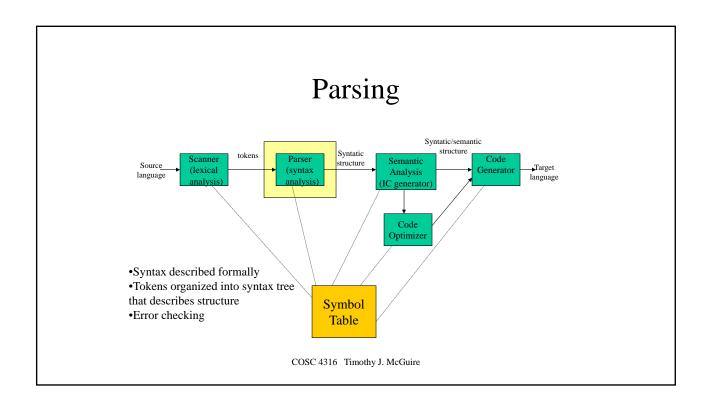
# Lecture 4b: LL Parsing

#### **COSC 4316**

(grateful acknowledgement to Robert van Engelen and Elizabeth White for some of the material from which these slides have been adapted)



#### Constructing an LL(1) Predictive Parsing Table

```
• for each entry M[A,a] do
M[A,a] \leftarrow \emptyset
endfor
• for each production A \rightarrow \alpha do
for each a \in FIRST(\alpha) do
M[A,a] \leftarrow M[A,a] \cup \{A \rightarrow \alpha \}
endfor
if \epsilon \in FIRST(\alpha) then
for each b \in FOLLOW(A) do
add A \rightarrow \alpha \text{ to } M[A,b]
endfor
endfor
endfor
```

**for** each entry 
$$M[A,a] = \emptyset$$
 **do**  $M[A,a] \leftarrow \textit{error}$  **endfor**

3

## Example

$$E \rightarrow E + T/T$$

$$T \rightarrow T + F/F$$

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

$$E_R \rightarrow + T E_R \mid \varepsilon$$

$$T \rightarrow F T_R$$

$$T_R \rightarrow * F T_R \mid \varepsilon$$

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

$A \rightarrow \alpha$	FIRST(α)	FOLLOW(A)
$E \to T E_R$	( id	\$)
$E_R \rightarrow + T E_R$	+	\$)
$E_R \rightarrow \varepsilon$	3	\$)
$T \rightarrow F T_R$	( id	+\$)
$T_R \rightarrow *FT_R$	*	+\$)
$T_R \rightarrow \varepsilon$	3	+\$)
$F \rightarrow (E)$	(	*+\$)
$F \rightarrow id$	id	* + \$ )

COSC 4316, Timothy J. McGuire

4

$A \rightarrow \alpha$	FIRST(α)	FOLLOW(A)
$E \rightarrow T E_R$	( id	\$)
$E_R \rightarrow + T E_R$	+	\$)
$E_R \rightarrow \varepsilon$	3	\$)
$T \rightarrow F T_R$	( id	+ \$ )
$T_R \to *FT_R$	*	+ \$ )
$T_R \rightarrow \varepsilon$	з	+ \$ )
$F \rightarrow (E)$	(	*+\$)
$F \rightarrow id$	id	*+\$)

	id	+	*	(	)	\$
Ε	$E \rightarrow T E_R$			$E \rightarrow T E_R$		
$E_R$		$E_R \rightarrow + T E_R$				
Т	$T \rightarrow F T_R$			$T \to F T_R$		
$T_R$			$T_R \to *FT_R$			
F	$F \rightarrow id$			$F \rightarrow (E)$		

 $\begin{array}{c} \textbf{for} \ \text{each production} \ A \to \alpha \ \textbf{do} \\ & \textbf{for} \ \text{each} \ a \in \mathsf{FIRST}(\alpha) \ \textbf{do} \\ & \text{add} \ A \to \alpha \ \text{to} \ M[A,a] \\ & \textbf{enddo} \\ & \textbf{if} \ \epsilon \in \mathsf{FIRST}(\alpha) \ \textbf{then} \\ & \textbf{for} \ \text{each} \ b \in \mathsf{FOLLOW}(A) \ \textbf{do} \\ & \text{add} \ A \to \alpha \ \text{to} \ M[A,b] \\ & \textbf{enddo} \\ & \textbf{enddo} \\ & \textbf{Mark} \ \text{each undefined entry in} \ M \ \text{error} \end{array}$ 

COSC 4316, Timothy J. McGuire

$A \rightarrow \alpha$	FIRST(α)	FOLLOW(A)
$E \rightarrow T E_R$	( id	\$)
$E_R \rightarrow + T E_R$	+	\$)
$E_R \rightarrow \varepsilon$	3	\$)
$T \rightarrow F T_R$	( id	+ \$ )
$T_R \rightarrow *FT_R$	*	+ \$ )
$T_R \rightarrow \varepsilon$	3	+ \$ )
$F \rightarrow (E)$	(	*+\$)
$F \rightarrow id$	id	*+\$)

	id	+	*	(	)	\$
E	$E \to T E_R$			$E \rightarrow T E_R$		
$E_R$		$E_R \rightarrow + T E_R$			$E_R \rightarrow \varepsilon$	$E_R \rightarrow \varepsilon$
T	$T \rightarrow F T_R$			$T \rightarrow F T_R$		
$T_R$		$T_R \rightarrow \varepsilon$	$T_R \to *FT_R$		$T_R \rightarrow \varepsilon$	$T_R \rightarrow \varepsilon$
F	$F \rightarrow id$			$F \rightarrow (E)$		

COSC 4316, Timothy J. McGuire

6

5

# Non-LL(1) Examples

Grammar	Not LL(1) because:
$S \rightarrow S \mathbf{a} \mid \mathbf{a}$	Left recursive
$S \rightarrow \mathbf{a} S \mid \mathbf{a}$	$FIRST(\mathbf{a} S) \cap FIRST(\mathbf{a}) \neq \emptyset$
$S \rightarrow \mathbf{a} R \mid \varepsilon$	
$R \to S \mid \varepsilon$	For $R: S \Rightarrow^* \varepsilon$ and $\varepsilon \Rightarrow^* \varepsilon$
$S \rightarrow \mathbf{a} R \mathbf{a}$	For <i>R</i> :
$R \to S \mid \varepsilon$	$FIRST(S) \cap FOLLOW(R) \neq \emptyset$

# LL(1) Grammars are Unambiguous

Ambiguous grammar

$$S \rightarrow \mathbf{i} E \mathbf{t} S SR \mid \mathbf{a}$$
  
 $SR \rightarrow \mathbf{e} S \mid \varepsilon$ 

 $E \rightarrow \mathbf{b}$ 



$A \rightarrow \alpha$	FIRST(α)	FOLLOW(A)
$S \rightarrow \mathbf{i} E \mathbf{t} S S_R$	i	e \$
$S \rightarrow \mathbf{a}$	a	e \$
$S_R \to \mathbf{e} S$	e	e \$
$S_R \rightarrow \varepsilon$	ε	e \$
$E \rightarrow \mathbf{b}$	b	t

<u>Err</u>	<u>or: auni</u>	icate tar	ne entrv			
	a	b	e	i	t	\$
S	$S \rightarrow \mathbf{a}$			$S \rightarrow \mathbf{i} E \mathbf{t} S S_R$		
$S_R$		(	$S_R \to \varepsilon$ $S_R \to \mathbf{e} S$	)		$S_R \rightarrow \varepsilon$
E		$E \rightarrow \mathbf{b}$				

#### Error Recovery in Predictive Parsing

- Panic Mode Recovery:
  - (Skip symbols until a set of synchronizing tokens appears)
- How to choose a synchronizing set? One option (among several):
  - 1. Place all terminal symbols in FOLLOW(*A*) into the synchronizing set for *A*
  - 2. Wherever *A* has an undefined entry corresponding to a synchronizing token, add a synchronizing action to that entry.
  - 3. If a synch action is found, pop the current nonterminal *A* from the stack and skip until a synchronizing token is found, *or*
  - 4. Don't pop *A*, but skip input until a symbol in FIRST(*A*) is found.

COSC 4316 Timothy J. McGuire

#### Panic Mode Recovery

Add synchronizing actions to undefined entries based on FOLLOW

FOLLOW(E) = { ) \$ } FOLLOW(ER) = { ) \$ } FOLLOW(T) = { + ) \$ }

Pro: Can be automated

Cons: Error messages are needed

FOLLOW(T) = { + ) \$ } FOLLOW(TR) = { + ) \$ } FOLLOW(F) = { + \* ) \$ }

_	<b>01</b> 11	nessages a	re meeded				
		id	+	*	(		\$
	E	$E \rightarrow T E_R$			$E \to T E_R$	synch	synch
	$E_R$		$E_R \rightarrow + T E_R$			$E_R \rightarrow \varepsilon$	$E_R \rightarrow \varepsilon$
	T	$T \rightarrow F T_R$	synch		$T \to F T_R$	synch	synch
	$T_R$		$T_R \rightarrow \varepsilon$	$T_R \rightarrow *F T_R$		$T_R \rightarrow \varepsilon$	$T_R \rightarrow \varepsilon$
	F	$F \rightarrow id$	synch	synch	$F \rightarrow (E)$	synch	synch

synch: the driver pops current nonterminal A and skips input till synch token or skips input until one of FIRST(A) is found

10

## Phrase-Level Recovery

Change input stream by inserting missing tokens

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

Pro: Can be automated

Cons: Recovery not always intuitive

Can then	continue	here
----------	----------	------

	id	+	*	(	)	\$
E	$E \to T E_R$			$E \rightarrow T E_R$	synch	synch
$E_R$		$E_R \to + T E_R$			$E_R \rightarrow \varepsilon$	$E_R \rightarrow \varepsilon$
T	$T \rightarrow F T_R$	synch		$T \to F T_R$	synch	synch
$T_R$	insert *	$T_R \to \varepsilon$	$T_R \rightarrow *F T_R$		$T_R \rightarrow \varepsilon$	$T_R \rightarrow \varepsilon$
F	$F \rightarrow id$	synch	synch	$F \rightarrow (E)$	synch	synch

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

11

#### **Error Productions**

$$\begin{split} E &\rightarrow T \, E_R \\ E_R &\rightarrow + T \, E_R \mid \epsilon \\ T &\rightarrow F \, T_R \\ TR &\rightarrow * F \, T_R \mid \epsilon \\ F &\rightarrow (E) \mid \mathbf{id} \end{split}$$

Add "error production":  $T_R \rightarrow F T_R$ 

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

Pro: Powerful recovery method Cons: Cannot be automated

	id	+	*	(	)	\$
E	$E \rightarrow T E_R$			$E \rightarrow T E_R$	synch	synch
$E_R$		$E_R \rightarrow + T E_R$			$E_R \rightarrow \varepsilon$	$E_R \rightarrow \varepsilon$
T	$T \rightarrow F T_R$	synch		$T \to F T_R$	synch	synch
$T_R$	$T_R \to F T_R$	$T_R \to \varepsilon$	$T_R \to *FT_R$		$T_R \rightarrow \varepsilon$	$T_R \rightarrow \varepsilon$
F	$F \rightarrow id$	synch	synch	$F \rightarrow (E)$	synch	synch

12

#### LL(k) parsing

also known as the lookahead

- Process input *k* symbols at a time.
- Initially, 'current' non-terminal is start symbol.
- Algorithm
  - Loop until no more input
    - Given next k input tokens and 'current' non-terminal T, choose a rule R (T  $\rightarrow$  ...)
    - For each element X in rule R from left to right,
       if X is a non-terminal, we will need to 'expand' X
       else if symbol X is a terminal, see if next input symbol matches X; if so, update from the input
- Typically, we consider **LL(1)**

COSC 4316 Timothy J. McGuire

## Two Approaches

- Recursive Descent parsing
  - Code tailored to the grammar
- Table Driven predictive parsing
  - Table tailored to the grammar
  - General Algorithm

Both algorithms driven by the tokens coming from the lexer.

# Writing a Recursive Descent Parser

• Generate a procedure for each non-terminal.

Use next token from yylex() (lookahead) to choose (PREDICT) which production to 'mimic'.

- for non-terminal X, call procedure X()
- for terminals X, call 'match(X)'

```
Ex: B \rightarrow b C D

B() {
    if (lookahead == 'b')
        { match('b'); C(); D(); }
        else ...
}
```

COSC 4316 Timothy J. McGuire

## Writing a Recursive Descent Parser

#### Also need the following:

```
match(symbol)
{
    if (symbol == lookahead)
        lookahead = yylex()
    else error()
}
main()
{
    lookahead = yylex();
    S();    /* S is the start symbol */
    if (lookahead == EOF) then accept
    else reject
}
error()
{
    ...
}
```

#### Back to grammar **S**() { $S \rightarrow a B$ **if** (lookahead == a ) { match(a);B(); } $S \rightarrow b C$ else if (lookahead == b) { match(b); C(); } else error("expecting a or b"); **B**() { if (lookahead == b) $B \rightarrow b b C$ {match(b); match(b); C();} else error(); **C**() { if (lookahead == c) $C \rightarrow c c$ { match(c) ; match(c) ;} else error(); COSC 4316 Timothy J. McGuire

# Parsing abbcc

S

```
Remaining input: abbcc
Call S() from main()
S() {
  if (lookahead == a ) { match(a);B(); }
                                               S \rightarrow a B
 else if (lookahead == b) { match(b); C(); } S \rightarrow b C
 else error("expecting a or b");
```

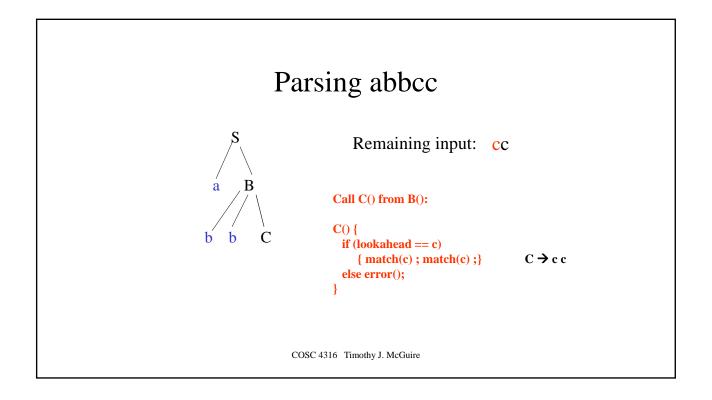
```
Parsing abbcc

Remaining input: bbcc

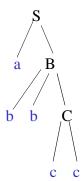
Call B() from A():

B() {
    if (lookahead == b)
    {match(b); match(b); C();}    B \rightarrow b b C else error();
    }

COSC 4316 Timothy J. McGuire
```



# Parsing abbcc



Remaining input:

COSC 4316 Timothy J. McGuire

#### How do we find the lookaheads?

- Can compute PREDICT sets from FIRST and FOLLOW for LL(1) parsing:
- PREDICT(A  $\rightarrow \alpha$ )

 $= (FIRST(\alpha) - \{\epsilon\}) \cup FOLLOW(A) \qquad \text{if $\epsilon$ in $FIRST(\alpha)$} \\ = FIRST(\alpha) \qquad \qquad \text{if $\epsilon$ not in $FIRST(\alpha)$}$ 

#### NOTE: ε never in PREDICT sets

For LL(k) grammars, the PREDICT sets for the productions associated with a given non-terminal must be disjoint.

## Example

Production	Predict
E → T E'	$= FIRST(T) = \{(,id\}$
E' → + T E'	{+}
E' <b>→</b> ε	$= FOLLOW(E') = \{\$,\}$
$T \rightarrow F T'$	$= FIRST(F) = \{(,id\}$
T' → * F T'	{*}
T' → ε	$= FOLLOW(T') = \{+,\$,\}$
F → id	{id}
$F \rightarrow (E)$	{(}

```
FIRST(F) = {(,id}

FIRST(T) = {(,id}

FIRST(E) = {(,id}

FIRST(T') = {*,ε}

FIRST(E') = {+,ε}

FOLLOW(E) = {$,)}

FOLLOW(E') = {$,)}

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

FOLLOW(T') = {+,$,}
```

Assume E is the start symbol

```
E() {
                                                                   E \rightarrow TE'
  if (lookahead in {(,id } ) { T(); E_prime(); }
  else error("E expecting ( or identifier");
}
E_prime() {
                                                                   E' \rightarrow + T E'
   if (lookahead in {+}) {match(+); T(); E_prime();}
                                                                   E' \rightarrow e
   else if (lookahead in {),end_of_file}) return;
   else error("E_prime expecting +, ) or end of file");
}
T() {
                                                                   T \rightarrow F T'
   if (lookahead in {(,id}) { F(); T_prime(); }
   else error("T expecting ( or identifier");
}
                                  COSC 4316 Timothy J. McGuire
```

```
T_{prime()} \{ \\ & \text{if (lookahead in $\{*\}) \{match(*); F(); T_{prime();}\}} \quad \textbf{T'} \rightarrow * \textbf{F} \textbf{T'} \\ & \text{else if (lookahead in $\{+,), \text{end_of_file}\}) return;} \quad \textbf{T'} \rightarrow \epsilon \\ & \text{else error("T_prime expecting $*, ) or end of file");} \} \\ & F() \{ \\ & \text{if (lookahead in $\{id\}) match(id);} \quad \textbf{F} \rightarrow \text{id} \\ & \text{else if (lookahead in $\{(\}) \{ match((); E(); match(()); \} \} } \quad \textbf{F} \rightarrow (E) \\ & \text{else error("F expecting ( or identifier");}} \} \\ \end{aligned}
```

# Parsing a + b \* c

COSC 4316 Timothy J. McGuire

E Remaining input: a+b\*c

# Parsing a + b \* c

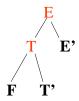


Remaining input: a+b\*c

```
E() {
    if (lookahead in {(,id } ) { T(); E_prime(); }
    else error("E expecting ( or identifier");
}
```

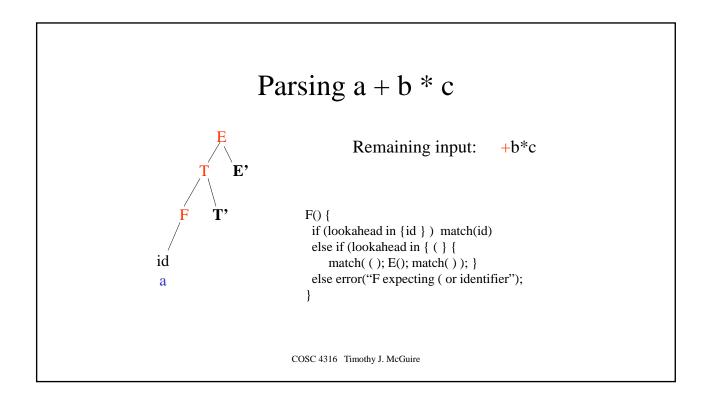
COSC 4316 Timothy J. McGuire

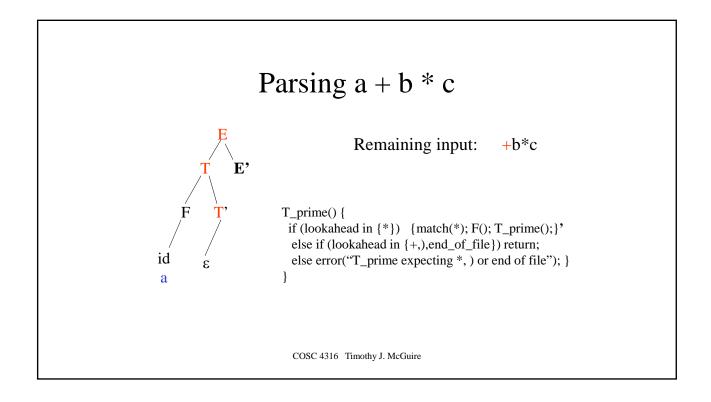
# Parsing a + b \* c

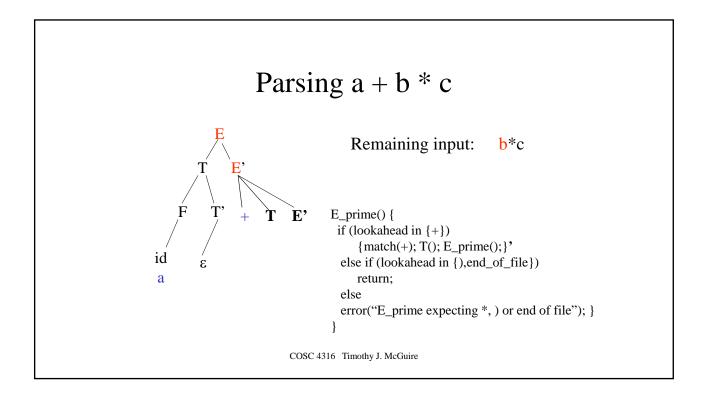


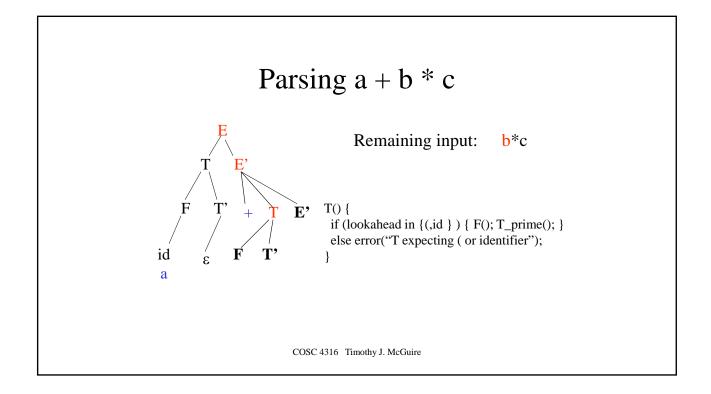
Remaining input: a+b\*c

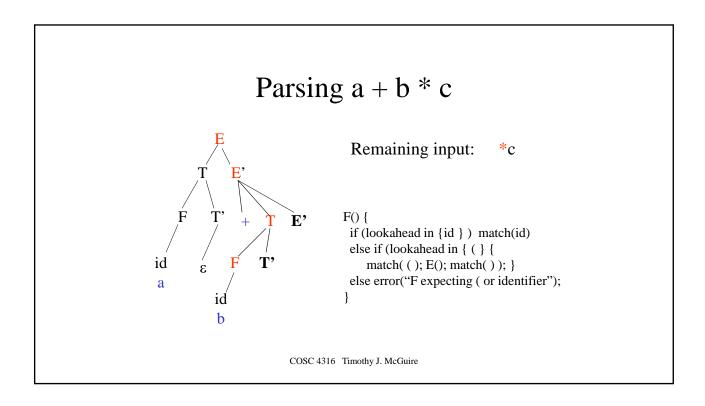
```
T() {
    if (lookahead in {(,id } ) { F(); T_prime(); }
    else error("T expecting ( or identifier");
}
```

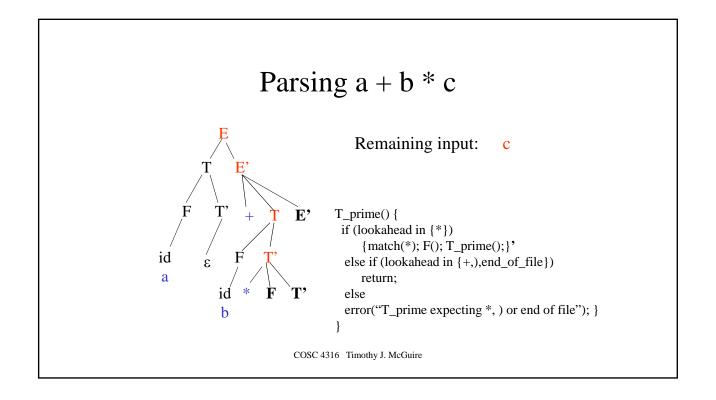


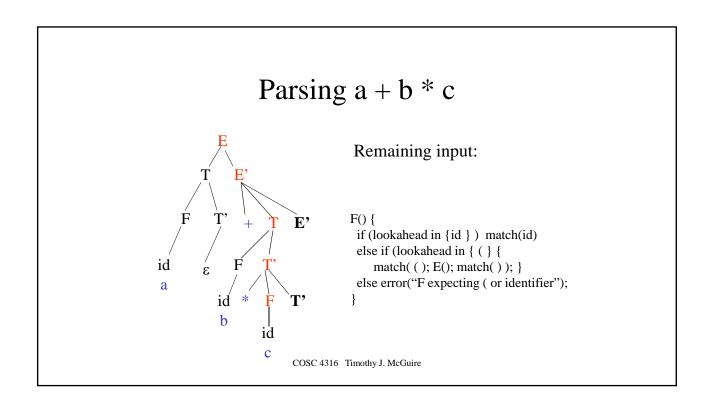


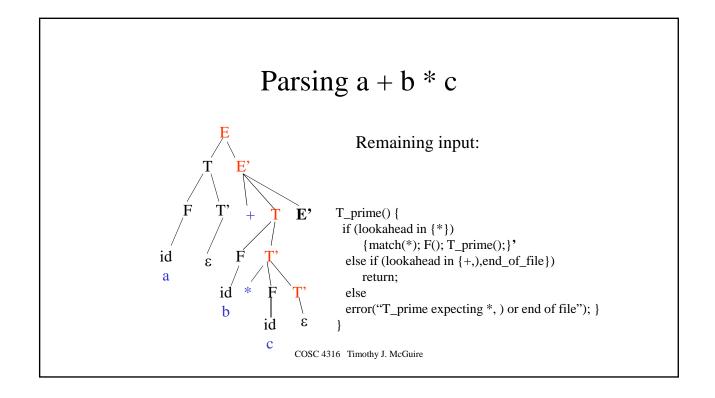






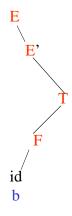






# Parsing a + b \* cRemaining input: F T' + T E' E\_prime() { if (lookahead in {+}) {match(+); T(); E\_prime();}' else if (lookahead in {),end\_of\_file}) return; else error("E\_prime expecting \*, ) or end of file"); } C COSC 4316 Timothy J. McGuire

## Stacks in Recursive Descent Parsing



- Runtime stack
- Procedure activations correspond to a path in parse tree from root to some interior node

#### Two Approaches

- Recursive Descent parsing
  - Code tailored to the grammar
- Table Driven predictive parsing
  - Table tailored to the grammar
  - General Algorithm

Both algorithms driven by the tokens coming from the lexer.

COSC 4316 Timothy J. McGuire

#### LL(1) Predictive Parse Tables

An LL(1) Parse table is a mapping T:

 $N \times T \rightarrow production P or error$ 

1. For all productions A  $\rightarrow \alpha$  do

For each terminal t in Predict(A  $\rightarrow \alpha$ ),

 $T[A,t] = A \rightarrow \alpha$ 

2. Every undefined table entry is an error.

## Using LL(1) Parse Tables

#### **ALGORITHM**

INPUT: token sequence to be parsed, followed by '\$' (end of file)

#### **DATA STRUCTURES:**

- Parse stack: Initialized by pushing '\$' and then pushing the start symbol
- Parse table T

COSC 4316 Timothy J. McGuire

# Algorithm: Predictive Parsing

# Example

N\T	+	*	(	)	ID	\$
Е						
E'						
T						
T'						
F						

Production	Predict
1: E → T E'	{(,id}
2: E' → + T E'	{+}
3: E' → ε	{\$,)}
4: T → F T'	{(,id}
5: T' → * F T'	{*}
6: T' <b>→</b> ε	{+,\$,)}
7: F → id	{id}
8: F → (E)	{(}

COSC 4316 Timothy J. McGuire

N∖T	+	*	(	)	ID	\$
Е			1		1	
E'	2			3		3
Т			4		4	
T'	6	5		6		6
F			8		7	

Production	Predict
1: E → T E'	{(,id}
2: E' → + T E'	{+}
3: E' → ε	{\$,)}
4: T → F T'	{(,id}
5: T' → * F T'	{*}
6: T' → ε	{+,\$,)}
7: F → id	{id}
8: F → (E)	{(}

							Stack	Input	Action
							\$E	a+b*c\$	E → T E'
NT/T	+	*	(	)	ID	\$			
Е			1		1				
E'	2			3		3			
T			4		4				
T'	6	5		6		6			
F			8		7				
A	SSUM	ne E i	s the	start	svmh	pol			

ΨΕ	*c\$ E <del>&gt;</del>	
(ALIVER ALIVER A	· '	<b>→</b> T E'
NT/T + * ( ) ID \$ \$E'T a+b	*c\$ T →	→ F T'
E 1 1 1		
E' 2 3 3		
T 4 4 4		
T' 6 5 6 6		
F 8 7		

NT/T	+	*	(	)	ID	\$
Е			1		1	
E'	2			3		3
T			4		4	
T'	6	5		6		6
F			8		7	

Stack	Input	Action
\$E	a+b*c\$	E → T E'
\$E'T	a+b*c\$	T → F T'
\$E'T'F	a+b*c\$	F → id

NT/T	+	*	(	)	ID	\$
Е			1		1	
E'	2			3		3
Т			4		4	
T'	6	5		6		6
F			8		7	

Stack	Input	Action
\$E	a+b*c\$	E <b>→</b> T E'
\$E'T	a+b*c\$	$T \rightarrow F T'$
\$E'T'F	a+b*c\$	F → id
\$E'T'id	a+b*c\$	match

NT/T	+	*	(	)	ID	\$
Е			1		1	
E'	2			3		3
Т			4		4	
T'	6	5		6		6
F			8		7	

Stack	Input	Action
\$E	a+b*c\$	E → T E'
\$E'T	a+b*c\$	T → F T'
\$E'T'F	a+b*c\$	F → id
\$E'T'id	a+b*c\$	match
\$E'T'	+b*c\$	Τ' → ε

NT/T	+	*	(	)	ID	\$
Е			1		1	
E'	2			3		3
Т			4		4	
T'	6	5		6		6
F			8		7	

Stack	Input	Action
\$E	a+b*c\$	E → T E'
\$E'T	a+b*c\$	T → F T'
\$E'T'F	a+b*c\$	F → id
\$E'T'id	a+b*c\$	match
\$E'T'	+b*c\$	Τ' → ε
\$E'	+b*c\$	E' → + T E'

NT/T	+	*	(	)	ID	\$
Е			1		1	
E'	2			3		3
Т			4		4	
T'	6	5		6		6
F			8		7	

Stack	Input	Action
\$E	a+b*c\$	E → T E'
\$E'T	a+b*c\$	T → F T'
\$E'T'F	a+b*c\$	F → id
\$E'T'id	a+b*c\$	match
\$E'T'	+b*c\$	Τ' → ε
\$E'	+b*c\$	E' → + T E'
\$E'T+	+b*c\$	match

NT/T	+	*	(	)	ID	\$
Е			1		1	
E'	2			3		3
T			4		4	
T'	6	5		6		6
F			8		7	

Stack	Input	Action
\$E	a+b*c\$	E → T E'
\$E'T	a+b*c\$	T → F T'
\$E'T'F	a+b*c\$	F → id
\$E'T'id	a+b*c\$	match
\$E'T'	+b*c\$	Τ' → ε
\$E'	+b*c\$	E' → + T E'
\$E'T+	+b*c\$	match
\$E' T	b*c\$	$T \rightarrow FT'$

NT/T	+	*	(	)	ID	\$
Е			1		1	
E'	2			3		3
T			4		4	
T'	6	5		6		6
F			8		7	

Stack	Input	Action
\$E	a+b*c\$	E → T E'
\$E'T	a+b*c\$	T → F T'
\$E'T'F	a+b*c\$	F → id
\$E'T'id	a+b*c\$	match
\$E'T'	+b*c\$	Τ' → ε
\$E'	+b*c\$	E' → + T E'
\$E'T+	+b*c\$	match
\$E' T	b*c\$	T → F T'
\$E'T'F	b*c\$	F → id

NT/T	+	*	(	)	ID	\$
Е			1		1	
E'	2			3		3
Т			4		4	
T'	6	5		6		6
F			8		7	

Stack	Input	Action
\$E	a+b*c\$	E → T E'
\$E'T	a+b*c\$	T → F T'
\$E'T'F	a+b*c\$	F → id
\$E'T'id	a+b*c\$	match
\$E'T'	+b*c\$	Τ' → ε
\$E'	+b*c\$	E' → + T E'
\$E'T+	+b*c\$	match
\$E' T	b*c\$	T → F T'
\$E'T'F	b*c\$	F → id
\$E'T id	b*c\$	match

# Parsing a + b \* c

Stack	Input	Action
\$E	a+b*c\$	E → T E'
\$E'T		$T \rightarrow F T'$
\$E'T'F		F → id
\$E'T'id		match
\$E'T'	+b*c\$	T' <b>→</b> ε
\$E'		E' → + T E'
\$E'T+		match
\$E'T	b*c\$	$T \rightarrow F T'$

Stack	Input	Action
\$E'T'F		F → id
\$E'T'id		match
\$E'T'	*c\$	$T' \rightarrow * F T'$
\$E'T'F*		match
\$E'T'F	c\$	F → id
\$E'T'id		match
\$E'T'	\$	T' <b>→</b> ε
\$E'		E' <b>→</b> ε
\$		accept