet,$ $\quad$ **$**$]$

I_4:

$[L \rightarrow {*}\bullet R,$ $\quad$ **=/$**$]$ $\text{goto}(I_4,R)=I_7$

$[R \rightarrow \bullet L,$ $\quad$ **=/$**$]$ $\text{goto}(I_4,L)=I_9$

$[L \rightarrow \bullet{*}R,$ $\quad$ **=/$**$]$ $\text{goto}(I_4,{*})=I_4$

$[L \rightarrow \bullet\mathbf{id},$ $\quad$ **=/$**$]$ $\text{goto}(I_4,\mathbf{id})=I_5$

I_5:

$[L \rightarrow \mathbf{id}\bullet,$ $\quad$ **=/$**$]$

I_6: $[S \rightarrow L\mathbf{=}\bullet R,$ $\quad$ **$**$]$ $\text{goto}(I_6,R)=I_8$

$[R \rightarrow \bullet L,$ $\quad$ **$**$]$ $\text{goto}(I_6,L$

$[L \rightarrow \bullet{*}R,$ $\quad$ **$**$]$ $\text{goto}(I_6,{*})=I_4$

$[L \rightarrow \bullet\mathbf{id},$ $\quad$ **$**$]$ $\text{goto}(I_6,\mathbf{id})=I_5$

I_7:

$[L \rightarrow {*}R\bullet,$ $\quad$ **=/$**$]$

I_8:

$[S \rightarrow L\mathbf{=}R\bullet,$ $\quad$ **$**$]$

I_9:

$[R \rightarrow L\bullet,$ $\quad$ **=/$**$]$

Shorthand
for two items

$[R \rightarrow L\bullet,$ $\quad$ **=**$]$
$[R \rightarrow L\bullet,$ $\quad$ **$**$]$

Example LALR(1) Parsing Table

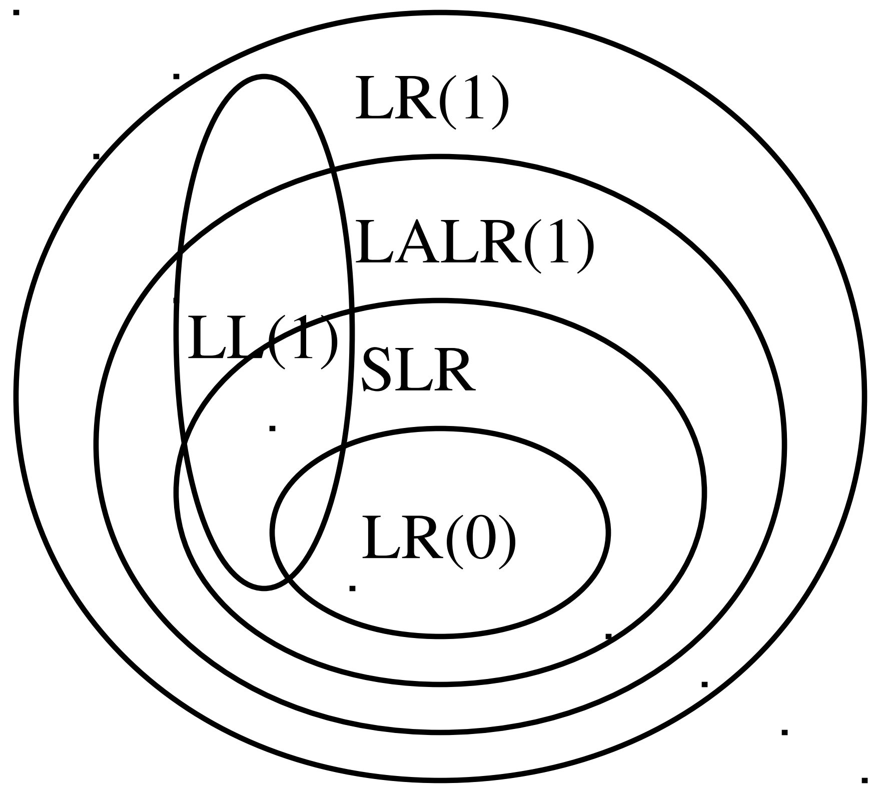
Grammar:
1. $S' \to S$
2. $S \to L = R$
3. $S \to R$
4. $L \to * R$
5. $L \to \mathbf{id}$
6. $R \to L$

	id	*	=	$	*S*	*L*	*R*
0	s5	s4			1	2	3
1				acc			
2			s6	r6			
3				r3			
4	s5	s4				9	7
5			r5	r5			
6	s5	s4				9	8
7			r4	r4			
8				r2			
9			r6	r6			

LL, SLR, LR, LALR Summary

- LL parse tables computed using FIRST/FOLLOW
  - Nonterminals $\times$ terminals $\rightarrow$ productions
  - Computed using FIRST/FOLLOW
- LR parsing tables computed using closure/goto
  - LR states $\times$ terminals $\rightarrow$ shift/reduce actions
  - LR states $\times$ nonterminals $\rightarrow$ goto state transitions
- A grammar is
  - LL(1) if its LL(1) parse table has no conflicts
  - SLR if its SLR parse table has no conflicts
  - LALR(1) if its LALR(1) parse table has no conflicts
  - LR(1) if its LR(1) parse table has no conflicts

LL, SLR, LR, LALR Grammars



Dealing with Ambiguous Grammars

1. $S' \rightarrow E$
2. $E \rightarrow E$ **+** E
3. $E \rightarrow$ **id**

	id	**+**	**$**	*E*
0	s2			1
1		s3	acc	
2		r3	r3	
3	s2			4
4		s3/r2	r2	

Shift/reduce conflict:
action[4,**+**] = shift 4
action[4,**+**] = reduce $E \rightarrow E$ **+** E

stack	input
$ 0	**id+id+id$**
…	…
$ 0 E 1 **+** 3 E 4	**+id$**

When shifting on **+**:
yields right associativity
id+(id+id)

When reducing on **+**:
yields left associativity
(id+id)+id

Using Associativity and Precedence to Resolve Conflicts

- Left-associative operators: reduce
- Right-associative operators: shift
- Operator of higher precedence on stack: reduce
- Operator of lower precedence on stack: shift

$S' \to E$
$E \to E$ **+** E
$E \to E$ ***** E
$E \to$ **id**

stack	input	
$ 0	**id*id+id$**	
…	…	
$ 0 E 1 ***** 3 E 5	**+id$**	reduce $E \to E$ ***** E

Error Detection in LR Parsing

- Canonical LR parser uses full LR(1) parse tables and will never make a single reduction before recognizing the error when a syntax error occurs on the input
- SLR and LALR may still reduce when a syntax error occurs on the input, but will never shift the erroneous input symbol

Error Recovery in LR Parsing

- Panic mode
  - Pop until state with a goto on a nonterminal *A* is found, (where *A* represents a major programming construct), push *A*
  - Discard input symbols until one is found in the FOLLOW set of *A*
- Phrase-level recovery
  - Implement error routines for every error entry in table
- Error productions
  - Pop until state has error production, then shift on stack
  - Discard input until symbol is encountered that allows parsing to continue