CS 335: Bottom-Up Parsing

Swarnendu Biswas

Department of Computer Science and Engineering,
Indian Institute of Technology Kanpur

Sem 2023-24-II

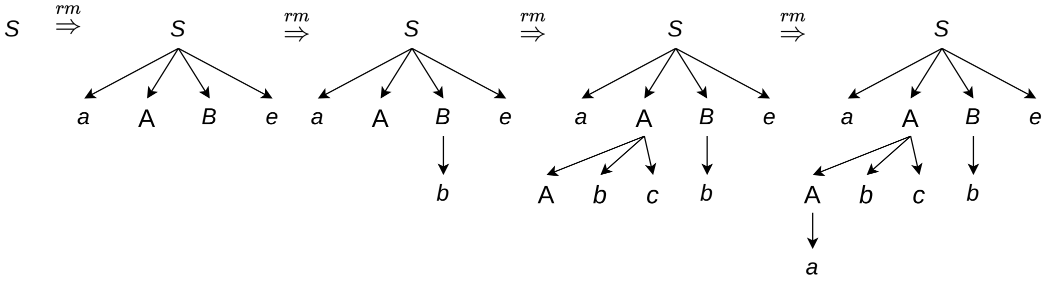
Rightmost Derivation of *abbcde*

Grammar

$$S \rightarrow aABe$$
$$A \rightarrow Abc \mid b$$
$$B \rightarrow d$$

Input string: *abbcde*

$$\begin{aligned} S &\rightarrow aABe \\ &\rightarrow aAde \\ &\rightarrow aAbcde \\ &\rightarrow abbcde \end{aligned}$$



Bottom-Up Parsing

Definition

Bottom-up parsing constructs the parse tree starting from the leaves and working up toward the root

Grammar

$$S \rightarrow aABe$$
$$A \rightarrow Abc \mid b$$
$$B \rightarrow d$$

Input string: *abbcde*

$S \rightarrow aABe$	$abbcde$
$\rightarrow aAde$	$\rightarrow aAbcde$
$\rightarrow aAbcde$	$\rightarrow aAde$
$\rightarrow abbcde$	$\rightarrow aABe$
	$\rightarrow S$

$\uparrow$ reverse of rightmost derivation

Bottom-Up Parsing

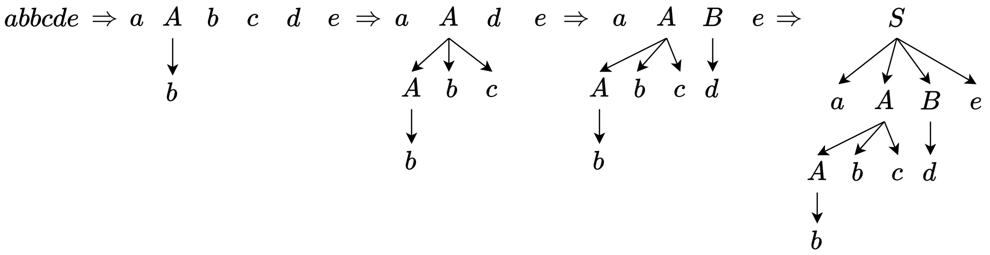
Grammar

$$S \rightarrow aABe$$
$$A \rightarrow Abc \mid b$$
$$B \rightarrow d$$

Input string: *abbcde*

$S \rightarrow aABe$	$abbcde$
$\rightarrow aAde$	$\rightarrow aAbcde$
$\rightarrow aAbcde$	$\rightarrow aAde$
$\rightarrow abbcde$	$\rightarrow aABe$
	$\rightarrow S$

reverse of rightmost derivation



Reduction

Bottom-up parsing reduces a string *w* to the start symbol *S*

At each reduction step, a chosen substring that is the RHS (or body) of a production is replaced by the LHS (or head) nonterminal

rightmost derivation

$$S \overset{rm}{\Rightarrow} \gamma_0 \overset{rm}{\Rightarrow} \gamma_1 \overset{rm}{\Rightarrow} \cdots \overset{rm}{\Rightarrow} \gamma_{n-1} \overset{rm}{\Rightarrow} \gamma_n = w$$

bottom-up parser

Handle

- Handle is a substring that matches the body of a production
- Reducing the handle is one step in the reverse of the rightmost derivation

$$E \rightarrow E + T \mid T$$
$$T \rightarrow T * F \mid F$$
$$F \rightarrow (E) \mid \textbf{id}$$

Right sentential form	Handle	Reducing Production
$\textbf{id}_1 * \textbf{id}_2$	$\textbf{id}_1$	$F \rightarrow \textbf{id}$
$F * \textbf{id}_2$	F	$T \rightarrow F$
$T * \textbf{id}_2$	$\textbf{id}_2$	$F \rightarrow \textbf{id}$
$T * F$	$T * F$	$T \rightarrow T * F$
T	T	$E \rightarrow T$
E		

Handle

- Handle is a substring that matches the body of a production
- Reducing the handle is one step in the reverse of the rightmost derivation

$$E \rightarrow E + T \mid T$$
$$T \rightarrow T * F \mid F$$
$$F \rightarrow (E) \mid \textbf{id}$$

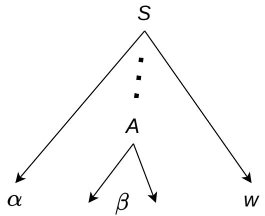
Right sentential form	Handle	Reducing Production
$\textbf{id}_1 * \textbf{id}_2$	$\textbf{id}_1$	$F \rightarrow \textbf{id}$
$F * \textbf{id}_2$	F	$T \rightarrow F$
$T * \textbf{id}_2$	$\textbf{id}_2$	$F \rightarrow \textbf{id}$
$T * F$	$T * F$	$T \rightarrow T * F$
T	T	$E \rightarrow T$
E		

Although T is the body of the production $E \rightarrow T$, T is not a handle in the sentential form $T * \textbf{id}_2$

The leftmost substring that matches the body of some production need not be a handle

Handle

- If $S \underset{rm}{\overset{*}{\Longrightarrow}} \alpha A w \underset{rm}{\Longrightarrow} \alpha\beta w$, then $A \rightarrow \beta$ is a handle of $\alpha\beta w$
- String w right of a handle must contain only terminals



A handle $A \rightarrow \beta$ in the parse tree for $\alpha\beta w$

- If grammar G is unambiguous, then every right sentential form has only one handle
- If G is ambiguous, then there can be more than one rightmost derivation of $\alpha\beta w$

Shift-Reduce Parsing

Shift-Reduce Parsing

- The input string being parsed consists of two parts
  - Left part is a string of terminals and nonterminals, and is stored in a stack
  - Right part is a string of terminals to be read from an input buffer
  - Bottom of the stack and end of the input are represented by $
- Shift-reduce parsing is a type of bottom-up parsing with **two primary actions**, shift and reduce
  - Shift-Reduce actions

    Shift Shift the next input symbol from the right string onto the top of the stack

    Reduce Identify a string on top of the stack that is the body of a production and replace the body with the head
  - Other actions are accept and error

Shift-Reduce Parsing

- Initial

Stack	Input
\$	w\$

Shift $\Downarrow^*$ Reduce

- Goal

Stack	Input
\$$S$	\$

Example of Shift-Reduce Parsing

$$E \rightarrow E + T \mid T$$
$$T \rightarrow T * F \mid F$$
$$F \rightarrow (E) \mid \textbf{id}$$

Stack	Input	Action
$\$$	$\textbf{id}_1 * \textbf{id}_2\$$	Shift
$\$\textbf{id}_1$	$*\textbf{id}_2\$$	Reduce by $F \rightarrow \textbf{id}$
$\$F$	$*\textbf{id}_2\$$	Reduce by $T \rightarrow F$
$\$T$	$*\textbf{id}_2\$$	Shift
$\$T*$	$\textbf{id}_2\$$	Shift
$\$T * \textbf{id}_2$	$\$$	Reduce by $F \rightarrow \textbf{id}$
$\$T * F$	$\$$	Reduce by $T \rightarrow T * F$
$\$T$	$\$$	Reduce by $E \rightarrow T$
$\$E$	$\$$	Accept

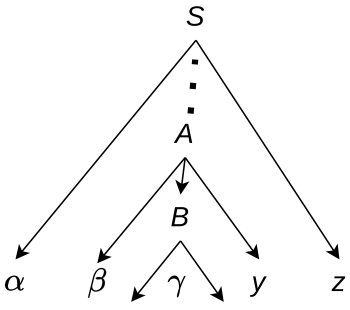
or report an error in case of syntax error

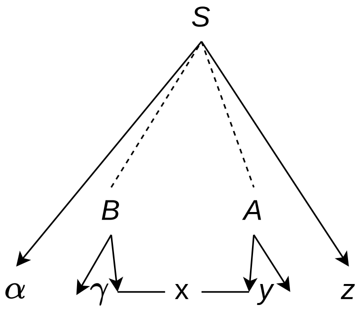
Handle on Top of Stack

Is the following scenario possible?

Stack	Input	Action
$\dots$		
$\$\alpha\beta\gamma$	$w\$$	Reduce by $A \rightarrow \gamma$
$\$\alpha\beta A$	$w\$$	Reduce by $B \rightarrow \beta$
$\$\alpha BA$	$w\$$	$\dots$
$\dots$		

Possible Choices in Rightmost Derivation





1. $S \underset{rm}{\Longrightarrow} \alpha Az \underset{rm}{\Longrightarrow} \alpha\beta Byz \underset{rm}{\Longrightarrow} \alpha\beta\gamma yz$

2. $S \underset{rm}{\Longrightarrow} \alpha BxAz \underset{rm}{\Longrightarrow} \alpha Bxyz \underset{rm}{\Longrightarrow} \alpha\gamma xyz$

Handle on Top of Stack

Is the following scenario possible?

Stack	Input	Action
$\ldots$		
$\$\alpha\beta\gamma$	$w\$$	Reduce by $A \rightarrow \gamma$

Handle will always eventually appear on **top of the stack**, never inside

Shift-Reduce Actions

Shift shift the next input symbol from the right string onto the top of the stack

Reduce identify a string on top of the stack that is the body of a production, and replace the body with the head

> How do you decide when to shift and when to reduce?

Steps in Shift-Reduce Parsers

General shift-reduce technique

- If there is no handle on the stack, then shift
- If there is a handle on the stack, then reduce

Bottom-up parsing is essentially the process of identifying handles and reducing them

- Different bottom-up parsers differ in the way they **detect** handles

Challenges in Bottom-up Parsing

Which action do you pick when both shift and reduce are valid?

Implies a shift-reduce conflict

Which rule to use if reduction is possible by more than one rule?

Implies a reduce-reduce conflict

Example of a Shift-Reduce Conflict

$$E \rightarrow E + E \mid E * E \mid \textbf{id}$$

$\textbf{id} + \textbf{id} * \textbf{id}$

Stack	Input	Action
$\$$	$\textbf{id} + \textbf{id} * \textbf{id}\$$	Shift
$\ldots$		
$\$E + E$	$*\textbf{id}\$$	Reduce by $E \rightarrow E + E$
$\$E$	$*\textbf{id}\$$	Shift
$\$E*$	$\textbf{id}\$$	Shift
$\$E * \textbf{id}$	$\$$	Reduce by $E \rightarrow \textbf{id}$
$\$E * E$	$\$$	Reduce by $E \rightarrow E * E$
$\$E$	$\$$	

$c + C$

Stack	Input	Action
$\$$	$\textbf{id} + \textbf{id} * \textbf{id}\$$	Shift
$\ldots$		
$\$E + E$	$*\textbf{id}\$$	Shift
$\$E + E*$	$\textbf{id}\$$	Shift
$\$E + E * \textbf{id}$	$\$$	Reduce by $E \rightarrow \textbf{id}$
$\$E + E * E$	$\$$	Reduce by $E \rightarrow E * E$
$\$E + E$	$\$$	Reduce by $E \rightarrow E + E$
$\$E$	$\$$	

Resolving Shift-Reduce Conflict

$$
\begin{aligned}
\mathit{Stmt} \rightarrow\ & \textbf{if } \mathit{Expr} \textbf{ then } \mathit{Stmt} \\
\mid\ & \textbf{if } \mathit{Expr} \textbf{ then } \mathit{Stmt} \textbf{ else } \mathit{Stmt} \\
\mid\ & \mathit{other}
\end{aligned}
$$

Stack	Input	Action
...		
$\$ \ldots$ **if** *Expr* **then** *Stmt*	**else** $\ldots$	

What is a correct thing to do for this grammar – shift or reduce? We can prioritize shifts.

Reduce-Reduce Conflict

$$M \rightarrow R + R \mid R + c \mid R$$
$$R \rightarrow c$$

$c + c$

Stack	Input	Action
$\$$	$c + c\$$	Shift
$\$c$	$+c\$$	Reduce by $R \rightarrow c$
$\$R$	$+c\$$	Shift
$\$R+$	$c\$$	Shift
$\$R+c$	$\$$	Reduce by $R \rightarrow c$
$\$R+R$	$\$$	Reduce by $R \rightarrow R + R$
$\$M$	$\$$	

id $+$ **id** $*$ **id**

Stack	Input	Action
$\$$	$c + c\$$	Shift
$\$c$	$+c\$$	Reduce by $R \rightarrow c$
$\$R$	$+c\$$	Shift
$\$R+$	$c\$$	Shift
$\$R+c$	$\$$	Reduce by $M \rightarrow R + c$
$\$M$	$\$$	

LR Parsing

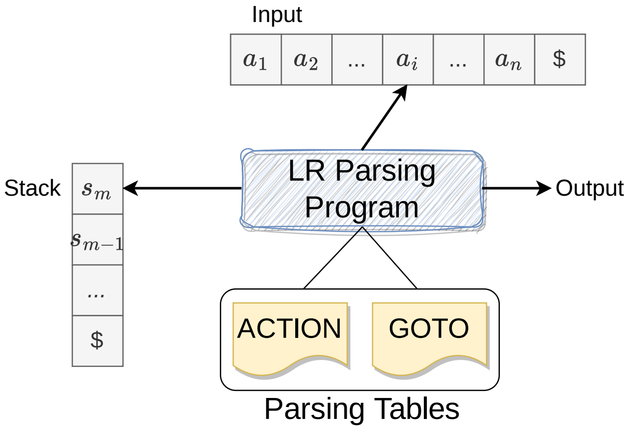
LR(k) Parsing

- Popular bottom-up parsing scheme
  - L is for left-to-right scan of input, R is for **reverse of rightmost** derivation, k is the number of lookahead symbols
- LR parsers are table-driven, like the non-recursive LL parser
- LR grammar is one for which we can construct an LR parsing table
- Popularity of LR Parsing
  - + Most general non-backtracking shift-reduce parsing method
  - + Can recognize almost all language constructs with CFGs
  - + Works for a superset of grammars parsed with predictive or LL parsers

LR(k) Parsing

- Popular bottom-up parsing scheme
  - L is for left-to-right scan of input, R is for **reverse of rightmost** derivation, k is the number of lookahead symbols
- LR parsers are table-driven, like the non-recursive LL parser
- LR grammar is one for which we can construct an LR parsing table
- Popularity of LR Parsing
  - \+ Most general non-backtracking shift-reduce parsing method
  - \+ Can recognize almost all language constructs with CFGs
  - \+ Works for a superset of grammars parsed with predictive or LL parsers
    - LL(k) parsing predicts which production to use having seen only the first k tokens of the right-hand side
    - LR(k) parsing can decide after it has seen input tokens corresponding to the entire right-hand side of the production

Block Diagram of LR Parser



The LR parsing driver is the same for all LR parsers, only the parsing tables (i.e., ACTION and GOTO) change across parser types

Steps in LR Parsing

- Remember the basic questions: **when to shift** and **when to reduce**!
- An LR parser makes shift-reduce decisions by maintaining states
- Information is encoded in a DFA constructed using a canonical LR(0) collection
  1. Augmented grammar G' with new start symbol S' and rule $S' \rightarrow S$
  2. Define helper functions Closure() and Goto()

LR(0) Item

- An LR(0) item of a grammar G is a production of G with a dot ($\bullet$) at some position in the body
- An item indicates how much of a production we have seen
  - Symbols on the left of "$\bullet$" are already on the stack
  - Symbols on the right of "$\bullet$" are expected in the input
- $A \rightarrow \bullet XYZ$ indicates that we expect a string derivable from XYZ next in the input
- $A \rightarrow X \bullet YZ$ indicates that we saw a string derivable from X in the input, and we expect a string derivable from YZ next in the input
- $A \rightarrow \epsilon$ generates only one item $A \rightarrow \bullet$

Production	**Items**
$A \rightarrow XYZ$	$A \rightarrow \bullet XYZ$ $A \rightarrow X \bullet YZ$ $A \rightarrow XY \bullet Z$ $A \rightarrow XYZ\bullet$

Closure Operation

- Let I be a set of items for a grammar G
- Closure(I) is constructed as follows
  (i) Add every item in I to Closure(I)
  (ii) If $A \rightarrow \alpha \bullet B\beta$ is in Closure(I) and $B \rightarrow \gamma$ is a rule in G, then add $B \rightarrow \bullet\gamma$ to Closure(I) if not already added
  (iii) Repeat until no more new items can be added to Closure(I)

A substring derivable from $B\beta$ will have a prefix derivable from B by applying one the B productions

$$
\begin{aligned}
E' &\rightarrow E \\
E &\rightarrow E + T \mid T \\
T &\rightarrow T * F \mid F \\
F &\rightarrow (E) \mid \textbf{id}
\end{aligned}
$$

Suppose $I = \{E' \rightarrow \bullet E\}$

$$
\begin{aligned}
\text{Closure}(I) = \{E' &\rightarrow \bullet E, \\
E &\rightarrow \bullet E + T, \\
E &\rightarrow \bullet T, \\
T &\rightarrow \bullet T * F, \\
T &\rightarrow \bullet F, \\
F &\rightarrow \bullet (E), \\
F &\rightarrow \bullet \textbf{id}\}
\end{aligned}
$$

Goto Operation

- Suppose I is a set of items and X is a grammar symbol
- Goto(I, X) is the closure of set all items $[A \rightarrow \alpha X \bullet \beta]$ such that $[A \rightarrow \alpha \bullet X\beta]$ is in I
  - If I is a set of items for some valid prefix α, then Goto(I, X) is the set of valid items for prefix αX

Intuitively, Goto(I, X) gives the transition of the state I under input X in the LR(0) automaton

$$
\begin{aligned}
E' &\rightarrow E \\
E &\rightarrow E + T \mid T \\
T &\rightarrow T * F \mid F \\
F &\rightarrow (E) \mid \textbf{id}
\end{aligned}
$$

Suppose

$$
\begin{aligned}
I = \{&E' \rightarrow E\bullet, \\
&E \rightarrow E \bullet +T\}
\end{aligned}
$$

$$
\begin{aligned}
\text{Goto}(I, +) = \{&E \rightarrow E + \bullet T, \\
&T \rightarrow \bullet T * F, \\
&T \rightarrow \bullet F, \\
&F \rightarrow \bullet (E), \\
&F \rightarrow \bullet \textbf{id}\}
\end{aligned}
$$

Algorithm to Compute LR(0) Canonical Collection

$C = \text{Closure}\left(\{S^{'} \rightarrow \bullet S\}\right)$

```
repeat
  for each set of items I ∈ C
    for each grammar symbol X
      if Goto(I,X) is not empty and not in C
        add Goto(I,X) to C
until no new sets of items are added to C
```

Example Computation of LR(0) Canonical Collection

$$\begin{aligned} I_0 &= \text{Closure}(E' \rightarrow \bullet E) \\ &= \{E' \rightarrow \bullet E, \\ &\quad E \rightarrow \bullet E + T, \\ &\quad E \rightarrow \bullet T, \\ &\quad T \rightarrow \bullet T * F, \\ &\quad T \rightarrow \bullet F, \\ &\quad F \rightarrow \bullet (E), \\ &\quad F \rightarrow \bullet \mathbf{id}\} \end{aligned}$$

$$\begin{aligned} I_1 &= \text{Goto}(I_0, E) \\ &= \{E' \rightarrow E \bullet, \\ &\quad E \rightarrow E \bullet +T\} \end{aligned}$$

$$\begin{aligned} I_2 &= \text{Goto}(I_0, T) \\ &= \{E \rightarrow T \bullet, \\ &\quad T \rightarrow T \bullet *F\} \end{aligned}$$

$$\begin{aligned} I_3 &= \text{Goto}(I_0, F) \\ &= \{T \rightarrow F \bullet\} \end{aligned}$$

$$\begin{aligned} I_4 &= \text{Goto}(I_0, `(\text{'}) \\ &= \{F \rightarrow (\bullet E), \\ &\quad E \rightarrow \bullet E + T, \\ &\quad E \rightarrow \bullet T, \\ &\quad T \rightarrow \bullet T * F, \\ &\quad T \rightarrow \bullet F, \\ &\quad F \rightarrow \bullet (E), \\ &\quad F \rightarrow \bullet \mathbf{id}\} \end{aligned}$$

$$\begin{aligned} I_5 &= \text{Goto}(I_0, \mathbf{id}) \\ &= \{F \rightarrow \mathbf{id} \bullet\} \end{aligned}$$

$$\begin{aligned} I_6 &= \text{Goto}(I_1, +) \\ &= \{E \rightarrow E + \bullet T, \\ &\quad T \rightarrow \bullet T * F, \\ &\quad T \rightarrow \bullet F, \\ &\quad F \rightarrow \bullet (E), \\ &\quad F \rightarrow \bullet \mathbf{id}\} \end{aligned}$$

$$\begin{aligned} I_7 &= \text{Goto}(I_2, *) \\ &= \{T \rightarrow T * \bullet F, \\ &\quad F \rightarrow \bullet (E), \\ &\quad F \rightarrow \bullet \mathbf{id}\} \end{aligned}$$

$$\begin{aligned} I_8 &= \text{Goto}(I_4, E) \\ &= \{E \rightarrow E \bullet +T, \\ &\quad F \rightarrow (E \bullet)\} \end{aligned}$$

$$\begin{aligned} I_9 &= \text{Goto}(I_6, T) \\ &= \{E \rightarrow E + T \bullet, \\ &\quad T \rightarrow T \bullet *F\} \end{aligned}$$

$$\begin{aligned} I_{10} &= \text{Goto}(I_7, F) \\ &= \{T \rightarrow T * F \bullet\} \end{aligned}$$

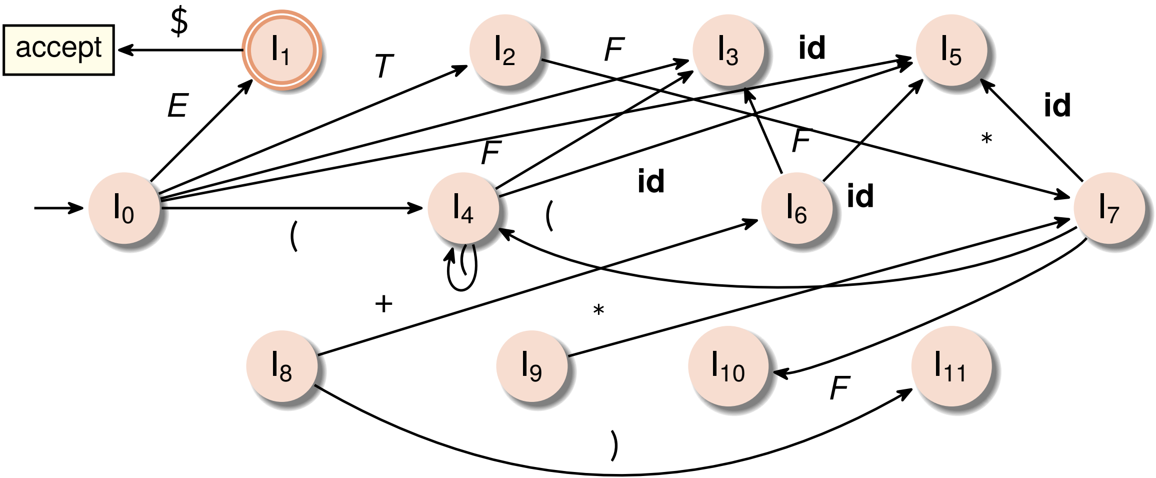
$$\begin{aligned} I_{11} &= \text{Goto}(I_8, `)\text{'}) \\ &= \{F \rightarrow (E) \bullet\} \end{aligned}$$

$$\begin{aligned} I_2 &= \text{Goto}(I_4, T) \\ I_3 &= \text{Goto}(I_4, F) \\ I_4 &= \text{Goto}(I_4, `(\text{'}) \\ I_5 &= \text{Goto}(I_4, \mathbf{id}) \\ I_3 &= \text{Goto}(I_6, F) \\ I_4 &= \text{Goto}(I_6, `(\text{'}) \\ I_5 &= \text{Goto}(I_6, \mathbf{id}) \\ I_4 &= \text{Goto}(I_7, `(\text{'}) \\ I_5 &= \text{Goto}(I_7, \mathbf{id}) \\ I_6 &= \text{Goto}(I_8, +) \\ I_7 &= \text{Goto}(I_9, *) \end{aligned}$$

LR(0) Automaton

- Canonical LR(0) collection is used for constructing the LR(0) automaton for parsing
- States represent sets of LR(0) items in the canonical LR(0) collection
  - Start state is $\text{Closure}\left(\{S' \rightarrow \bullet S\}\right)$, where S' is the start symbol of the augmented grammar
  - State j refers to the state corresponding to the set of items I_j
- By construction, all transitions to state j is for the same symbol X
  - Each state, except the start state, has a unique grammar symbol associated with it

LR(0) Automaton



Use of LR(0) Automaton

- How can the LR(0) automaton help with shift-reduce decisions?
- Suppose string γ of grammar symbols takes the automaton from start state S_0 to state S_j
  - Shift on next input symbol a if S_j has a transition on a
  - Otherwise, reduce
    - Items in state S_j help decide which production to use

Structure of LR Parsing Table

- Assume S_i is top of the stack and a_i is the current input symbol
- Parsing table consists of two parts: an ACTION and a GOTO function
- ACTION table is indexed by state and terminal symbols; ACTION$[S_i, a_i]$ can have four values
  (i) Shift a_i to the stack, go to state S_j
  (ii) Reduce by rule k
  (iii) Accept
  (iv) Error (empty cell in the table)
- GOTO table is indexed by state and nonterminal symbols

Constructing LR(0) Parsing Table

(i) Construct LR(0) canonical collection $C = \{I_0, I_1, \ldots, I_n\}$ for grammar $G^{'}$

(ii) State i is constructed from I_i

  (a) If $[A \rightarrow \alpha \bullet A\beta] \in I_i$ and $\text{GOTO}(I_i, a) = I_j$, then set $\text{ACTION}[i, a]$ = "Shift j"

  (b) If $[A \rightarrow \alpha\bullet] \in I_i$, then set $\text{ACTION}[i, a]$ = "Reduce $A \rightarrow \alpha$" for all a

  (c) If $[S^{'} \rightarrow S\bullet] \in I_i$, then set $\text{ACTION}[i, \$]$ = "Accept"

(iii) If $\text{GOTO}(I_i, A) = I_j$, then $\text{GOTO}[i, A] = j$

(iv) All entries left undefined are "errors"

LR(0) Parsing Table

State	ACTION						GOTO		
	id	+	*	(	)	$	E	T	F
0	$s5$			$s4$			1	2	3
1		$s6$				Accept			
2	$r2$	$r2$	$s7, r2$	$r2$	$r2$	$r2$			
3	$r4$	$r4$	$r4$	$r4$	$r4$	$r4$			
4	$s5$			$s4$			8	2	3
5	$r6$	$r6$	$r6$	$r6$	$r6$	$r6$			
6	$s5$			$s4$				9	3
7	$s5$			$s4$					10
8		$s6$				$s11$			
9	$r1$	$r1$	$s7, r1$	$r1$	$r1$	$r1$			
10	$r3$	$r3$	$r3$	$r3$	$r3$	$r3$			
11	$r5$	$r5$	$r5$	$r5$	$r5$	$r5$			

LR Parser Configurations

- A LR parser configuration is a pair $\langle s_0 s_1 \ldots s_m,\ a_i a_{i+1} \ldots a_n\$\rangle$
  - The left half is stack content, and the right half is the remaining input
- Configuration represents the right sentential form $X_1 X_2 \ldots X_m a_i a_{i+1} \ldots a_n$

LR Parsing Algorithm

(i) If ACTION$[s_m, a_i] = sj$, then the new configuration is $\langle s_0 s_1 \dots s_m s_j,\ a_{i+1} \dots a_n \rangle$

(ii) If ACTION$[s_m, a_i] =$ reduce $A \rightarrow \beta$, then the new configuration is $\langle s_0 s_1 \dots s_{m-r} s,\ a_i a_{i+1} \dots a_n \rangle$, where $r = |\beta|$ and $s =$ GOTO$[s_{m-r}, A]$

(iii) If ACTION$[s_m, a_i] =$ Accept, then parsing is successful

(iv) If ACTION$[s_m, a_i] =$ error, then parsing has discovered an error

LR Parsing Program

```
Let a be the first symbol in w$
while (1)
  Let s be the top of the stack
  if ACTION[s,a] == shift t
    push t onto the stack
    let a be the next input symbol
  else if ACTION[s,a] = reduce A → β
    // Reduce with the production A → β
    pop |β| symbols of the stack
    let state t now be the top of the stack
    push GOTO[t,A] onto the stack
  else if ACTION[s,a] == Accept
    break // parsing is complete
  else
    invoke error recovery
```

Shift-Reduce Parser with LR(0) Automaton

Stack	Input	Action
$\$0$	$\mathbf{id} * \mathbf{id}\$$	Shift
$\$0\,\mathbf{id}\,5$	$* \mathbf{id}\$$	Reduce by $F \rightarrow \mathbf{id}$
$\$0\,F\,3$	$* \mathbf{id}\$$	Reduce by $T \rightarrow F$
$\$0\,T\,2$	$* \mathbf{id}\$$	Shift
$\$0\,T\,2 * 7$	$\mathbf{id}\$$	Shift
$\$0\,T\,2 * 7\,\mathbf{id}\,5$	$\$$	Reduce by $F \rightarrow \mathbf{id}$
$\$0\,T\,2 * 7\,F\,10$	$\$$	Reduce by $T \rightarrow T * F$
$\$0\,T\,2$	$\$$	Reduce by $E \rightarrow T$
$\$0\,E\,1$	$\$$	Accept

popped 5 and pushed 3 because $I_3 = \text{Goto}(I_0, F)$

While the stack consisted of only symbols in the shift-reduce parser, here the stack also contains states from the LR(0) automaton

Viable Prefix

- Consider $E \overset{rm}{\Longrightarrow} T \overset{rm}{\Longrightarrow} T * F \overset{rm}{\Longrightarrow} T * \textbf{id} \overset{rm}{\Longrightarrow} F * \textbf{id} \overset{rm}{\Longrightarrow} \textbf{id} * \textbf{id}$
- **id**$*$ is a prefix of a right sentential form but can never appear on the stack
  - Always reduce $F \rightarrow \textbf{id}$ before shifting $*$ (see previous slide)
- Not all prefixes of a right sentential form can appear on the stack
  - LR parser will not shift past the handle
- A **viable prefix** is a prefix of a right sentential form that can appear on the stack of a shift-reduce parser
  - α is a viable prefix if $\exists w$ such that αw is a right sentential form
- There is no error as long as the parser has viable prefixes on the stack

Example of a Viable Prefix

$$S \rightarrow X_1X_2X_3X_4$$
$$A \rightarrow X_1X_2$$

Stack	Input
$\$$	$X_1X_2X_3\$$
$\$X_1$	$X_2X_3\$$
$\$X_1X_2$	$X_3\$$
$\$A$	$X_3\$$
$\$AX_3$	$\$$

$X_1X_2X_3$ can never appear on the stack

- Suppose there is a production $A \rightarrow \beta_1\beta_2$ and $\alpha\beta_1$ is on the stack
  - $\beta_2 \neq \epsilon$ implies that the handle $\beta_1\beta_2$ is not at the top of the stack yet, so shift
  - $\beta_2 = \epsilon$ implies that the parser can reduce by the handle $A \rightarrow \beta_1$

Challenges with LR(0) Parsing

An LR(0) parser works only if each state with a reduce action has only one possible reduce action and no shift action

Ok	Shift-Reduce Conflict	Reduce-Reduce Conflict
$\{L \rightarrow L, S\bullet\}$	$\{L \rightarrow L, S\bullet,$ $L \rightarrow S\bullet, L\}$	$\{L \rightarrow S, L\bullet,$ $L \rightarrow S\bullet\}$

Takes shift/reduce decisions without any lookahead token

Lacks the power to parse programming language grammars

Canonical Collection of Sets of LR(0) Items

Consider the following grammar for adding numbers

Left associative	Right associative	Shift-Reduce Conflict
$S \rightarrow S + E \mid E$ $E \rightarrow \textbf{num}$	$S \rightarrow E + S \mid E$ $E \rightarrow \textbf{num}$	$\{S \rightarrow E \bullet +S,$ $S \rightarrow E\bullet\}$

$$
\begin{aligned}
\text{FIRST}\,(S) &= \{\textbf{num}\} \\
\text{FIRST}\,(E) &= \{\textbf{num}\} \\
\text{FOLLOW}\,(S) &= \{\$\} \\
\text{FOLLOW}\,(E) &= \{+, \$\}
\end{aligned}
$$

$$
\begin{aligned}
I_0 &= \text{Closure}(\{S' \rightarrow \bullet S\}) \\
&= \{S' \rightarrow \bullet S, \\
&\quad S \rightarrow \bullet E + S, \\
&\quad S \rightarrow \bullet E, \\
&\quad E \rightarrow \bullet \textbf{num}\}
\end{aligned}
$$

$$
\begin{aligned}
I_1 &= \text{Goto}(I_0, S) \\
&= \{S' \rightarrow S\bullet\}
\end{aligned}
$$

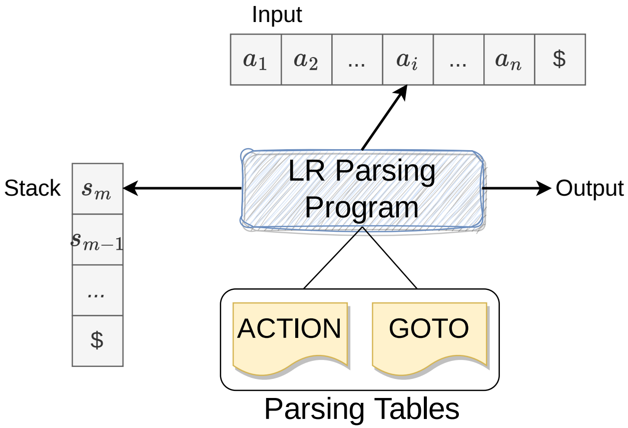
$$
\begin{aligned}
I_2 &= \text{Goto}(I_0, E) \\
&= \{S \rightarrow E \bullet +S, \\
&\quad S \rightarrow E\bullet\}
\end{aligned}
$$

$$
\begin{aligned}
I_3 &= \text{Goto}(I_0, \textbf{num}) \\
&= \{E \rightarrow \textbf{num}\bullet\}
\end{aligned}
$$

$$
\begin{aligned}
I_4 &= \text{Goto}(I_2, +) \\
&= \{S \rightarrow E + \bullet S\}
\end{aligned}
$$

Simple LR Parsing

Block Diagram of LR Parser



The LR parsing driver is the same for all LR parsers, only the parsing tables (i.e., ACTION and GOTO) change across parser types

SLR(1) Parsing

- Uses LR(0) items and LR(0) automaton, extends LR(0) parser to eliminate a few conflicts
  - For each reduction $A \rightarrow \beta$, look at the next symbol c
  - Apply reduction only if $c \in \mathrm{FOLLOW}(\mathrm{A})$ or $c = \epsilon$ and $S \overset{*}{\Rightarrow} \gamma A$

Constructing SLR Parsing Table

(i) Construct LR(0) canonical collection $C = \{I_0, I_1, \ldots, I_n\}$ for grammar G'

(ii) State i is constructed from I_i

  (a) If $[A \rightarrow \alpha \bullet A\beta] \in I_i$ and $\text{GOTO}(I_i, a) = I_j$, then set $\text{ACTION}[i, a] =$ "Shift j"

  (b) If $[A \rightarrow \alpha\bullet] \in I_i$, then set $\text{ACTION}[i, a] =$ "Reduce $A \rightarrow \alpha$" for all a in FOLLOW(A)

  (c) If $[S' \rightarrow S\bullet] \in I_i$, then set $\text{ACTION}[i, \$] =$ "Accept"

(iii) If $\text{GOTO}(I_i, A) = I_j$, then $\text{GOTO}[i, A] = j$

(iv) All entries left undefined are "errors"

constraints on when reductions are applied

SLR Parsing for Expression Grammar

Rule #	Rule
1	$E \longrightarrow E + T$
2	$E \longrightarrow T$
3	$T \longrightarrow T * F$
4	$T \longrightarrow F$
5	$F \longrightarrow (E)$
6	$F \longrightarrow \textbf{id}$

$\text{FIRST}\,(E) = \{(, \textbf{id}\}$

$\text{FIRST}\,(T) = \{(, \textbf{id}\}$

$\text{FIRST}\,(F) = \{(, \textbf{id}\}$

$\text{FOLLOW}\,(E) = \{\$, +,)\}$

$\text{FOLLOW}\,(T) = \{\$, +,)\}$

$\text{FOLLOW}\,(F) = \{\$, +, *,)\}$

- *sj* means shift and stack state *j*
- *rj* means reduce by rule \$*j*
- Accept means accept
- blank means error

Canonical Collection of Sets of LR(0) Items

$$
\begin{aligned}
I_0 &= \text{Closure}(E^{'} \rightarrow \bullet E) \\
&= \{E^{'} \rightarrow \bullet E, \\
&\quad E \rightarrow \bullet E + T, \\
&\quad E \rightarrow \bullet T, \\
&\quad T \rightarrow \bullet T * F, \\
&\quad T \rightarrow \bullet F, \\
&\quad F \rightarrow \bullet\,(E)\,, \\
&\quad F \rightarrow \bullet \textbf{id}\}
\end{aligned}
$$

$$
\begin{aligned}
I_1 &= \text{Goto}(I_0, E) \\
&= \{E^{'} \rightarrow E \bullet, \\
&\quad E \rightarrow E \bullet +T\}
\end{aligned}
$$

$$
\begin{aligned}
I_2 &= \text{Goto}(I_0, T) \\
&= \{E \rightarrow T \bullet, \\
&\quad T \rightarrow T \bullet *F\}
\end{aligned}
$$

$$
\begin{aligned}
I_3 &= \text{Goto}(I_0, F) \\
&= \{T \rightarrow F \bullet\}
\end{aligned}
$$

$$
\begin{aligned}
I_4 &= \text{Goto}(I_0, \text{`('}) \\
&= \{F \rightarrow (\bullet E)\,, \\
&\quad E \rightarrow \bullet E + T, \\
&\quad E \rightarrow \bullet T, \\
&\quad T \rightarrow \bullet T * F, \\
&\quad T \rightarrow \bullet F, \\
&\quad F \rightarrow \bullet\,(E)\,, \\
&\quad F \rightarrow \bullet \textbf{id}\}
\end{aligned}
$$

$$
\begin{aligned}
I_5 &= \text{Goto}(I_0, \textbf{id}) \\
&= \{F \rightarrow \textbf{id} \bullet\}
\end{aligned}
$$

$$
\begin{aligned}
I_6 &= \text{Goto}(I_1, +) \\
&= \{E \rightarrow E + \bullet T, \\
&\quad T \rightarrow \bullet T * F, \\
&\quad T \rightarrow \bullet F, \\
&\quad F \rightarrow \bullet (E), \\
&\quad F \rightarrow \bullet \textbf{id}\}
\end{aligned}
$$

$$
\begin{aligned}
I_7 &= \text{Goto}(I_2, *) \\
&= \{T \rightarrow T * \bullet F, \\
&\quad F \rightarrow \bullet (E), \\
&\quad F \rightarrow \bullet \textbf{id}\}
\end{aligned}
$$

$$
\begin{aligned}
I_8 &= \text{Goto}(I_4, E) \\
&= \{E \rightarrow E \bullet +T, \\
&\quad F \rightarrow (E \bullet)\}
\end{aligned}
$$

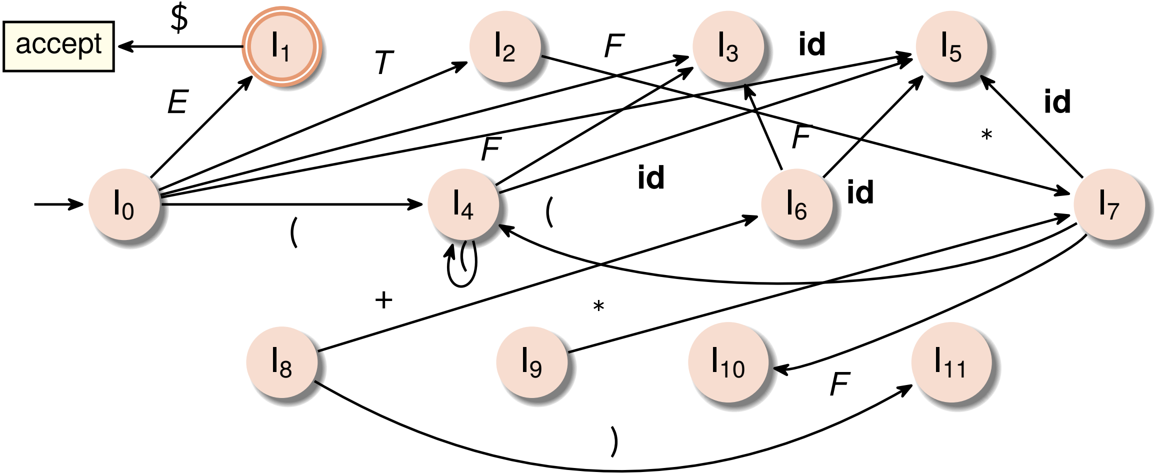
$$
\begin{aligned}
I_9 &= \text{Goto}(I_6, T) \\
&= \{E \rightarrow E + T \bullet, \\
&\quad T \rightarrow T \bullet *F\}
\end{aligned}
$$

$$
\begin{aligned}
I_{10} &= \text{Goto}(I_7, F) \\
&= \{T \rightarrow T * F \bullet\}
\end{aligned}
$$

$$
\begin{aligned}
I_{11} &= \text{Goto}(I_8, \text{`)'}) \\
&= \{F \rightarrow (E) \bullet\}
\end{aligned}
$$

$$
\begin{aligned}
I_2 &= \text{Goto}(I_4, T) \\
I_3 &= \text{Goto}(I_4, F) \\
I_4 &= \text{Goto}(I_4, \text{`('}) \\
I_5 &= \text{Goto}(I_4, \textbf{id}) \\
I_3 &= \text{Goto}(I_6, F) \\
I_4 &= \text{Goto}(I_6, \text{`('}) \\
I_5 &= \text{Goto}(I_6, \textbf{id}) \\
I_4 &= \text{Goto}(I_7, \text{`('}) \\
I_5 &= \text{Goto}(I_7, \textbf{id}) \\
I_6 &= \text{Goto}(I_8, +) \\
I_7 &= \text{Goto}(I_9, *)
\end{aligned}
$$

LR(0) Automaton



SLR Parsing Table

State	ACTION						GOTO		
	id	+	*	(	)	$	E	T	F
0	$s5$			$s4$			1	2	3
1		$s6$				Accept			
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		$r5$	$r5$		$r5$	$r5$			

Moves of an LR Parser on $\textbf{id} * \textbf{id} + \textbf{id}$

Stack	Input	Action
$\$0$	$\textbf{id} * \textbf{id} + \textbf{id}\$$	Shift
$\$0\ \textbf{id}\ 5$	$* \textbf{id} + \textbf{id}\$$	Reduce by $F \longrightarrow \textbf{id}$
$\$0\ F\ 3$	$* \textbf{id} + \textbf{id}\$$	Reduce by $T \longrightarrow F$
$\$0\ T\ 2$	$* \textbf{id} + \textbf{id}\$$	Shift
$\$0\ T\ 2 * 7$	$\textbf{id} + \textbf{id}\$$	Shift
$\$0\ T\ 2 * 7\ \textbf{id}\ 5$	$+ \textbf{id}\$$	Reduce by $F \longrightarrow \textbf{id}$
$\$0\ T\ 2 * 7\ F\ 10$	$+ \textbf{id}\$$	Reduce by $T \longrightarrow T * F$
$\$0\ T\ 2$	$+ \textbf{id}\$$	Reduce by $E \longrightarrow T$
$\$0\ E\ 1$	$+ \textbf{id}\$$	Shift
$\$0\ E\ 1 + 6$	$\textbf{id}\$$	Shift
$\$0\ E\ 1 + 6\ \textbf{id}\ 5$	$\$$	Reduce by $F \longrightarrow \textbf{id}$
$\$0\ E\ 1 + 6\ F\ 3$	$\$$	Reduce by $T \longrightarrow F$
$\$0\ E\ 1 + 6\ T\ 9$	$\$$	Reduce by $E \longrightarrow E + T$
$\$0\ E\ 1$	$\$$	Accept

Limitations of SLR Parsing

- If an SLR parse table for a grammar does not have multiple entries in any cell, then the grammar is unambiguous
- Every SLR(1) grammar is unambiguous, but there are unambiguous grammars that are not SLR(1)

Example to Highlight Limitations of SLR Parsing

Unambiguous grammar

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

FIRST $(S) = \{*, \mathbf{id}\}$
FIRST $(L) = \{*, \mathbf{id}\}$
FIRST $(R) = \{*, \mathbf{id}\}$
FOLLOW $(S) = \{\$, =\}$
FOLLOW $(L) = \{\$, =\}$
FOLLOW $(R) = \{\$, =\}$

Example derivation

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

Canonical LR(0) Collection

$$
\begin{aligned}
I_0 &= \text{Closure}(S' \rightarrow \bullet S) \\
&= \{S' \rightarrow \bullet S, \\
&\quad S \rightarrow \bullet L = R, \\
&\quad S \rightarrow \bullet R, \\
&\quad L \rightarrow \bullet * R, \\
&\quad L \rightarrow \bullet \textbf{id}, \\
&\quad R \rightarrow \bullet L\}
\end{aligned}
$$

$$
\begin{aligned}
I_1 &= \text{Goto}(I_0, S) \\
&= \{S' \rightarrow S\bullet\}
\end{aligned}
$$

$$
\begin{aligned}
I_2 &= \text{Goto}(I_0, L) \\
&= \{S \rightarrow L\bullet = R, \\
&\quad R \rightarrow L\bullet\}
\end{aligned}
$$

$$
\begin{aligned}
I_3 &= \text{Goto}(I_0, R) \\
&= \{S \rightarrow R\bullet\}
\end{aligned}
$$

$$
\begin{aligned}
I_4 &= \text{Goto}(I_0, R) \\
&= \{L \rightarrow * \bullet R, \\
&\quad R \rightarrow \bullet L, \\
&\quad L \rightarrow \bullet * R, \\
&\quad L \rightarrow \bullet \textbf{id}\}
\end{aligned}
$$

$$
\begin{aligned}
I_5 &= \text{Goto}(I_0, \textbf{id}) \\
&= \{L \rightarrow \bullet \textbf{id}\}
\end{aligned}
$$

$$
\begin{aligned}
I_6 &= \text{Goto}(I_2, =) \\
&= \{S \rightarrow L = \bullet R, \\
&\quad R \rightarrow \bullet L, \\
&\quad L \rightarrow \bullet * R, \\
&\quad L \rightarrow \textbf{id}\}
\end{aligned}
$$

$$
\begin{aligned}
I_7 &= \text{Goto}(I_4, R) \\
&= \{L \rightarrow *R\bullet\}
\end{aligned}
$$

$$
\begin{aligned}
I_8 &= \text{Goto}(I_4, L) \\
&= \{R \rightarrow L\bullet\}
\end{aligned}
$$

$$
\begin{aligned}
I_9 &= \text{Goto}(I_6, R) \\
&= \{S \rightarrow L = R\bullet\}
\end{aligned}
$$

SLR Parsing Table

State	ACTION				GOTO		
	=	*	**id**	$\$$	S	L	R
0		$s4$	$s5$		1	2	3
1				Accept			
2	$s6, r6$			$r6$			
3							
4		$s4$	$s5$			8	7
5	$r5$			$r5$			
6		$s4$	$s5$			8	9
7	$r4$			$r4$			
8	$r6$			$r6$			
9				$r2$			

Shift-Reduce Conflict with SLR Parsing

$$
\begin{aligned}
I_0 &= \text{Closure}(S' \rightarrow \bullet S) \\
&= \{S' \rightarrow \bullet S, \\
&\quad S \rightarrow \bullet L = R, \\
&\quad S \rightarrow \bullet R, \\
&\quad L \rightarrow \bullet * R, \\
&\quad L \rightarrow \bullet \textbf{id},
\end{aligned}
$$

I_1 =

$$
\begin{aligned}
I_2 &= \text{Goto}(I_0, L) \\
&= \{S \rightarrow L \bullet = R, \\
&\quad R \rightarrow L \bullet\}
\end{aligned}
$$

$$
\begin{aligned}
I_3 &= \text{Goto}(I_0, R) \\
&= \{S \rightarrow R \bullet\}
\end{aligned}
$$

$$
\begin{aligned}
I_4 &= \text{Goto}(I_0, R) \\
&= \{L \rightarrow * \bullet R, \\
&\quad R \rightarrow \bullet L,
\end{aligned}
$$

$$
\begin{aligned}
I_6 &= \text{Goto}(I_2, =) \\
&= \{S \rightarrow L = \bullet R, \\
&\quad R \rightarrow \bullet L, \\
&\quad L \rightarrow \bullet * R, \\
&\quad L \rightarrow \textbf{id}\}
\end{aligned}
$$

$$
\begin{aligned}
I_7 &= \text{Goto}(I_4, R) \\
&= \{L \rightarrow * R \bullet\}
\end{aligned}
$$

$$
\begin{aligned}
I_8 &= \text{Goto}(I_4, L) \\
&= \{R \rightarrow L \bullet\}
\end{aligned}
$$

$$
\begin{aligned}
I_9 &= \text{Goto}(I_6, R) \\
&= \{S \rightarrow L = R \bullet\}
\end{aligned}
$$

(i) ACTION$[2, =]$ = Shift 6, or

(ii) ACTION$[2, =]$ = Reduce $R \rightarrow L$ because $= \in$ FOLLOW(R)

Moves of an SLR Parser on **id** = **id**

Stack	Input	Action
\$0	**id** = **id**	Shift 5
\$0**id**5	= **id**	Reduce by $L \rightarrow$ **id**
\$0L2	= **id**	Reduce by $R \rightarrow L$
\$0R3	= **id**	Error

No right sentential form begins with $R = \ldots$

Stack	Input	Action
\$0	**id** = **id**\$	Shift 5
\$0**id**5	= **id**\$	Reduce by $L \rightarrow$ **id**
\$0L2	= **id**\$	Shift 6
\$0L2 = 6	**id**\$	Shift 5
\$0L2 = 6**id**5	\$	Reduce by $L \rightarrow$ **id**
\$0L2 = 6L8	\$	Reduce by $R \rightarrow L$
\$0L2 = 6R9	\$	Reduce by $S \rightarrow L = R$
\$0S1	\$	Accept

Moves of an SLR Parser on **id** = **id**

Stack	Input	Action
\$0	**id** = **id**	Shift 5
\$0**id**5	= **id**	Reduce by $L \rightarrow$ **id**
\$0L2	= **id**	Reduce by $R \rightarrow L$
\$0R3	= **id**	Error

Stack	Input	Action
\$0	**id** = **id**\$	Shift 5
\$0**id**5	= **id**\$	Reduce by $L \rightarrow$ **id**
\$0L2	= **id**\$	Shift 6
\$0L2 = 6	**id**\$	Shift 5
		$\rightarrow$ **id**
		$\rightarrow L$
		$\rightarrow L = R$

State i calls for a reduction by $A \rightarrow \alpha$ if the set of items I_i contains items $[A \rightarrow \alpha \bullet]$ and $a \in$ FOLLOW(A)

- Suppose βA is a viable prefix at the top of the stack
- There may be no right sentential form where a follows βA
  - An LR parser should not reduce by $A \rightarrow \alpha$ in such cases

Moves of an SLR Parser on **id** = **id**

Stack	Input	Action
$0	**id** = **id**	Shift 5
$0**id**5	= **id**	Reduce by $L \to$ **id**
0L$2	= **id**	Reduce by $R \to L$
0R$3	= **id**	Error

Stack	Input	Action
$0	**id** = **id**$	Shift 5
$0**id**5	= **id**$	Reduce by $L \to$ **id**
0L$2	= **id**$	Shift 6
0L$2 = 6	**id**$	Shift 5
0L$2 = 6**id**5	$	Reduce by $L \to$ **id**
		$\to L$
		$\to L = R$

SLR parser **cannot remember** the **left context**

- SLR(1) states only tell us about the sequence on top of the stack, not what is below on the stack

References

- A. Aho et al. Compilers: Principles, Techniques, and Tools. Chapter 4.5–4.8, 2nd edition, Pearson Education.
- K. Cooper and L. Torczon. Engineering a Compiler. Chapter 3.4–3.6, 2nd edition, Morgan Kaufmann.