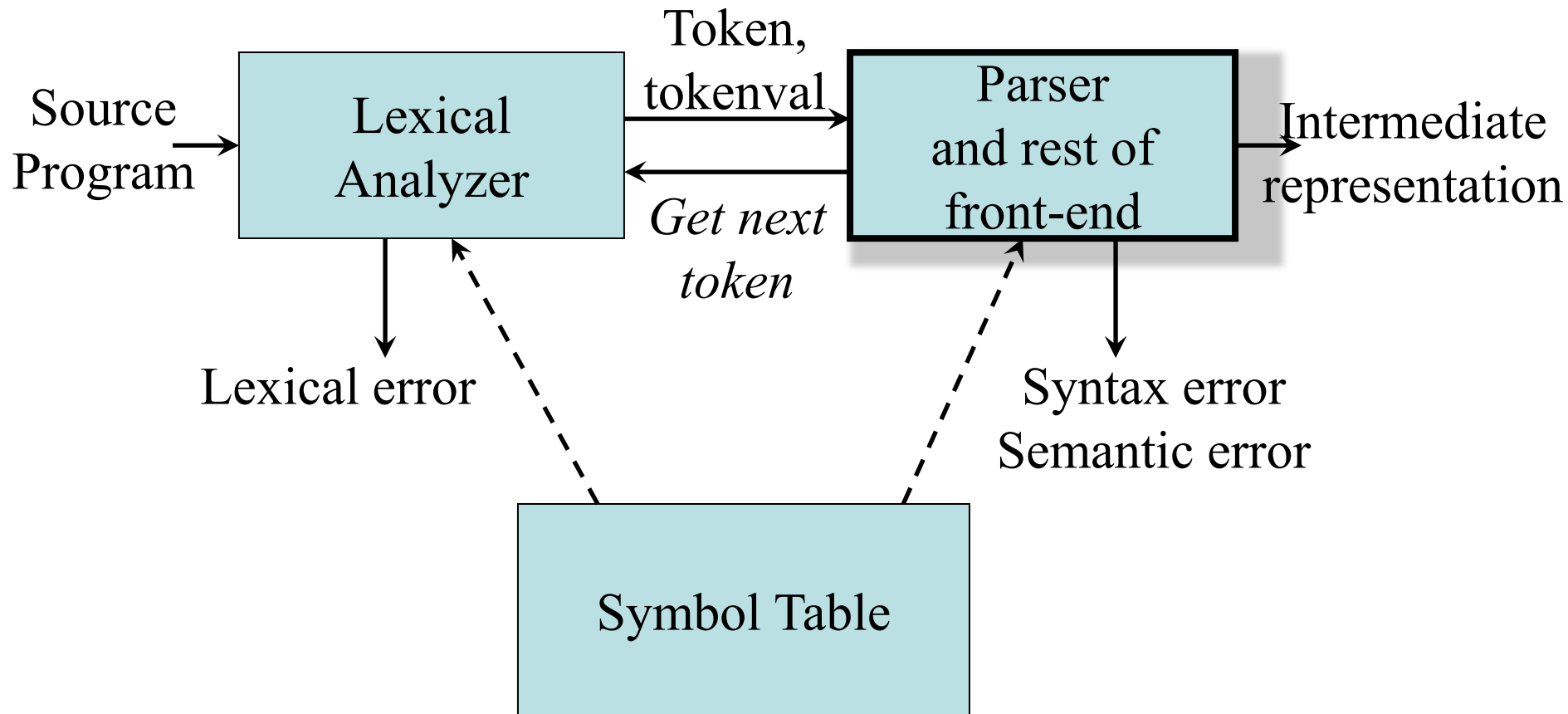


# Syntax Analysis

# Position of a Parser in the Compiler Model



# The Parser

- A parser implements a C-F grammar as a recognizer of strings
- The role of the parser in a compiler is twofold:
  1. To check syntax (= string recognizer)
    - And to report syntax errors accurately
  2. To invoke semantic actions
    - For static semantics checking, e.g. type checking of expressions, functions, etc.
    - For syntax-directed translation of the source code to an intermediate representation

# Syntax-Directed Translation

- One of the major roles of the parser is to produce an *intermediate representation* (IR) of the source program using *syntax-directed translation* methods
- Possible IR output:
  - Abstract syntax trees (ASTs)
  - Control-flow graphs (CFGs) with triples, three-address code, or register transfer list notation
  - WHIRL (SGI Pro64 compiler) has 5 IR levels!

# Error Handling

- A good compiler should assist in identifying and locating errors
  - *Lexical errors*: important, compiler can easily recover and continue
  - *Syntax errors*: most important for compiler, can almost always recover
  - *Static semantic errors*: important, can sometimes recover
  - *Dynamic semantic errors*: hard or impossible to detect at compile time, runtime checks are required
  - *Logical errors*: hard or impossible to detect

# Viable-Prefix Property

- The *viable-prefix property* of parsers allows early detection of syntax errors
  - Goal: detection of an error *as soon as possible* without further consuming unnecessary input
  - How: detect an error as soon as the prefix of the input does not match a prefix of any string in the language

Prefix { ...  
**for** ( ; )  
 ...

↓ Error is detected here

Prefix { ...  
**DO 10 I = 1; 0**  
 ...

Error is detected here ↓

# Error Recovery Strategies

- *Panic mode*
  - Discard input until a token in a set of designated synchronizing tokens is found
- *Phrase-level recovery*
  - Perform local correction on the input to repair the error
- *Error productions*
  - Augment grammar with productions for erroneous constructs
- *Global correction*
  - Choose a minimal sequence of changes to obtain a global least-cost correction

# Grammars (Recap)

- Context-free grammar is a 4-tuple  $G = (N, T, P, S)$  where
  - $T$  is a finite set of tokens (*terminal* symbols)
  - $N$  is a finite set of *nonterminals*
  - $P$  is a finite set of *productions* of the form
$$\alpha \rightarrow \beta$$
where  $\alpha \in (N \cup T)^* N (N \cup T)^*$  and  $\beta \in (N \cup T)^*$
  - $S \in N$  is a designated *start symbol*



# Notational Conventions Used

- Terminals

$$a, b, c, \dots \in T$$

specific terminals: **0**, **1**, **id**, **+**

- Nonterminals

$$A, B, C, \dots \in N$$

specific nonterminals: *expr*, *term*, *stmt*

- Grammar symbols

$$X, Y, Z \in (N \cup T)$$

- Strings of terminals

$$u, v, w, x, y, z \in T^*$$

- Strings of grammar symbols

$$\alpha, \beta, \gamma \in (N \cup T)^*$$

# Derivations (Recap)

- The *one-step derivation* is defined by
 
$$\alpha A \beta \Rightarrow \alpha \gamma \beta \quad \text{id+id}$$
 where  $A \rightarrow \gamma$  is a production in the grammar
- In addition, we define
  - $\Rightarrow$  is *leftmost*  $\Rightarrow_{lm}$  if  $\alpha$  does not contain a nonterminal
  - $\Rightarrow$  is *rightmost*  $\Rightarrow_{rm}$  if  $\beta$  does not contain a nonterminal
  - Transitive closure  $\Rightarrow^*$  (zero or more steps)
  - Positive closure  $\Rightarrow^+$  (one or more steps)
- The *language generated by  $G$*  is defined by
 
$$L(G) = \{w \in T^* \mid S \Rightarrow^+ w\}$$

# Derivation (Example)

Grammar  $G = (\{E\}, \{+, *, (, ), -, \mathbf{id}\}, P, E)$  with productions  $P =$

$$E \rightarrow E + E$$

$$| E * E$$

$$| ( E )$$

$$| - E$$

$$| \mathbf{id}$$

Example derivations:

$$E \Rightarrow - E \Rightarrow - \mathbf{id}$$

$$E \Rightarrow_{rm} E + E \Rightarrow_{rm} E + \mathbf{id} \Rightarrow_{rm} \mathbf{id} + \mathbf{id}$$

$$E \Rightarrow^* E$$

$$E \Rightarrow^* \mathbf{id} + \mathbf{id}$$

$$E \Rightarrow_{lm} \Rightarrow E * E \Rightarrow \mathbf{id} * E \Rightarrow \mathbf{id} * E + E \Rightarrow \mathbf{id} * \mathbf{id} + E \Rightarrow \mathbf{id} * \mathbf{id} + \mathbf{id}$$

# Chomsky Hierarchy: Language Classification

- A grammar  $G$  is said to be
  - *Regular* if it is *right linear* where each production is of the form
 
$$A \rightarrow w B \quad \text{or} \quad A \rightarrow w$$
 or *left linear* where each production is of the form
 
$$A \rightarrow B w \quad \text{or} \quad A \rightarrow w$$
  - *Context free* if each production is of the form
 
$$A \rightarrow \alpha$$
 where  $A \in N$  and  $\alpha \in (N \cup T)^*$
  - *Context sensitive* if each production is of the form
 
$$\alpha A \beta \rightarrow \alpha \gamma \beta$$
 where  $A \in N$ ,  $\alpha, \gamma, \beta \in (N \cup T)^*$ ,  $|\gamma| > 0$
  - *Unrestricted*

# Chomsky Hierarchy

$$\mathbb{L}(\textit{regular}) \subset \mathbb{L}(\textit{context free}) \subset \mathbb{L}(\textit{context sensitive}) \subset \mathbb{L}(\textit{unrestricted})$$

Where  $\mathbb{L}(T) = \{ L(G) \mid G \text{ is of type } T \}$

That is: the set of all languages  
generated by grammars  $G$  of type  $T$

Examples:

Every *finite language* is regular! (construct a FSA for strings in  $L(G)$ )

$L_1 = \{ \mathbf{a}^n \mathbf{b}^n \mid n \geq 1 \}$  is context free

$L_2 = \{ \mathbf{a}^n \mathbf{b}^n \mathbf{c}^n \mid n \geq 1 \}$  is context sensitive

# Parsing

- *Universal* (any C-F grammar)
  - Cocke-Younger-Kasimi
  - Earley
- *Top-down* (C-F grammar with restrictions)
  - Recursive descent (predictive parsing)
  - LL (Left-to-right, Leftmost derivation) methods
- *Bottom-up* (C-F grammar with restrictions)
  - Operator precedence parsing
  - LR (Left-to-right, Rightmost derivation) methods
    - SLR, canonical LR, LALR

# Top-Down Parsing

- LL methods (Left-to-right, Leftmost derivation) and recursive-descent parsing

Grammar:

$$E \rightarrow T + T$$

$$T \rightarrow ( E )$$

$$T \rightarrow - E$$

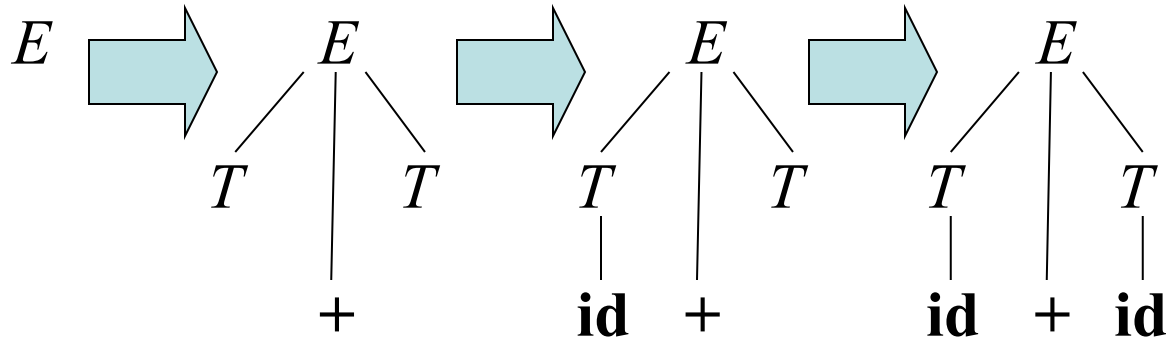
$$T \rightarrow \mathbf{id}$$

Leftmost derivation:

$$E \Rightarrow_{lm} T + T$$

$$\Rightarrow_{lm} \mathbf{id} + T$$

$$\Rightarrow_{lm} \mathbf{id} + \mathbf{id}$$



# Left Recursion (Recap)

- Productions of the form

$$\begin{array}{c} A \rightarrow A \alpha \\ | \beta \\ | \gamma \end{array}$$

are left recursive

- When one of the productions in a grammar is left recursive then a predictive parser loops forever on certain inputs



# A General Systematic Left Recursion Elimination Method

*Input: Grammar  $G$  with no cycles or  $\varepsilon$ -productions*

Arrange the nonterminals in some order  $A_1, A_2, \dots, A_n$

**for**  $i = 1, \dots, n$  **do**

**for**  $j = 1, \dots, i-1$  **do**

        replace each

$$A_i \rightarrow A_j \gamma$$

    with

$$A_i \rightarrow \delta_1 \gamma \mid \delta_2 \gamma \mid \dots \mid \delta_k \gamma$$

    where

$$A_j \rightarrow \delta_1 \mid \delta_2 \mid \dots \mid \delta_k$$

**enddo**

    eliminate the *immediate left recursion* in  $A_i$

**enddo**

# Immediate Left-Recursion Elimination

Rewrite every left-recursive production

$$\begin{array}{lcl}
 A \rightarrow A \alpha & A \rightarrow S \alpha \mid \gamma \\
 \mid A \delta & S \rightarrow A \beta \\
 \mid \beta \\
 \mid \gamma & S \rightarrow S \alpha \beta \mid \gamma \beta
 \end{array}$$

into a right-recursive production:

$$\begin{array}{lcl}
 A \rightarrow \beta A_R & S \rightarrow \gamma \beta S' \\
 \mid \gamma A_R \\
 A_R \rightarrow \alpha A_R & S' \rightarrow \alpha \beta S' \mid \varepsilon \\
 \mid \delta A_R \\
 \mid \varepsilon
 \end{array}$$

# Example Left Recursion Elim.

$$\left. \begin{array}{l} A \rightarrow B C \mid \mathbf{a} \\ B \rightarrow C A \mid A \mathbf{b} \\ C \rightarrow A B \mid C C \mid \mathbf{a} \end{array} \right\} \text{Choose arrangement: } A, B, C$$

$i = 1$ : nothing to do

$$\begin{aligned} i = 2, j = 1: & \quad B \rightarrow C A \mid \underline{A} \mathbf{b} \\ \Rightarrow & \quad B \rightarrow C A \mid \underline{B C} \mathbf{b} \mid \underline{a} \mathbf{b} \\ \Rightarrow_{(\text{imm})} & \quad B \rightarrow C A B_R \mid \mathbf{a} \mathbf{b} B_R \\ & \quad B_R \rightarrow C \mathbf{b} B_R \mid \varepsilon \end{aligned}$$

$$\begin{aligned} i = 3, j = 1: & \quad C \rightarrow \underline{A} B \mid C C \mid \mathbf{a} \\ \Rightarrow & \quad C \rightarrow \underline{B C} B \mid \underline{a} B \mid C C \mid \mathbf{a} \end{aligned}$$

$$\begin{aligned} i = 3, j = 2: & \quad C \rightarrow \underline{B} C B \mid \mathbf{a} B \mid C C \mid \mathbf{a} \\ \Rightarrow & \quad C \rightarrow \underline{C A B_R} C B \mid \underline{\mathbf{a} \mathbf{b} B_R} C B \mid \mathbf{a} B \mid \underline{C} C \mid \mathbf{a} \\ \Rightarrow_{(\text{imm})} & \quad C \rightarrow \mathbf{a} \mathbf{b} B_R C B C_R \mid \mathbf{a} B C_R \mid \mathbf{a} C_R \\ & \quad C_R \rightarrow A B_R C B C_R \mid C C_R \mid \varepsilon \end{aligned}$$

# Left Factoring

- When a nonterminal has two or more productions whose right-hand sides start with the same grammar symbols, the grammar is not LL(1) and cannot be used for predictive parsing
- Replace productions

$$A \rightarrow \alpha \beta_1 \mid \alpha \beta_2 \mid \dots \mid \alpha \beta_n \mid \gamma$$

with

$$A \rightarrow \alpha A_R \mid \gamma$$

$$A_R \rightarrow \beta_1 \mid \beta_2 \mid \dots \mid \beta_n$$

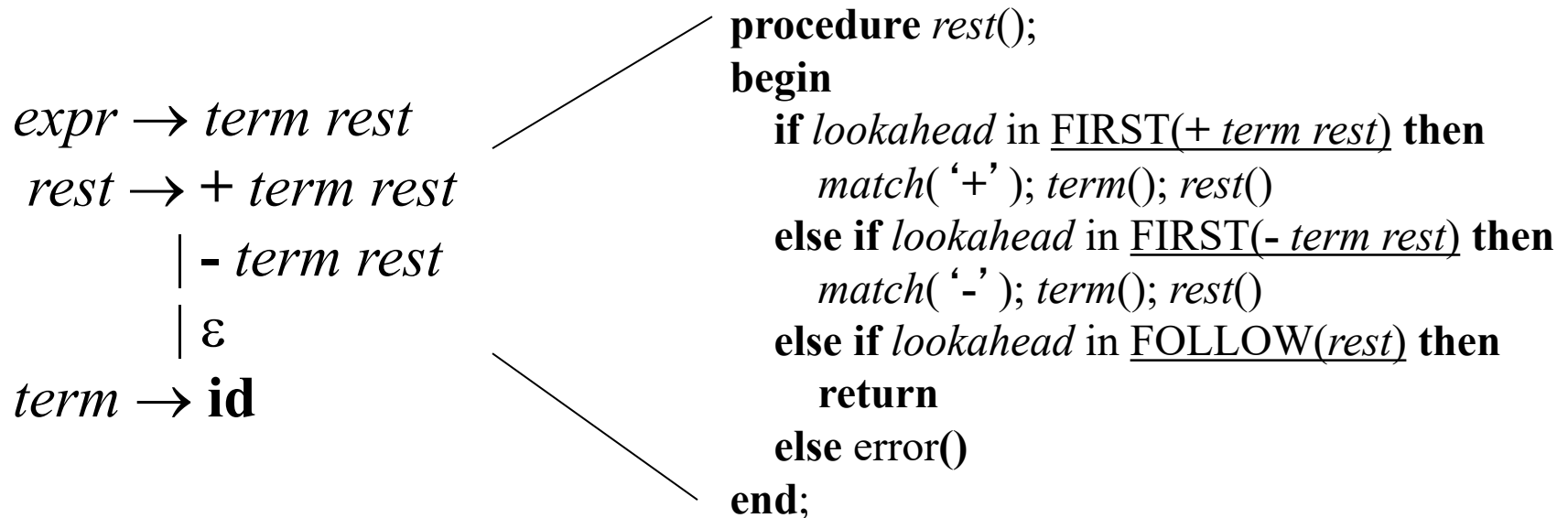
# Predictive Parsing

- Eliminate left recursion from grammar
- Left factor the grammar
- Compute FIRST and FOLLOW
- Two variants:
  - Recursive (recursive-descent parsing)
  - Non-recursive (table-driven parsing)

# Recursive-Descent Parsing (Recap)

- Grammar must be LL(1)
- Every nonterminal has one (recursive) procedure responsible for parsing the nonterminal's syntactic category of input tokens
- When a nonterminal has multiple productions, each production is implemented in a branch of a selection statement based on input look-ahead information

# Using FIRST and FOLLOW in a Recursive-Descent Parser



where  $\text{FIRST}(+ \text{ term rest}) = \{ + \}$   
 $\text{FIRST}(- \text{ term rest}) = \{ - \}$   
 $\text{FOLLOW}(\text{rest}) = \{ \$ \}$

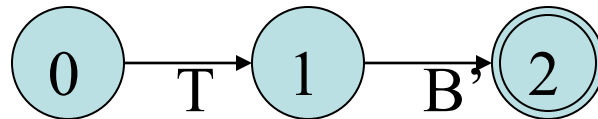
$B \rightarrow B \text{ or } T \mid T$

$T \rightarrow T \text{ and } F \mid F$

$F \rightarrow \text{not } B \mid (B) \mid \text{true} \mid \text{false} \mid \text{id}$

(id and true)

$B \rightarrow TB'$



$B' \rightarrow \text{or } TB' \mid \varepsilon$

$T \rightarrow FT'$

$T' \rightarrow \text{and } FT' \mid \varepsilon$

$F \rightarrow \text{not } B \mid (B) \mid \text{true} \mid \text{false} \mid \text{id}$



$S \rightarrow aBb$

$B \rightarrow cd|d$

$S() \{$

Match a;

$B();$

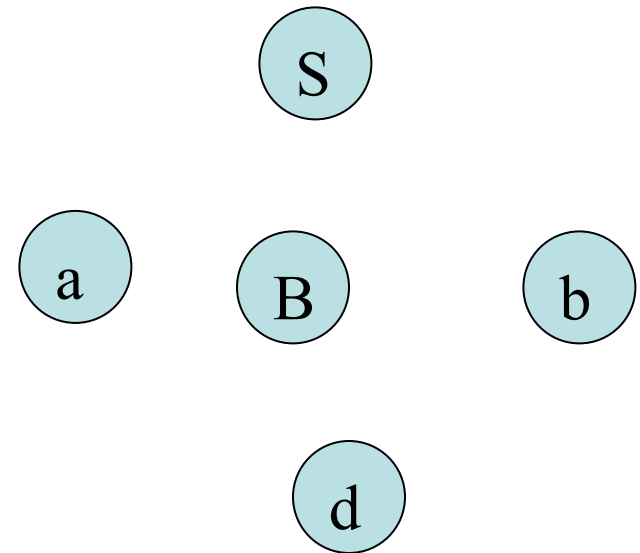
Match b;

$\}$

adb

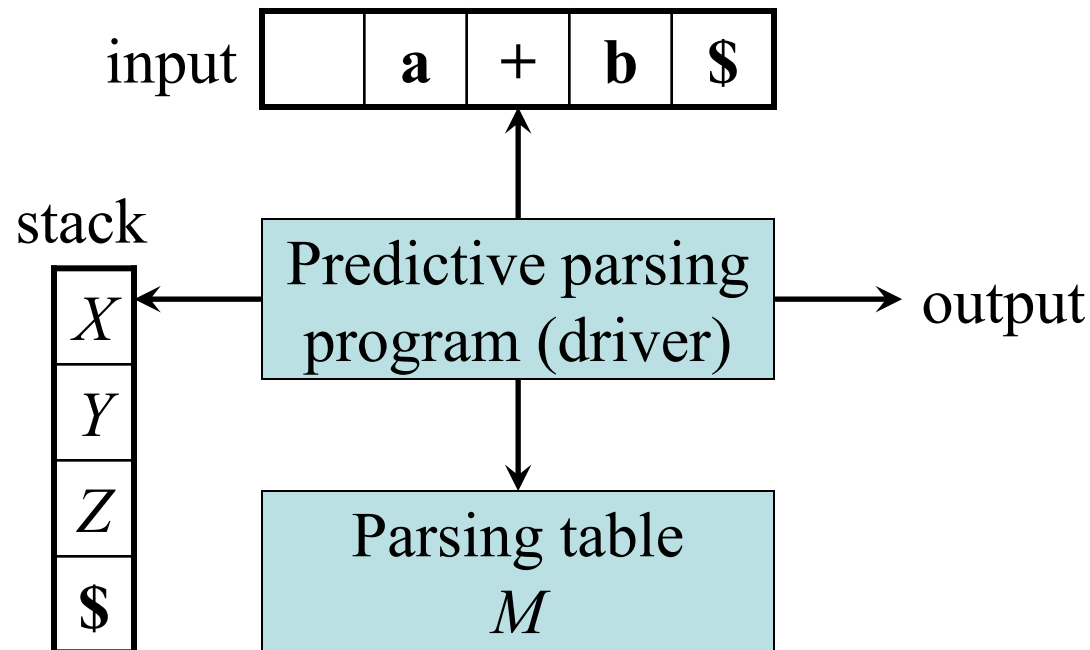
$B()$

(match c && match d) || match d;



# Non-Recursive Predictive Parsing: Table-Driven Parsing

- Given an LL(1) grammar  $G = (N, T, P, S)$  construct a table  $M[A, a]$  for  $A \in N, a \in T$  and use a *driver program* with a *stack*



# FIRST (Revisited)

- $\text{FIRST}(\alpha) = \{ \text{the set of terminals that begin all strings derived from } \alpha \}$

$$\text{FIRST}(a) = \{a\} \quad \text{if } a \in T$$

$$\text{FIRST}(\varepsilon) = \{\varepsilon\}$$

$$\text{FIRST}(A) = \cup_{A \rightarrow \alpha} \text{FIRST}(\alpha) \quad \text{for } A \rightarrow \alpha \in P$$

$$\text{FIRST}(X_1X_2\dots X_k) =$$

**if** for all  $j = 1, \dots, i-1 : \varepsilon \in \text{FIRST}(X_j)$  **then**

add non- $\varepsilon$  in  $\text{FIRST}(X_i)$  to  $\text{FIRST}(X_1X_2\dots X_k)$

**if** for all  $j = 1, \dots, k : \varepsilon \in \text{FIRST}(X_j)$  **then**

add  $\varepsilon$  to  $\text{FIRST}(X_1X_2\dots X_k)$

$$E \rightarrow E+T \mid T$$
$$T \rightarrow T * F \mid F$$
$$F \rightarrow (E) \mid \text{id}$$
$$E \rightarrow TE'$$
$$E' \rightarrow +TE' \mid \varepsilon$$
$$T \rightarrow FT'$$
$$T' \rightarrow *FT' \mid \varepsilon$$
$$F \rightarrow (E) \mid \text{id}$$
$$\text{First}(E') = +, \varepsilon$$
$$\text{First}(T') = *, \varepsilon$$
$$\text{First}(E) = \text{First}(T) = \text{First}(F) = (, \text{id}$$

$S \rightarrow ABCDE$

$A \rightarrow Ba \mid \varepsilon \quad A \rightarrow a \mid \varepsilon \quad (\text{coz } B \rightarrow \varepsilon)$

$B \rightarrow b \mid \varepsilon$

$C \rightarrow c$

$D \rightarrow d \mid \varepsilon$

$E \rightarrow De \mid \varepsilon \quad E \rightarrow e \mid \varepsilon \quad (\text{coz } D \rightarrow \varepsilon)$

$\text{First } (S) \{a, b, \varepsilon, c\}$

$\text{First } (A) \{a, b, \varepsilon\}$

$\text{First } (B) \{b, \varepsilon\}$

$\text{First } (C) \{c\}$

$\text{First } (D) \{d, \varepsilon\}$

$\text{First } (E) \{d, \varepsilon, e\}$

# FOLLOW

- $\text{FOLLOW}(A) = \{ \text{the set of terminals that can immediately follow nonterminal } A \}$

$\text{FOLLOW}(A) =$

- 2    **for** all  $(B \rightarrow \alpha A \beta) \in P$  **do**  
            $\text{FOLLOW}(A) = \text{FIRST}(\beta)$  except  $\{\epsilon\}$
- 3    **for** all  $(B \rightarrow A \beta) \in P$  and  $\beta \rightarrow \epsilon$  **do**  
            $\text{FOLLOW}(A) = \text{FOLLOW}(B)$
- 4    **for** all  $(B \rightarrow \alpha A) \in P$  **do**  
            $\text{FOLLOW}(A) = \text{FOLLOW}(B)$
- 1    **if**  $A$  is the start symbol  $S$  **then**  
           add  $\$$  to  $\text{FOLLOW}(A)$

$$E \rightarrow TE'$$

$$E' \rightarrow +TE' \mid \varepsilon$$

$$T \rightarrow FT'$$

$$T' \rightarrow *FT' \mid \varepsilon$$

$$F \rightarrow (E) \mid id$$

Rule	E	E'	T	T'	F
1	\$				
2	)		+		*
3			,\$)		+,),\$
4		,\$)		+,),\$	

# Find FIRST and FOLLOW

Example

$$S \rightarrow Bb \mid Cc \quad \{a, b, c, \varepsilon\} \{\$ \}$$

$$B \rightarrow aB \mid \varepsilon \quad \{a, \varepsilon\} \{b\}$$

$$C \rightarrow cC \mid \varepsilon \quad \{c, \varepsilon\} \{c\}$$

Example

$$S \rightarrow aABb \quad \{a\} \{\$ \}$$

$$A \rightarrow c \mid \varepsilon \quad \{c, \varepsilon\} \{d, \$, b\}$$

$$B \rightarrow d \mid \varepsilon \quad \{d, \varepsilon\} \{b\}$$

Example

$$S \rightarrow ACB \mid CbB \mid Ba$$

$$A \rightarrow da \mid BC$$

$$B \rightarrow g \mid \varepsilon$$

$$C \rightarrow h \mid \varepsilon$$

Example

$$S \rightarrow aBDh$$

$$B \rightarrow cC$$

$$C \rightarrow bC \mid \varepsilon$$

$$D \rightarrow EF$$

$$E \rightarrow g \mid \varepsilon$$

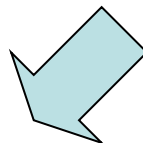
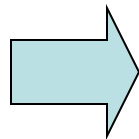


# Constructing an LL(1) Predictive Parsing Table

```
for each production  $A \rightarrow \alpha$  do  
    for each  $a \in \text{FIRST}(\alpha)$  do  
        add  $A \rightarrow \alpha$  to  $M[A, a]$   
    enddo  
    if  $\varepsilon \in \text{FIRST}(\alpha)$  then  
        for each  $b \in \text{FOLLOW}(A)$  do  
            add  $A \rightarrow \alpha$  to  $M[A, b]$   
        enddo  
    endif  
enddo  
Mark each undefined entry in  $M$  error
```

# Example Table

$E \rightarrow T E_R$   
 $E_R \rightarrow + T E_R \mid \varepsilon$   
 $T \rightarrow F T_R$   
 $T_R \rightarrow * F T_R \mid \varepsilon$   
 $F \rightarrow ( E ) \mid \mathbf{id}$



$A \rightarrow \alpha$	FIRST( $\alpha$ )	FOLLOW( $A$ )
$E \rightarrow T E_R$	( id	\$ )
$E_R \rightarrow + T E_R$	+	\$ )
$E_R \rightarrow \varepsilon$	$\varepsilon$	
$T \rightarrow F T_R$	( id	+ \$ )
$T_R \rightarrow * F T_R$	*	+ \$ )
$T_R \rightarrow \varepsilon$	$\varepsilon$	
$F \rightarrow ( E )$	(	* + \$ )
$F \rightarrow \mathbf{id}$	id	* + \$ )

	id	+	*	(	)	\$
$E$	$E \rightarrow T E_R$			$E \rightarrow T E_R$		
$E_R$		$E_R \rightarrow + T E_R$			$E_R \rightarrow \varepsilon$	$E_R \rightarrow \varepsilon$
$T$	$T \rightarrow F T_R$			$T \rightarrow F T_R$		
$T_R$		$T_R \rightarrow \varepsilon$	$T_R \rightarrow * F T_R$		$T_R \rightarrow \varepsilon$	$T_R \rightarrow \varepsilon$
$F$	$F \rightarrow \mathbf{id}$			$F \rightarrow ( E )$		

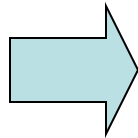
# LL(1) Grammars are Unambiguous

Ambiguous grammar

$$S \rightarrow \mathbf{i} E \mathbf{t} S S_R \mid \mathbf{a}$$

$$S_R \rightarrow \mathbf{e} S \mid \varepsilon$$

$$E \rightarrow \mathbf{b}$$



$A \rightarrow \alpha$	FIRST( $\alpha$ )	FOLLOW( $A$ )
$S \rightarrow \mathbf{i} E \mathbf{t} S S_R$	<b>i</b>	<b>e \$</b>
$S \rightarrow \mathbf{a}$	<b>a</b>	
$S_R \rightarrow \mathbf{e} S$	<b>e</b>	<b>e \$</b>
$S_R \rightarrow \varepsilon$	$\varepsilon$	
$E \rightarrow \mathbf{b}$	<b>b</b>	<b>t</b>

Error: duplicate table entry

	<b>a</b>	<b>b</b>	<b>e</b>	<b>i</b>	<b>t</b>	<b>\$</b>
$S$	$S \rightarrow \mathbf{a}$			$S \rightarrow \mathbf{i} E \mathbf{t} S S_R$		
$S_R$			$S_R \rightarrow \varepsilon$ $S_R \rightarrow \mathbf{e} S$			$S_R \rightarrow \varepsilon$
$E$		$E \rightarrow \mathbf{b}$				

# Predictive Parsing Program (Driver)

```
push($)  
push(S)  
a := lookahead  
repeat  
     $X := \text{pop}()$   
    if  $X$  is a terminal or  $X = \$$  then  
        match(X) // moves to next token and a := lookahead  
    else if  $M[X, a] = X \rightarrow Y_1 Y_2 \dots Y_k$  then  
        push( $Y_k, Y_{k-1}, \dots, Y_2, Y_1$ ) // such that  $Y_1$  is on top  
        ... invoke actions and/or produce IR output ...  
    else    error()  
    endif  
until  $X = \$$ 
```

# Example Table-Driven Parsing

$E \rightarrow T E_R$   
 $E_R \rightarrow + T E_R \mid \varepsilon$   
 $T \rightarrow F T_R$   
 $T_R \rightarrow * F T_R \mid \varepsilon$   
 $F \rightarrow ( E ) \mid \text{id}$

Stack	Input	Production applied
$\$E$	<u>id</u> +id*id\$	$E \rightarrow T E_R$
$\$E_R T$	<u>id</u> +id*id\$	$T \rightarrow F T_R$
$\$E_R T_R F$	<u>id</u> +id*id\$	$F \rightarrow \text{id}$
$\$E_R T_R \text{id}$	<u>id</u> +id*id\$	<b>advance the lookahead</b>
$\$E_R T_R$	+ <u>id</u> *id\$	$T_R \rightarrow \varepsilon$
$\$E_R$	+ <u>id</u> *id\$	$E_R \rightarrow + T E_R$
$\$E_R T +$	+ <u>id</u> *id\$	<b>advance the lookahead</b>
$\$E_R T$	<u>id</u> *id\$	$T \rightarrow F T_R$
$\$E_R T_R F$	<u>id</u> *id\$	$F \rightarrow \text{id}$
$\$E_R T_R \text{id}$	<u>id</u> *id\$	<b>advance the lookahead</b>
$\$E_R T_R$	* <u>id</u> \$	$T_R \rightarrow * F T_R$
$\$E_R T_R F *$	* <u>id</u> \$	<b>advance the lookahead</b>
$\$E_R T_R F$	<u>id</u> \$	$F \rightarrow \text{id}$
$\$E_R T_R \text{id}$	<u>id</u> \$	<b>advance the lookahead</b>
$\$E_R T_R$	\$	$T_R \rightarrow \varepsilon$
$\$E_R$	\$	$E_R \rightarrow \varepsilon$
\$	\$	Accept

# Panic Mode Recovery

Add synchronizing actions to undefined entries based on FOLLOW

Pro: Can be automated

Cons: Error messages are needed

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

$\text{FOLLOW}(E_R) = \{ ) \$ \}$

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

$\text{FOLLOW}(T_R) = \{ + ) \$ \}$

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

	id	+	*	(	)	\$
$E$	$E \rightarrow T E_R$			$E \rightarrow T E_R$	<i>synch</i>	<i>synch</i>
$E_R$		$E_R \rightarrow + T E_R$			$E_R \rightarrow \epsilon$	$E_R \rightarrow \epsilon$
$T$	$T \rightarrow F T_R$	<i>synch</i>		$T \rightarrow F T_R$	<i>synch</i>	<i>synch</i>
$T_R$		$T_R \rightarrow \epsilon$	$T_R \rightarrow * F T_R$		$T_R \rightarrow \epsilon$	$T_R \rightarrow \epsilon$
$F$	$F \rightarrow \text{id}$	<i>synch</i>	<i>synch</i>	$F \rightarrow ( E )$	<i>synch</i>	<i>synch</i>

***synch***: the driver pops current nonterminal  $A$  and skips input till *synch* token or skips input until one of  $\text{FIRST}(A)$  is found

# Phrase-Level Recovery

Change input stream by inserting missing tokens

For example: **id id** is changed into **id \* id**

Pro: Can be fully automated

Cons: Recovery not always intuitive

Can then continue here

	<b>id</b>	<b>+</b>	<b>*</b>	<b>(</b>	<b>)</b>	<b>\$</b>
$E$	$E \rightarrow T E_R$			$E \rightarrow T E_R$	<i>synch</i>	<i>synch</i>
$E_R$		$E_R \rightarrow + T E_R$			$E_R \rightarrow \epsilon$	$E_R \rightarrow \epsilon$
$T$	$T \rightarrow F T_R$	<i>synch</i>		$T \rightarrow F T_R$	<i>synch</i>	<i>synch</i>
$T_R$	<i>insert *</i>	$T_R \rightarrow \epsilon$	$T_R \rightarrow * F T_R$		$T_R \rightarrow \epsilon$	$T_R \rightarrow \epsilon$
$F$	$F \rightarrow \text{id}$	<i>synch</i>	<i>synch</i>	$F \rightarrow ( E )$	<i>synch</i>	<i>synch</i>

*insert \**: driver inserts missing **\*** and retries the production

# Error Productions

$$\begin{aligned}
 E &\rightarrow T E_R \\
 E_R &\rightarrow + T E_R \mid \varepsilon \\
 T &\rightarrow F T_R \\
 T_R &\rightarrow * F T_R \mid \varepsilon \\
 F &\rightarrow ( E ) \mid \mathbf{id}
 \end{aligned}$$

Add “*error production*”:

$$T_R \rightarrow F T_R$$

to ignore missing \*, e.g.: **id id**

Pro: Powerful recovery method

Cons: Manual addition of productions

	<b>id</b>	+	*	(	)	\$
$E$	$E \rightarrow T E_R$			$E \rightarrow T E_R$	<i>synch</i>	<i>synch</i>
$E_R$		$E_R \rightarrow + T E_R$			$E_R \rightarrow \varepsilon$	$E_R \rightarrow \varepsilon$
$T$	$T \rightarrow F T_R$	<i>synch</i>		$T \rightarrow F T_R$	<i>synch</i>	<i>synch</i>
$T_R$	$T_R \rightarrow F T_R$	$T_R \rightarrow \varepsilon$	$T_R \rightarrow * F T_R$		$T_R \rightarrow \varepsilon$	$T_R \rightarrow \varepsilon$
$F$	$F \rightarrow \mathbf{id}$	<i>synch</i>	<i>synch</i>	$F \rightarrow ( E )$	<i>synch</i>	<i>synch</i>