Class Information

• Homework 2 due tomorrow.
• Homework 3 will be posted early next week.
Basic Idea:

• The parse tree is constructed from the root, expanding non-terminal nodes on the tree’s frontier following a leftmost derivation.
• The input program is read from left to right, and input tokens are read (consumed) as the program is parsed.
• The next non-terminal symbol is replaced using one of its rules. The particular choice has to be unique and uses parts of the input (partially parsed program), for instance the first token of the remaining input.
Example:

\[ S ::= a \ S \ b | \varepsilon \]

How can we parse (automatically construct a leftmost derivation) the input string \texttt{a a a b b b} using a PDA (push-down automaton) and only the first symbol of the remaining input?

INPUT: \texttt{a a a b b b eof}
LL(1) Parsing Example

S ::= a S b | ε

Remaining Input:

a a a b b b

Sentential Form:

S

Applied Production:
S ::= a S b | ε

Remaining Input: a a a b b b

Sentential Form: a S b

Applied Production: S ::= a S b
LL(1) Parsing Example

S ::= a S b | ε

Remaining Input: a a a b b b

Sentential Form: a S b

Applied Production: Match!
LL(1) Parsing Example

\[ S ::= a \ S \ b \mid \varepsilon \]

Remaining Input:

\[ a \ a \ b \ b \ b \]

Sentential Form:

\[ a \ S \ b \]

Applied Production:
LL(1) Parsing Example

S ::= a S b | ε

Remaining Input:

a a b b b

Sentential Form:

a a S b b

Applied Production:

S ::= a S b
LL(1) Parsing Example

\[ S ::= a\ S\ b \mid \varepsilon \]

Remaining Input: \[ \text{[a]}a\ b\ b\ b \]

Sentential Form: \[ a\ a\ S\ b\ b \]

Applied Production:

Match!
LL(1) Parsing Example

\[ S ::= a \ S \ b | \epsilon \]

Remaining Input: 
\[ a \ b \ b \ b \]

Sentential Form: 
\[ a \ a \ S \ b \ b \]

Applied Production:
LL(1) Parsing Example

S ::= a S b | \( \varepsilon \)

Remaining Input: 

a b b b

Sentential Form: 

a a a S b b b

Applied Production: 

S ::= a S b
LL(1) Parsing Example

S ::= a S b | ε

Remaining Input: a b b b

Sentential Form: a a a S b b b

Applied Production: Match!
LL(1) Parsing Example

\[ S ::= a \ S \ b | \varepsilon \]

Remaining Input:
\[ b \ b \ b \]

Sentential Form:
\[ a \ a \ a \ S \ b \ b \ b \]

Applied Production:
LL(1) Parsing Example

S ::= a S b | ε

Remaining Input:
   b b b

Sentential Form:
   a a a b b b

Applied Production:
   S ::= ε
LL(1) Parsing Example

\[ S ::= a \; S \; b \mid \varepsilon \]

Remaining Input:
\[ bb b \]

Sentential Form:
\[ a \; a \; a \; b \; b \; b \]

Applied Production:

Match!
LL(1) Parsing Example

\[ S ::= a \ S \ b \mid \varepsilon \]

Remaining Input: \( b \ b \)

Sentential Form: \( a \ a \ a \ b \ b \ b \)

Applied Production:
**LL(1) Parsing Example**

\[
S ::= a \ S \ b \mid \varepsilon
\]

Remaining Input: `b b`

Sentential Form: `a a a b b b`

Applied Production:

Match!
LL(1) Parsing Example

S ::= a S b | ε

Remaining Input: b

Sentential Form: a a a b b b

Applied Production:
LL(1) Parsing Example

S ::= a S b | ε

Remaining Input: b

Sentential Form: a a a b b b

Applied Production:

Match!

ε
LL(1) Parsing Example

$S ::= a S b \mid \varepsilon$

Remaining Input:

Sentential Form:

a a a b b b

Applied Production:
Another LL(1) Parsing Example

Consider this example grammar:

\[
\begin{align*}
\text{id\_list} & \ ::= \text{id} \ \text{id\_list\_tail} \\
\text{id\_list\_tail} & \ ::= , \ \text{id} \ \text{id\_list\_tail} \\
\text{id\_list\_tail} & \ ::= ;
\end{align*}
\]

How to parse the following input string?

\[
\text{A, B, C;}
\]
Another LL(1) Parsing Example

id_list ::= id id_list_tail
id_list_tail ::= , id id_list_tail
id_list_tail ::= ;

id_list

Remaining Input: A, B, C;

Sentential Form:
  id_list

Applied Production:
Another LL(1) Parsing Example

id_list ::= id id_list_tail
id_list_tail ::= , id id_list_tail
id_list_tail ::= ;

Remaining Input: A, B, C;

Sentential Form:
\textbf{id}(A) \ \textbf{id\_list\_tail}

Applied Production:
\textbf{id\_list} ::= \textbf{id} \ \textbf{id\_list\_tail}
Another LL(1) Parsing Example

id_list ::= id id_list_tail
id_list_tail ::= , id id_list_tail
id_list_tail ::= ;

Remaining Input: A, B, C;

Sentential Form: id(A) id_list_tail

Applied Production:
Another LL(1) Parsing Example

id_list ::= id id_list_tail
id_list_tail ::= , id id_list_tail
id_list_tail ::= ;

Remaining Input: , B , C ;

Sentential Form: id(A) id_list_tail

Applied Production:
Another LL(1) Parsing Example

id_list ::= id id_list_tail
id_list_tail ::= , id id_list_tail
id_list_tail ::= ;

Remaining Input: , B , C ;

Sentential Form: id(A) , id(B) id_list_tail

Applied Production: id_list_tail ::= , id id_list_tail
Another LL(1) Parsing Example

id_list ::= id id_list_tail
id_list_tail ::= , id id_list_tail
id_list_tail ::= ;

Remaining Input: [B, C ;]

id(A) id_list_tail
id(B) id_list_tail

Match!

Sentential Form: id(A), id(B) id_list_tail

Applied Production:
Another LL(1) Parsing Example

**Production Rules:**

\[
\begin{align*}
id\_list & ::= \textbf{id} \ id\_list\_tail \\
id\_list\_tail & ::= , \textbf{id} \ id\_list\_tail \\
id\_list\_tail & ::= ;
\end{align*}
\]

**Remaining Input:**

\[B, C;\]

**Sentential Form:**

\[\text{id}(A), \text{id}(B) id\_list\_tail\]

**Applied Production:**

- Start with `id_list`
- Apply the production `id_list ::= id id_list_tail`
- Apply the production `id_list_tail ::= , id id_list_tail`
- Apply the production `id_list_tail ::= ;`
Another LL(1) Parsing Example

id_list ::= id id_list_tail
id_list_tail ::= , id id_list_tail
id_list_tail ::= ;

Remaining Input: B, C;

Sentential Form: id(A), id(B) id_list_tail

Applied Production:
Another LL(1) Parsing Example

id_list ::= id id_list_tail
id_list_tail ::= , id id_list_tail
id_list_tail ::= ;

Remaining Input: , C ;

Sentential Form:
\textbf{id(A)} , \textbf{id(B)} id_list_tail

Applied Production:
Another LL(1) Parsing Example

id_list ::= \texttt{id} id_list_tail
id_list_tail ::= , \texttt{id} id_list_tail
id_list_tail ::= ;

Remaining Input:
, C;

Sentential Form:
id(A), id(B), id(C) id_list_tail

Applied Production:
id_list_tail ::= , \texttt{id} id_list_tail
Another LL(1) Parsing Example

id_list ::= id id_list_tail
id_list_tail ::= , id id_list_tail
id_list_tail ::= ;

Remaining Input: [ ], C ;

Sentential Form:
  id(A) , id(B) , id(C) id_list_tail

Applied Production:

Match!
id_list ::= id id_list_tail
id_list_tail ::= , id id_list_tail
id_list_tail ::= ;

Remaining Input:  
C ;

Sentential Form:  
id(A), id(B), id(C) id_list_tail

Applied Production:
**Another LL(1) Parsing Example**

```
id_list ::= id id_list_tail
id_list_tail ::= , id id_list_tail
id_list_tail ::= ;
```

Remaining Input: 

```
C ;
```

Sentential Form: 

```
id(A) , id(B) , id(C) id_list_tail
```

Applied Production: 

```
```

Match!
Another LL(1) Parsing Example

id_list ::= id id_list_tail
id_list_tail ::= , id id_list_tail
id_list_tail ::= ;

id_list
  --- id(A)
  --- id_list_tail
    --- ,
    --- id(B)
    --- id_list_tail
      --- ,
      --- ,
      --- id(C)
      --- id_list_tail

Remaining Input:
  ;

Sentential Form:
  id(A) , id(B) , id(C) id_list_tail

Applied Production:
Another LL(1) Parsing Example

id_list ::= id id_list_tail
id_list_tail ::= , id id_list_tail
id_list_tail ::= ;

(id(A), id(B), id(C));

Applied Production:
id_list_tail ::= ;

Remaining Input:
;
Another LL(1) Parsing Example

id_list ::= id id_list_tail
id_list_tail ::= , id id_list_tail
id_list_tail ::= ;

Remaining Input:

; ;

Sentential Form:
id(A) , id(B) , id(C) ;

Applied Production:

; ;

Match!
Another LL(1) Parsing Example

\[
id_list ::= \text{id} \ id\_list\_tail \\
id\_list\_tail ::= , \text{id} \ id\_list\_tail \\
id\_list\_tail ::= ;
\]

Remaining Input:

```
Sentential Form:
\text{id}(A) , \text{id}(B) , \text{id}(C) ;
```

Applied Production:
Predictive Parsing

Basic idea:

For any two productions $A ::= \alpha \mid \beta$, we would like a distinct way of choosing the correct production to expand.

For some rhs $\alpha \in G$, define $\text{FIRST}(\alpha)$ as the set of tokens that appear as the first symbol in some string derived from $\alpha$.

That is

$x \in \text{FIRST}(\alpha) \iff \alpha \Rightarrow^* x\gamma$ for some $\gamma$, and
Revisiting the id_list Example

id_list ::= id id_list_tail
id_list_tail ::= , id id_list_tail
id_list_tail ::= ;

id(A) | id_list_tail

Remaining Input: , B, C;

Applied Production:
Revisiting the id_list Example

id_list ::= id id_list_tail
id_list_tail ::= , id id_list_tail
id_list_tail ::= ;

Remaining Input: , B , C ;

Applied Production:
 idi_list_tail ::= , id id_list_tail
Revisiting the id_list Example

id_list ::= id id_list_tail
id_list_tail ::= , id id_list_tail
id_list_tail ::= ;

Remaining Input: , B , C ;

Applied Production:

id_list_tail ::= , id id_list_tail
Revisiting the id_list Example

id_list ::= id id_list_tail
id_list_tail ::= , id id_list_tail
id_list_tail ::= ;

Remaining Input: [ , ] B , C ;

Applied Production:
Revisiting the id_list Example

id_list ::= id id_list_tail
id_list_tail ::= , id id_list_tail
id_list_tail ::= ;

Remaining Input: [B, C];

Applied Production:

id_list_tail ::= ;

Mismatch!
Given id_list_tail as the first non-terminal to expand in the tree:

If the first token of remaining input is , we choose the rule

\[ id_list_tail ::= , \text{id} \ id_list_tail \]

If the first token of remaining input is ; we choose the rule

\[ id_list_tail ::= ; \]
Key Property:

Whenever two productions $A ::= \alpha$ and $A ::= \beta$ both appear in the grammar, we would like

- $FIRST(\alpha) \cap FIRST(\beta) = \emptyset$
Revisiting the id_list Example

id_list ::= id id_list_tail
id_list_tail ::= , id id_list_tail
id_list_tail ::= ;

Remaining Input: , B, C;

id_list

id(A)  id_list_tail

\[
\text{FIRST}(, \text{id id_list_tail}) = \{ , \}
\]
\[
\text{FIRST}(;) = \{ ; \}
\]
\[
\text{FIRST}(, \text{id id_list_tail} \cap \text{FIRST}(;) = \emptyset
\]

Given id_list_tail as the first non-terminal to expand in the tree:

If the first token of remaining input is , we choose the rule

id_list_tail ::= , id id_list_tail

If the first token of remaining input is ; we choose the rule

id_list_tail ::= ;
Key Property:

Whenever two productions $A ::= \alpha$ and $A ::= \beta$ both appear in the grammar, we would like

- $FIRST(\alpha) \cap FIRST(\beta) = \emptyset$

This rule is intuitive. However, it is not correct, because it doesn’t handle $\varepsilon$ rules. How to handle $\varepsilon$ rules?
Revisiting the LL(1) Parsing Example

\[ S ::= a \ S \ b \mid \varepsilon \]

Remaining Input: \(\text{b b b}\)

Applied Production:
Revisiting the LL(1) Parsing Example

\[ S ::= a S b | \varepsilon \]

Remaining Input: \[ b b b \]

Applied Production: \[ S ::= a S b \]

Mismatch!
It only means \[ S ::= a S b \] is not the right production rule to use!
Revisiting the LL(1) Parsing Example

S ::= a S b | ε

Remaining Input: b b b

Applied Production:
Revisiting the LL(1) Parsing Example

\[ S ::= a \; S \; b \mid \varepsilon \]

Remaining Input: 
\[ b \; b \; b \]

Applied Production: 
\[ S ::= \varepsilon \]

\[ S ::= \varepsilon \] turns out to be the right rule later.

However, at this point: 
\[ \varepsilon \] does not match “b” either!
For a non-terminal A, define **FOLLOW**\( (A) \) as the set of terminals that can appear immediately to the right of A in some sentential form.

Thus, a non-terminal’s **FOLLOW** set specifies the tokens that can legally appear after it. A terminal symbol has no **FOLLOW** set.

**FIRST** and **FOLLOW** sets can be constructed automatically
Predictive Parsing

Key Property:

Whenever two productions $A ::= \alpha$ and $A ::= \beta$ both appear in the grammar, we would like

- $FIRST(\alpha) \cap FIRST(\beta) = \emptyset$

This would allow the parser to make a correct choice with a lookahead of only one symbol!
Key Property:

Whenever two productions $A ::= \alpha$ and $A ::= \beta$ both appear in the grammar, we would like

\cdot $\text{FIRST} (\alpha) \cap \text{FIRST} (\beta) = \emptyset$, and
\cdot if $\alpha \Rightarrow^* \varepsilon$, then $\text{FIRST} (\beta) \cap \text{FOLLOW} (A) = \emptyset$
\cdot Analogue case for $\beta \Rightarrow^* \varepsilon$. Note: due to first condition, at most one of $\alpha$ and $\beta$ can derive $\varepsilon$.

This would allow the parser to make a correct choice with a lookahead of only one symbol!
LL(1) Grammar

Define $FIRST^+(A ::= \delta)$ for rule $A ::= \delta$

- $FIRST(\delta) - \{ \varepsilon \} \cup \text{Follow}(A)$, if $\varepsilon \in FIRST(\delta)$
- $FIRST(\delta)$ otherwise

A Grammar is LL(1) iff
$(A ::= \alpha$ and $A ::= \beta)$ implies
\[ FIRST^+(A ::= \alpha) \cap FIRST^+(A ::= \beta) = \emptyset \]
Back to Our Example

Start ::= S eof
S ::= a S b | ε

FIRST(aSb) = {a}
FIRST(ε) = {ε}
FOLLOW(S) = {eof, b}

Is the grammar LL(1)?

FIRST⁺(S ::= aSb) = {a}
FIRST⁺(S ::= ε) = ( FIRST(ε) - {ε} ) ∪ FOLLOW(S) = {eof, b}

Define FIRST⁺(δ) for rule A ::= δ

- FIRST (δ) - { ε } U Follow (A), if ε ∈ FIRST(δ)
- FIRST (δ) otherwise
Table Driven LL(1) Parsing

Example:

\[
S ::= a \ S \ b \mid \varepsilon
\]

LL(1) parse table

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>eof</th>
<th>other</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>aSb</td>
<td>\varepsilon</td>
<td>\varepsilon</td>
<td>error</td>
</tr>
</tbody>
</table>

How to parse input \textbf{a a a b b b}?
Table Driven LL(1) Parsing

**Input:** a string \( w \) and a parsing table \( M \) for \( G \)

push eof
push Start Symbol
token ← next_token()

\( X \) ← top-of-stack
repeat
    if \( X \) is a terminal then
        if \( X == \) token then
            pop \( X \)
            token ← next_token()
        else error()
    else /* \( X \) is a non-terminal */
        if \( M[X, \text{token}] == X \rightarrow Y_1 Y_2 \ldots Y_k \) then
            pop \( X \)
            push \( Y_k, Y_{k-1}, \ldots, Y_1 \)
        else error()

\( X \) ← top-of-stack
until \( X = \text{eof} \)
if token != eof then error()
Now, a predictive parser looks like:

Rather than writing code, we build tables. Building tables can be automated!
Predictive Parsing

Now, a predictive parser looks like:

Rather than writing code, we build tables. Building tables can be automated!
Recursive Descent Parsing

Now, we can produce a recursive descent parser for our LL(1) grammar. Recursive descent is one of the simplest parsing techniques used in practical compilers:

- Each **non-terminal** has an associated parsing procedure that can recognize any sequence of tokens generated by that **non-terminal**
- There is a main routine to initialize all globals (e.g: *tokens*) and call the start symbol. On return, check whether token==eof, and whether errors occurred
- Within a parsing procedure, both **non-terminals** and terminals can be matched:
  - non-terminal A: call procedure for A
  - token t: compare t with current input token; if matched, consume input, otherwise, ERROR
- Parsing procedure may contain code that performs some useful “computations” (*syntax directed translation*)
Recursive Descent Parsing (pseudo code)

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>eof</th>
<th>other</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>aSb</td>
<td>ε</td>
<td>ε</td>
<td>error</td>
</tr>
</tbody>
</table>

main: {
    token := next_token( );
    if (S( ) and token == eof) print “accept” else print “error”; 
}
Recursive Descent Parsing (pseudo code)

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>eof</th>
<th>other</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>aSb</td>
<td>ε</td>
<td>ε</td>
<td>error</td>
</tr>
</tbody>
</table>

bool S: {
    switch token {
        case a:  token := next_token( );
                  call S( );
                  if (token == b) {
                    token := next_token( );
                    return true;
                  } else
                    return false;
        case b:
        case eof:  return true;
        break;
        default: return false;
    }
}

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>eof</th>
<th>other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Next Lecture

Next Time:

- Review of LL(1) parsing and syntax directed translation
- Read Scott, Chapter 2.3.1 - 2.3.2