

# Syntax Analysis

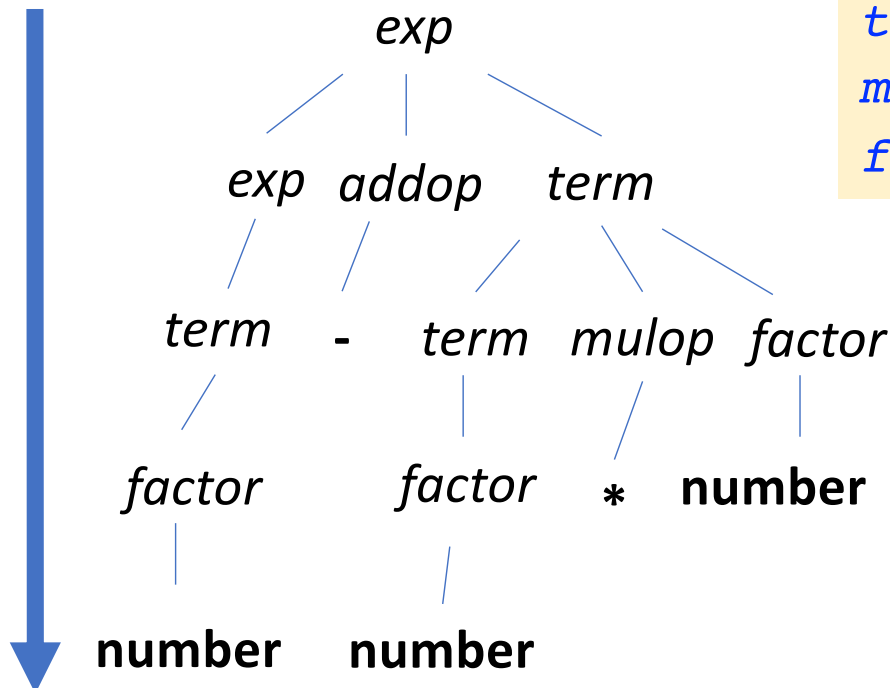
(Chapters 4 & 5)

# Top-Down Parsing (Chapter 4)

- Builds a parse tree top down, from the start nonterminal
- and creating tree nodes in **preorder** (a leftmost derivation)

34 - 3 \* 42

$exp \rightarrow exp \text{ addop } term \mid term$   
 $addop \rightarrow + \mid -$   
 $term \rightarrow term \text{ mulop } factor \mid factor$   
 $mulop \rightarrow *$   
 $factor \rightarrow ( exp ) \mid \mathbf{number}$

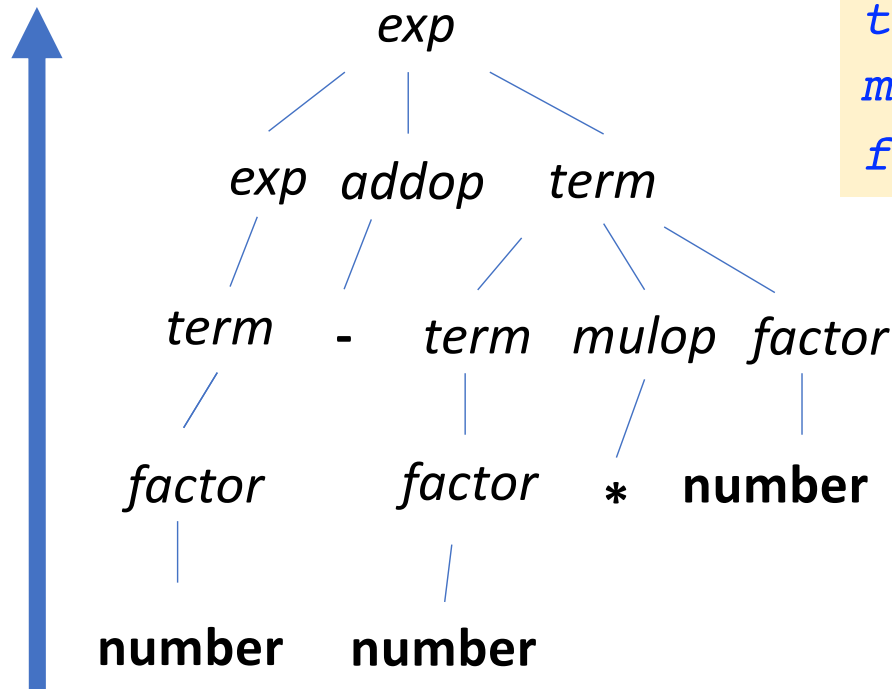


# Bottom-Up Parsing (Chapter 5)

- Builds a parse tree bottom up, from the leaf nodes

34 - 3 \* 42

$exp \rightarrow exp \text{ addop } term \mid term$   
 $addop \rightarrow + \mid -$   
 $term \rightarrow term \text{ mulop } factor \mid factor$   
 $mulop \rightarrow *$   
 $factor \rightarrow ( exp ) \mid \mathbf{number}$



# Top-Down Parsing

(Chapter 4)

# Top-Down Parsing

- Backtracking parsers

More Powerful

Exponential Time

- try different possibilities
- back up arbitrary number of input symbols once a try fails

- Predictive parsers

- use one or more lookahead symbols to narrow down the possibilities

34 - 3 \* 42

```
exp → exp addop term | term  
addop → + | -  
term → term mulop factor | factor  
mulop → *  
factor → ( exp ) | number
```

# Top-Down Parsing (Predictive)

- Recursive-descent parsing
  - versatile
  - better for a hand-written parser
- LL(1) parsing
  - scan from **left to right**, and perform **leftmost** derivation
  - look ahead at most **one** input symbol

34 - 3 \* 42

```
exp → exp addop term | term  
addop → + | -  
term → term mulop factor | factor  
mulop → *  
factor → ( exp ) | number
```

# Recursive Descent Parsing

# Recursive-Descent Parsing

- Basic Ideas

- for each nonterminal, define a function to recognize it

```
factor()  
{  
  switch(token)  
  case (:  
    match()  
    exp()  
    match()  
  case number:  
    match(number)  
  default:  
    error()  
}
```

Lookahead

Match and consume the symbol

*exp* → *exp addop term* | *term*

*addop* → + | -

*term* → *term mulop factor* | *factor*

*mulop* → \*

*factor* → ( *exp* ) | **number**



# Recursive-Descent Parsing

- Basic Ideas

- for each nonterminal, define a function to recognize it

```
match(expectedToken)
{
  if (token == expectedToken)
    token = getToken()
  else
    error()
}
```



advance the input

# Recursive-Descent Parsing

- Basic Ideas

- for each nonterminal, define a function to recognize it

```
factor()  
{  
  switch(token)  
  case (:  
    match()  
    exp()  
    match()  
  case number:  
    match(number)  
  otherwise  
    error()  
}
```

factor()  
↓  
exp()  
↓  
term()  
↓  
factor()

**Recursive &  
Descent**

# Recursive-Descent Parsing

- Exercise

- Write down the pseudocode for recognizing `if-stmt`

*if-stmt*  $\rightarrow$  **if** (*exp*) *stmt* **else** *stmt*

```
factor()  
{  
  switch(token)  
  case (:  
    match((  
      exp()  
      match()  
    case number:  
      match(number)  
    otherwise  
      error()  
  }  
}
```

*exp*  $\rightarrow$  *exp* *addop* *term* | *term*

*addop*  $\rightarrow$  + | -

*term*  $\rightarrow$  *term* *mulop* *factor* | *factor*

*mulop*  $\rightarrow$  \*

*factor*  $\rightarrow$  ( *exp* ) | **number**

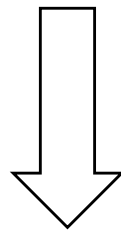
# Recursive-Descent Parsing

- EBNF

- extended BNF

```
if-stmt → if (exp) stmt  
| if (exp) stmt else stmt
```

rewrite



```
if-stmt → if (exp) stmt [else stmt]
```

means "optional"

```
ifStmt()  
{  
    match(if)  
    match()  
    exp()  
    match()  
    stmt()  
    if(token == else)  
        match(else)  
        stmt()  
}
```

# Recursive-Descent Parsing

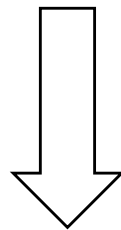
- EBNF

- extended BNF

left recursion

$exp \rightarrow exp \text{ addop } term$   
 $| term$

rewrite



$exp \rightarrow term \{ \text{addop } term \}$

means "repetition"

```
exp()  
{  
  term()  
  while(token == + or -)  
  {  
    match(token)  
    term()  
  }  
}
```

no "addop" call

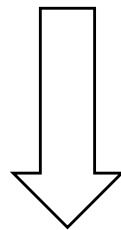
# Recursive-Descent Parsing

- EBNF

- extended BNF

*term* → *term mulop factor*  
| *factor*

rewrite



*term* → *factor { mulop factor }*

```
term()  
{  
  factor()  
  while(token == *)  
  {  
    match(token)  
    factor()  
  }  
}
```



# Recursive-Descent Parsing

- Calculation can be embedded in parsing
- Preserve left associativity

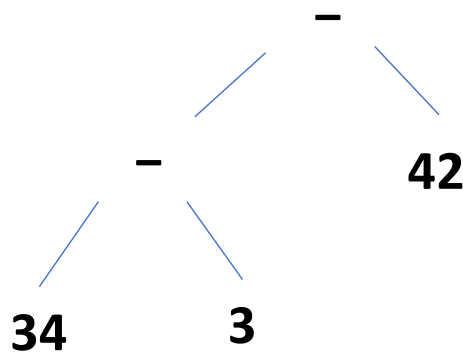
*exp* → *term* { *addop term* }

```
int exp()
{
    int temp = term()
    while(token == + or -) {
        switch(token)
            case + :
                match(+)
                temp += term()
            case - :
                match(-)
                temp -= term()
        }
    }
    return temp
}
```

# Recursive-Descent Parsing

- Tree construction can be embedded in parsing
- Example for generating a AST

*exp* → *term* { *addop term* }



34 - 3 - 42

```
treeNode exp()  
{  
    treeNode node = term()  
    while(token == + or -)  
    {  
        treeNode newnode  
        newnode = makeOpNode(token)  
        match(token)  
        newnode.leftChild = node  
        newnode.rightChild = term()  
        node = newnode  
    }  
    return node  
}
```



# LL(1) Parsing

# LL(1) Parsing

- Use a stack rather than recursive calls to build a tree
- Similar to running some pushdown automaton (PDA)
  - Begin by pushing the start nonterminal to the stack
  - Perform some actions based on the stack and next input symbol
  - Accept if both stack and input become empty

E.g.

tokens    grammar

( )

$S \rightarrow ( S ) S \mid \epsilon$

# LL(1) Parsing

- Example

<u>tokens</u>	<u>grammar</u>
( )	$S \rightarrow ( S ) S \mid \epsilon$

**\$: marks  
stack bottom**

**stack top:  
leftmost of RHS**

	Parsing Stack	Input	Action
1	\$ S	( ) \$	$S \rightarrow ( S ) S$
2	\$ S ) S (	( ) \$	match
3	\$ S ) S	) \$	$S \rightarrow \epsilon$
4	\$ S )	) \$	match
5	\$ S	\$	$S \rightarrow \epsilon$
6	\$	\$	match

# LL(1) Parsing

- Two Actions:

- If stack top is a nonterminal  $A$  and  $A \rightarrow \alpha$ , replace  $A$  with  $\alpha$  (**generate**)
- If stack top is a terminal (token), **match** it with input token
  - If matched, pop stack and advance input
  - Otherwise, throw an error

Error Input:

)

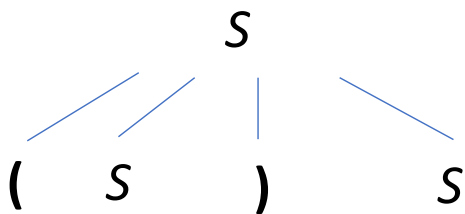
	Parsing Stack	Input	Action
1	$\$ S$	) $\$$	$S \rightarrow ( S ) S$
2	$\$ S ) S ($	) $\$$	mismatch

	Parsing Stack	Input	Action
1	$\$ S$	) $\$$	$S \rightarrow \epsilon$
2	$\$$	) $\$$	mismatch

# LL(1) Parsing

- Parse Tree Construction

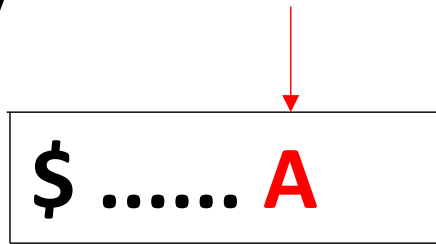
- root node is constructed at the beginning of the parse
- construct and attach tree nodes in each **generate** action



	Parsing Stack	Input	Action
1	\$ S	( ) \$	$S \rightarrow ( S ) S$
2	\$ S ) S (	( ) \$	match

# First and Follow Sets

# Why First ?



Stack



Tokens

**A** → **a B** | ~~**X**~~ **C** | **c D**

**X** → **Y y** | **Z z**

**Z** → **b b** | **z z**

# First Set: Definition

- Suppose  $\alpha$  is a string of terminals and nonterminals, **First( $\alpha$ )** consists of the first terminals that can be derived from  $\alpha$ .

if  $\alpha \Rightarrow^* a\beta$ , then  $a \in \text{First}(\alpha)$

if  $\alpha \Rightarrow^* \varepsilon$  (nullable), then  $\varepsilon \in \text{First}(\alpha)$

E.g.

**First(ABC)** = {a, c, d}

**First(BC)** = {b, c,  $\varepsilon$ }

## Another grammar

$A \rightarrow aB \mid CD$

$B \rightarrow Bb \mid \varepsilon$

$C \rightarrow c \mid \varepsilon$

$D \rightarrow d$



# First Set: Properties

1. If  $X$  is a terminal or  $\epsilon$ , then  $\text{First}(X) = \{X\}$
2. Suppose  $X$  is a nonterminal and  $X \rightarrow Y_1Y_2\dots Y_k$ 
  - if for some  $i$ ,  $Y_1\dots Y_{i-1} \Rightarrow^* \epsilon$ , then  $\text{First}(X) \supseteq \text{First}(Y_i) - \{\epsilon\}$
  - if  $Y_1\dots Y_k \Rightarrow^* \epsilon$ , then  $\epsilon \in \text{First}(X)$

Why exclude it ?

E.g.

$\text{First}(A) = \{a, c, d\}$

## Another grammar

$A \rightarrow aB \mid CD$

$B \rightarrow Bb \mid \epsilon$

$C \rightarrow c \mid \epsilon$

$D \rightarrow d$

# First Set: Algorithm

- Compute the First set for each nonterminal iteratively

```
for each nonterminal A
  First(A) = {}
while some First set changed
  for each A  $\rightarrow$   $X_1X_2\dots X_n$ 
    k = 1
    continue = true
    while continue == true and k<=n
      add First( $X_k$ ) -  $\{\epsilon\}$  to First(A)
      if  $\epsilon \notin$  First( $X_k$ )
        continue = false
      k++
    if continue == true
      add  $\epsilon$  to First(A)
```

Why iterative ?

# First Set: Algorithm

	A	B	C	D
init	{}	{}	{}	{}
round-1	{a}	{b}	{ $\epsilon$ }	{d, $\epsilon$ }
round-2	{a, $\epsilon$ , d}	{b}	{ $\epsilon$ }	{d, $\epsilon$ }
round-3	{a, $\epsilon$ , d}	{b}	{ $\epsilon$ }	{d, $\epsilon$ }

The same results, so iteration stops

$A \rightarrow aB \mid CD$   
 $B \rightarrow bC$   
 $C \rightarrow \epsilon$   
 $D \rightarrow d \mid \epsilon$

# Why Follow ?

$\$ \dots b A$

Stack

$b \dots$

Tokens

$A \rightarrow a B \mid \epsilon \mid c D$



if  $b$  can Follow  $A$

# Follow Set: Definition

- For a nonterminal **A**, if there exists a derivation from the start nonterminal  $S \Rightarrow^* \alpha A a \beta$ , then  $a \in \text{Follow}(A)$
- If  $S \Rightarrow^* \alpha A$ , then  $\$ \in \text{Follow}(A)$

E.g.

**\$ always in Follow set of start symbol**

**Follow(A) = {\$}**

**Follow(B) = {d}**

**Follow(C) = {b, \$}**

**Follow(D) = {e, \$}**

grammar

$A \rightarrow aBD \mid CC$

$B \rightarrow a$

$C \rightarrow bDe$

$D \rightarrow d$

# Follow Set: Definition

- For a nonterminal **A**, if there exists a derivation from the start nonterminal  $S \Rightarrow^* \alpha A a \beta$  ( $a \neq \epsilon$ ), then  $a \in \text{Follow}(A)$
- If  $S \Rightarrow^* \alpha A$ , then  $\$ \in \text{Follow}(A)$

**Exercise:**

**Follow(A) = { \$ }**

**Follow(B) = { d, \$ }**

**Follow(C) = { d, e, \$ }**

**Follow(D) = { e, \$ }**

Another grammar

$A \rightarrow aBD \mid CC$

$B \rightarrow a$

$C \rightarrow De$

$D \rightarrow d \mid \epsilon$

# Follow Set: Properties

1. If  $A$  is the start symbol,  $\$ \in \text{Follow}(A)$
2. For any nonterminal  $A$ ,  $\epsilon \notin \text{Follow}(A)$
3. If  $A \rightarrow \alpha B \gamma$  then  $\text{First}(\gamma) - \{\epsilon\} \subseteq \text{Follow}(B)$
4. If  $A \rightarrow \alpha B \gamma$  and  $\gamma \Rightarrow^* \epsilon$  then  $\text{Follow}(A) \subseteq \text{Follow}(B)$

# Follow Set: Algorithm

- Compute the Follow set for each nonterminal iteratively

```
for each nonterminal A
  if A is start-symbol
    Follow(A) = {$}
  else
    Follow(A) = {}
while some Follow set changed
  for each A  $\rightarrow$   $X_1X_2\dots X_n$ 
    for each  $X_i$  that is a nonterminal
      add First( $X_{i+1}X_{i+2}\dots X_n$ ) -  $\{\epsilon\}$  to Follow( $X_i$ )
      if  $\epsilon \in$  First( $X_{i+1}X_{i+2}\dots X_n$ )
        add Follow(A) to Follow( $X_i$ )
```

3rd property

4th property



# Follow Set: Algorithm

First	S	A	B	C
	{e}	{e}	{b}	{e}

Follow	S	A	B	C
init	{\$}	{}	{}	{}
round-1	{\$}	{b, a}	{\$}	{b, a}
round-2	{\$}	{b, a}	{\$}	{b, a}

The same results, so iteration stops

$S \rightarrow AB$   
 $A \rightarrow eC$   
 $B \rightarrow bAa$   
 $C \rightarrow e$

# Example: Compute First/Follow Set

```

$$E \rightarrow T E'$$

$$E' \rightarrow \text{addop } T E' \mid \varepsilon$$

$$\text{addop} \rightarrow + \mid -$$

$$T \rightarrow F T'$$

$$T' \rightarrow \text{mulop } F T' \mid \varepsilon$$

$$\text{mulop} \rightarrow *$$

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

```

First(E) = { (, id }

First(E') = { +, -,  $\varepsilon$  }

First(addop) = { +, - }

First(T) = { (, id }

First(T') = { \*,  $\varepsilon$  }

First(mulop) = { \* }

First(F) = { (, id }

Follow(E) = { \$, ) }

Follow(E') = { \$, ) }

Follow(addop) = { (, id }

Follow(T) = { \$, ), +, - }

Follow(T') = { \$, ), +, - }

Follow(mulop) = { (, id }

Follow(F) = { \$, ), +, -, \* }

# Example:

$E \rightarrow T E'$   
 $E' \rightarrow + T E' \mid \epsilon$   
 $T \rightarrow F T'$   
 $T' \rightarrow * F T' \mid \epsilon$   
 $F \rightarrow ( E ) \mid id$

	FIRST
E	{ (, id }
E'	{ +, $\epsilon$ }
T	{ (, id }
T'	{ *, $\epsilon$ }
F	{ (, id }

$F \rightarrow ( E ) \mid id$

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

$T' \rightarrow * F T' \mid \epsilon$

$FIRST(T') = \{ *, \epsilon \}$

$E' \rightarrow + T E' \mid \epsilon$

$FIRST(E') = \{ +, \epsilon \}$

$T \rightarrow F T'$

$FIRST(T) = FIRST(F) = \{ (, id \}$

$E \rightarrow T E'$

$FIRST(E) = FIRST(T) = \{ (, id \}$

# Example:

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

## FOLLOW(E)

E is the start symbol

\$  $\varepsilon$  FOLLOW(E)

F  $\rightarrow$  ( E )

)  $\varepsilon$  FOLLOW(E)

FOLLOW(E) = { \$, ) }

## FOLLOW(E')

$E \rightarrow T E'$  &  $E' \rightarrow + T E'$

.....E ■ .....  $\rightarrow$  .....TE' ■ .....

- FOLLOW(E) is contained in FOLLOW(E')

- { \$, ) } is contained in FOLLOW(E')

FOLLOW(E') = { \$, ) }

# Example:

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

	FIRST
E	{ (, id }
E'	{ +, $\varepsilon$ }
T	{ (, id }
T'	{ *, $\varepsilon$ }
F	{ (, id }

**FOLLOW(T)**

$E \rightarrow T E'$  &  $E' \rightarrow + T E'$

FIRST(E') – {  $\varepsilon$  } is contained in FOLLOW(T)

→ { + } is contained in FOLLOW(T)

$\varepsilon$  belongs to FIRST(E')

→ FOLLOW(E) is contained in FOLLOW(T)

→ { \$, ) } is contained in FOLLOW(T)

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

# Example:

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

	FIRST
E	{ (, id }
E'	{ +, $\varepsilon$ }
T	{ (, id }
T'	{ *, $\varepsilon$ }
F	{ (, id }

**FOLLOW(T')**

$T \rightarrow F T'$  &  $T' \rightarrow * F T'$

FOLLOW(T) is contained in FOLLOW(T')

$\rightarrow \{ +, \$, ) \}$  is contained in FOLLOW(T')

**FOLLOW(T') = { +, \$, ) }**

# Example:

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

	FIRST
E	{ (, id }
E'	{ +, $\epsilon$ }
T	{ (, id }
T'	{ *, $\epsilon$ }
F	{ (, id }

**FOLLOW(F)**

$T \rightarrow \mathbf{F} T' \quad \& \quad T' \rightarrow * \mathbf{F} T'$

FIRST(T') – {  $\epsilon$  } is contained in FOLLOW(F)

→ { \* } is contained in FOLLOW(F)

$\epsilon$  belongs to FIRST(T')

→ FOLLOW(T) is contained in FOLLOW(F)

→ { +, \$, ) } is contained in FOLLOW(F)

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

# Example:

$E \rightarrow TE'$   
 $E' \rightarrow +TE' \mid \varepsilon$   
 $T \rightarrow FT'$   
 $T' \rightarrow *FT' \mid \varepsilon$   
 $F \rightarrow (E) \mid id$

	FIRST
E	{ (, id }
E'	{ +, $\varepsilon$ }
T	{ (, id }
T'	{ *, $\varepsilon$ }
F	{ (, id }

	FOLLOW
E	{ \$, ) }
E'	{ \$, ) }
T	{ +, \$, ) }
T'	{ +, \$, ) }
F	{ *, +, \$, ) }



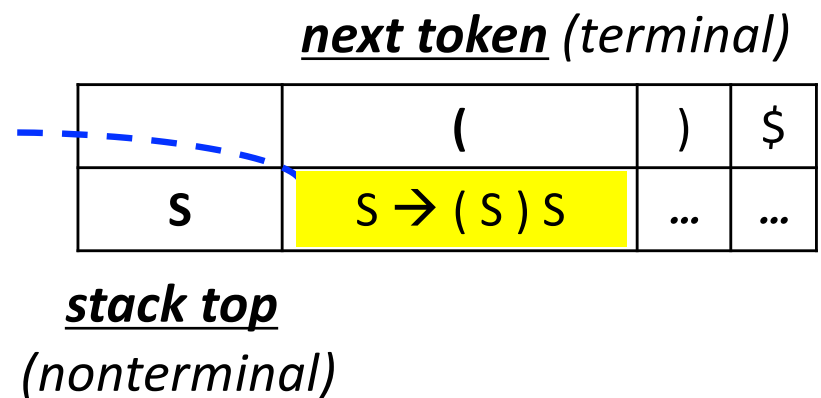
Back to LL(1) Parsing

# LL(1) Parsing

- Parsing Table

- if the stack top is **N**, and the lookahead token is **T**, then entry **[N, T]** in the table is the production rule to use

	Parsing Stack	Input	Action
1	...	...	...
2	...	...	...
3	\$ ... S	(...\$	$S \rightarrow (S)S$
4	...	...	...
5	...	...	...



# LL(1) Parsing

- Parsing Table Construction

Given  $A \rightarrow \alpha$

- for each token  $a$  in  $\text{First}(\alpha)$ , add  $A \rightarrow \alpha$  to the entry  $[A, a]$
- if  $\epsilon \in \text{First}(\alpha)$ , for each  $a$  in  $\text{Follow}(A)$ , add  $A \rightarrow \alpha$  to entry  $[A, a]$

$S \rightarrow ( S ) S$

$S \rightarrow \epsilon$

$\text{First}(( S ) S) = \{ ( \}$

$\text{First}(\epsilon) = \{ \epsilon \}$     $\text{Follow}(S) = \{ ), \$ \}$

	(	)	\$
S	$S \rightarrow ( S ) S$	$S \rightarrow \epsilon$	$S \rightarrow \epsilon$

# LL(1) Parsing

- Parsing Table Construction

Given  $A \rightarrow \alpha$

- for each token  $a$  in  $\text{First}(\alpha)$ , add  $A \rightarrow \alpha$  to the entry  $[A, a]$
- if  $\epsilon \in \text{First}(\alpha)$ , for each  $a$  in  $\text{Follow}(A)$ , add  $A \rightarrow \alpha$  to entry  $[A, a]$

## Exercise:

$S \rightarrow A$   
 $A \rightarrow ( A ) A$   
 $A \rightarrow \epsilon$

	(	)	\$
S	$S \rightarrow A$		$S \rightarrow A$
A	$A \rightarrow ( A ) A$	$A \rightarrow \epsilon$	$A \rightarrow \epsilon$

# LL(1) Parsing

- Parsing Table Construction: Example

r1  $E \rightarrow T E'$   
r2  $E' \rightarrow \text{addop } T E'$   
r3  $E' \rightarrow \epsilon$   
r4  $\text{addop} \rightarrow +$   
r5  $\text{addop} \rightarrow -$   
r6  $T \rightarrow F T'$   
r7  $T' \rightarrow \text{mulop } F T'$   
r8  $T' \rightarrow \epsilon$   
r9  $\text{mulop} \rightarrow *$   
r10  $F \rightarrow ( E )$   
r11  $F \rightarrow \text{id}$

	(	id	)	+	-	*	\$
E	r1	r1					
E'			r3	r2	r2		r3
addop				r4	r5		
T	r6	r6					
T'			r8	r8	r8	r7	r8
mulop						r9	
F	r10	r11					

# LL(1) Parsing

- LL(1) Grammar

- A grammar is an LL(1) grammar if the associated LL(1) parsing table has at most one production in each table entry
- Cannot be ambiguous
- A subset of CFG

$S \rightarrow A$   
 $A \rightarrow ( A ) A$   
 $A \rightarrow \epsilon$

	(	)	\$
S	$S \rightarrow A$		$S \rightarrow A$
A	$A \rightarrow ( A ) A$	$A \rightarrow \epsilon$	$A \rightarrow \epsilon$

# LL(1) Parsing: Algorithm

```
push start symbol  $S$  onto stack
while stack top  $\neq$   $\$$  and next token  $\neq$   $\$$ 
    if stack top is  $a$  and  $a ==$  next token
        pop stack
        advance input
    else if stack top is  $A$  and next token is  $a$ 
        and  $[A,a]$  has rule  $A \rightarrow X_1X_2\dots X_n$ 
            pop stack
            for  $i$  from  $n$  to  $1$ 
                push  $X_i$  onto stack
    else
        error
if stack top  $==$  next token  $==$   $\$$ 
    accept
else
    error
```

# Issues Related to LL(1)



# A Grammar is LL(1) iff

$A \rightarrow \alpha \mid \beta$



$\text{First}(\alpha) \cap \text{First}(\beta) = \Phi$

$A \rightarrow \alpha \mid \beta$

st  $\beta \Rightarrow^* \epsilon$



$\text{First}(\alpha) \cap \text{Follow}(A) = \Phi$

# Left Recursion

- Left recursion often makes the grammar non-LL(1)

```

exp → exp addop term | term
addop → + | -
term → term mulop factor | factor
mulop → * | /
factor → ( exp ) | number
    
```

First(exp addop term) = { (, **number** }

First(term) = { (, **number** }

	(	number	...
exp	exp → exp addop term exp → term	exp → exp addop term exp → term	

# Left Recursion

- **Rewriting:** break it into two rules: (i) generate base case first and (ii) generate the repetition using right recursion

$$A \rightarrow A \alpha \mid \beta$$



$$\begin{aligned} A &\rightarrow \beta A' \\ A' &\rightarrow \alpha A' \mid \varepsilon \end{aligned}$$

right recursion

$$A \rightarrow A\alpha_1 \mid A\alpha_2 \mid \dots \mid A\alpha_n \mid \beta_1 \mid \beta_2 \mid \dots \mid \beta_m$$



$$\begin{aligned} A &\rightarrow \beta_1 A' \mid \beta_2 A' \mid \dots \mid \beta_m A' \\ A' &\rightarrow \alpha_1 A' \mid \alpha_2 A' \mid \dots \mid \alpha_n A' \mid \varepsilon \end{aligned}$$

General Form

$\alpha$  and  $\beta$  are strings of terminals and nonterminals and  $\beta$  does not begin with  $A$

# Left Recursion

- **Exercise**

$$A \rightarrow A \alpha \mid \beta$$

$$\begin{aligned} A &\rightarrow \beta A' \\ A' &\rightarrow \alpha A' \mid \varepsilon \end{aligned}$$
$$exp \rightarrow exp \text{ addop } term \mid term$$

$$\begin{aligned} exp &\rightarrow term \ exp' \\ exp' &\rightarrow \text{addop } term \ exp' \mid \varepsilon \end{aligned}$$

# Left Recursion

- Example

```
exp → exp addop term | term
addop → + | -
term → term mulop factor | factor
mulop → *
factor → ( exp ) | number
```

```
exp → term exp'
exp' → addop term exp' | ε
addop → + | -
term → factor term'
term' → mulop factor term' | ε
mulop → *
factor → ( exp ) | number
```



# Left Factoring

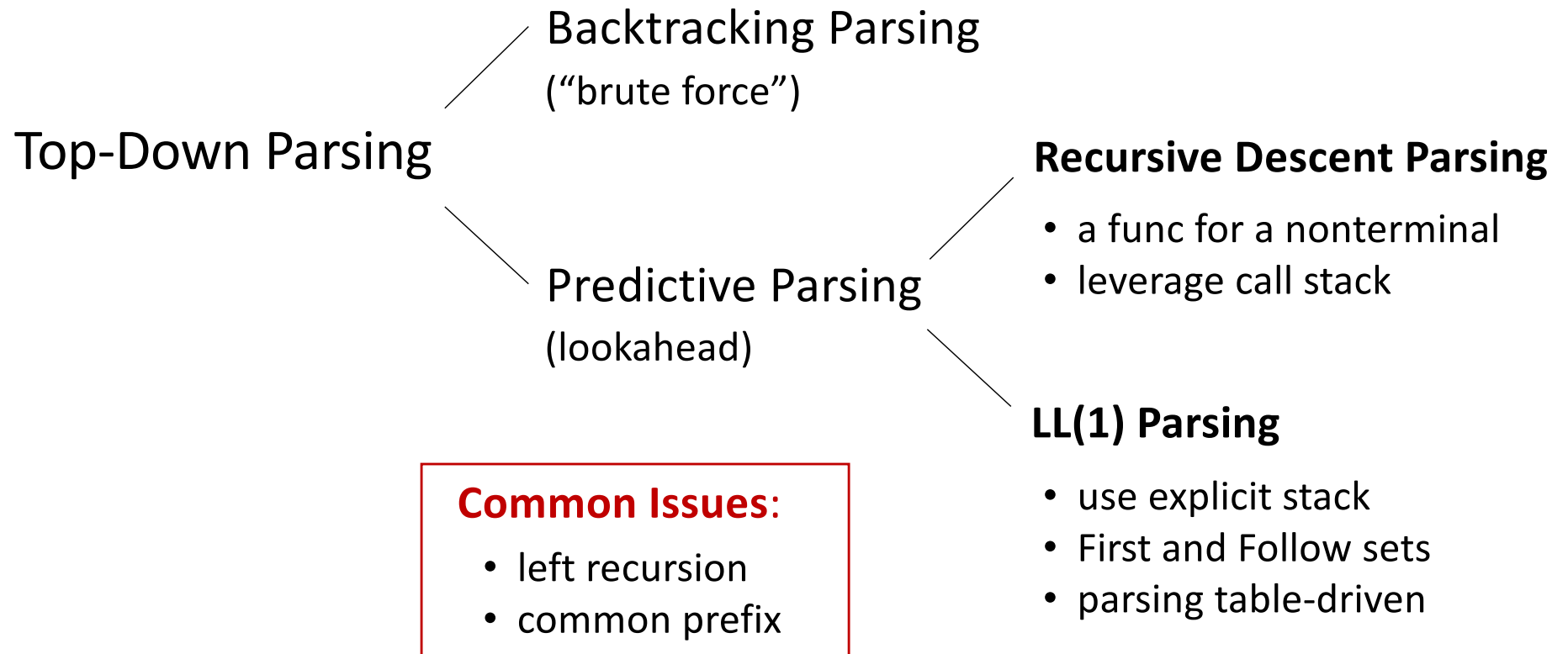
- **Issue:** when grammar rule choices share a common prefix, look ahead one symbol may not be sufficient to determine the rule
- **Rewriting:** take the common part out and add a new nonterminal

$$A \rightarrow \alpha \beta \mid \alpha \gamma$$

$$\begin{aligned} A &\rightarrow \alpha A' \\ A' &\rightarrow \beta \mid \gamma \end{aligned}$$
$$\begin{aligned} \textit{if-stmt} &\rightarrow \textit{if} (\textit{exp}) \textit{stmt} \\ &\mid \textit{if} (\textit{exp}) \textit{stmt} \textit{else} \textit{stmt} \end{aligned}$$

$$\begin{aligned} \textit{if-stmt} &\rightarrow \textit{if} (\textit{exp}) \textit{stmt} \textit{else-part} \\ \textit{else-part} &\rightarrow \textit{else} \textit{stmt} \mid \epsilon \end{aligned}$$

# Summary



# SAMPLE PROBLEMS



# Example

$S \rightarrow A \mid BC$

$A \rightarrow aA \mid \varepsilon$

$B \rightarrow bB \mid \varepsilon$

$C \rightarrow cC \mid dC \mid \varepsilon$

	FIRST	FOLLOW
S	a,b,c,d, $\varepsilon$	\$
A	a, $\varepsilon$	\$
B	b, $\varepsilon$	c,d,\$
C	c,d, $\varepsilon$	\$

	a	b	c	d	\$
S	$S \rightarrow A$	$S \rightarrow BC$	$S \rightarrow BC$	$S \rightarrow BC$	$S \rightarrow A$ $S \rightarrow BC$
A	$A \rightarrow aA$				$A \rightarrow \varepsilon$
B		$B \rightarrow bB$	$B \rightarrow \varepsilon$	$B \rightarrow \varepsilon$	$B \rightarrow \varepsilon$
C			$C \rightarrow cC$	$C \rightarrow dC$	$C \rightarrow \varepsilon$

# Example

LEXP  $\rightarrow$  ATOM | LIST  
ATOM  $\rightarrow$  num | id  
LIST  $\rightarrow$  ( LSEQ )  
LSEQ  $\rightarrow$  LSEQ LEXP | LEXP

LEXP  $\rightarrow$  ATOM | LIST  
ATOM  $\rightarrow$  num | id  
LIST  $\rightarrow$  ( LSEQ )  
LSEQ  $\rightarrow$  LEXP LSEQ'  
LSEQ'  $\rightarrow$  LEXP LSEQ' |  $\varepsilon$

	FIRST	FOLLOW
LEXP	num, id, (	\$, num, id, (, )
ATOM	num, id	\$, num, id, (, )
LIST	(	\$, num, id, (, )
LSEQ	num, id, (	)
LSEQ'	num, id, (, $\varepsilon$	)

# Example

LEXP  $\rightarrow$  ATOM | LIST

ATOM  $\rightarrow$  num | id

LIST  $\rightarrow$  ( LSEQ )

LSEQ  $\rightarrow$  LEXP LSEQ'

LSEQ'  $\rightarrow$  LEXP LSEQ' |  $\varepsilon$

	FIRST	FOLLOW
LEXP	num, id, (	\$, num, id, (, )
ATOM	num, id	\$, num, id, (, )
LIST	(	\$, num, id, (, )
LSEQ	num, id, (	)
LSEQ'	num, id, (, $\varepsilon$	)

LEXP  $\rightarrow$  ATOM | LIST

FIRST(ATOM)  $\cap$  FIRST(LIST) =  $\emptyset$

ATOM  $\rightarrow$  num | id

FIRST(num)  $\cap$  FIRST(id) =  $\emptyset$

LSEQ'  $\rightarrow$  LEXP LSEQ' |  $\varepsilon$

FIRST(LEXP)  $\cap$  FOLLOW(LSEQ') =  $\emptyset$

$\Rightarrow$  Grammar is LL(1)

# Example

$LEXP \rightarrow ATOM \mid LIST$

$ATOM \rightarrow num \mid id$

$LIST \rightarrow ( LSEQ )$

$LSEQ \rightarrow LEXP LSEQ'$

$LSEQ' \rightarrow LEXP LSEQ' \mid \varepsilon$

	FIRST	FOLLOW
LEXP	num, id, (	\$, num, id, (, )
ATOM	num, id	\$, num, id, (, )
LIST	(	\$, num, id, (, )
LSEQ	num, id, (	)
LSEQ'	num, id, (, $\varepsilon$	)

	num	id	(	)	\$
LEXP	$LEXP \rightarrow ATOM$	$LEXP \rightarrow ATOM$	$LEXP \rightarrow LIST$		
ATOM	$ATOM \rightarrow num$	$ATOM \rightarrow id$			
LIST			$LIST \rightarrow ( LSEQ )$		
LSEQ	$LSEQ \rightarrow LEXP LSEQ'$	$LSEQ \rightarrow LEXP LSEQ'$	$LSEQ \rightarrow LEXP LSEQ'$		
LSEQ'	$LSEQ' \rightarrow LEXP LSEQ'$	$LSEQ' \rightarrow LEXP LSEQ'$	$LSEQ' \rightarrow LEXP LSEQ'$	$LSEQ' \rightarrow \varepsilon$	