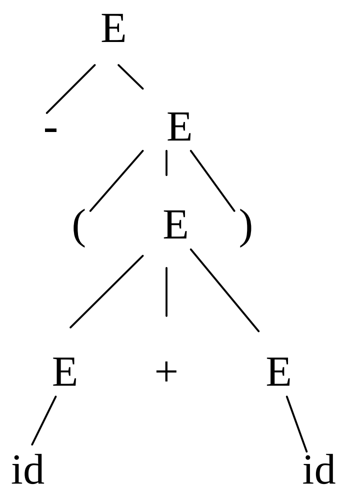
Lecture 5

Syntactic Analyzer (Parser)

Every programming language has rules that prescribe the syntactic structure of well formed programs. The syntax of programming language constructs can be described by context free grammars. In syntax analysis we are concerned with groping *tokens* into larger syntactic classes such as *expression* , *statements* , and *procedure*. The syntax analyzer (parser) outputs a *syntax tree*, in which its leaves are the *tokens* and every non-leaf node represents a syntactic *class* type. For example:-
Consider the following grammars:-

$$E \longrightarrow E+E \mid E*E \mid (E) \mid -E \mid id$$

Then the parse tree for **-(id+id)** is:-



<u>Syntax Error Handling</u> :-

Often much of the error detection and recovery in a compiler is central around the *parser*. One reason of this is that many errors are syntactic in nature. Errors where the token stream violates the structure of the language are determined by parser, such as an arithmetic expression with unbalanced parentheses.

<u>Derivations</u> :-

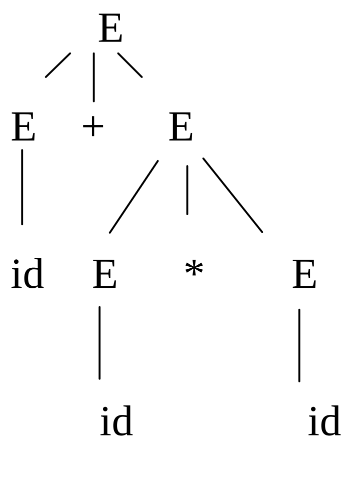
This derivational of view gives a precise description of the *top-down* construction of *parse tree*. The central idea here is that a *production* is treated as rewriting rule in which the *nonterminal* on the left is replaced by the string on the right side of the *production*. For example, consider the following grammar:

$E \rightarrow E+E$
$E \rightarrow E*E$
$E \rightarrow (E)$
$E \rightarrow -E$
$E \rightarrow id$

The *derivation* of the input string **id + id* id** is:

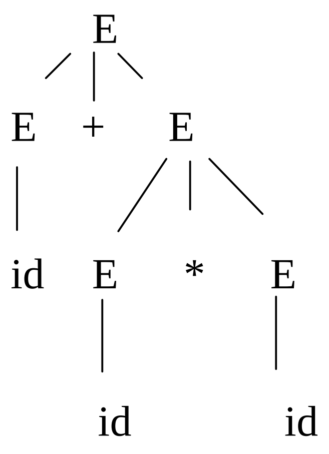
<u>Left-most derivation</u>	***<u>Right-most derivation</u>***
$\underline{E}$	$\underline{E}$
$\underline{E}+E$	$E+\underline{E}$
$id+\underline{E}$	$E+E*\underline{E}$
$id+\underline{E}*E$	$E+\underline{E}*id$
$id+id*\underline{E}$	$\underline{E}+id*id$
$id+id*id$	$id+id*id$

<u>Note</u>:- *parse tree* may by viewed as a graphical representation for a derivation :

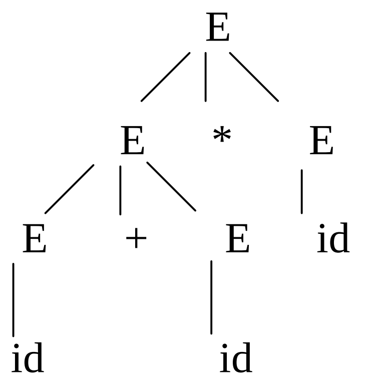


<u>Ambiguity</u> :-

A grammar that produce more that one parse tree for same sentence is said to be ***Ambiguous***. In the another way, by produced more that one *left-most derivation* or more that one *Right-most derivation* for the same sentence.



(1)



(2)

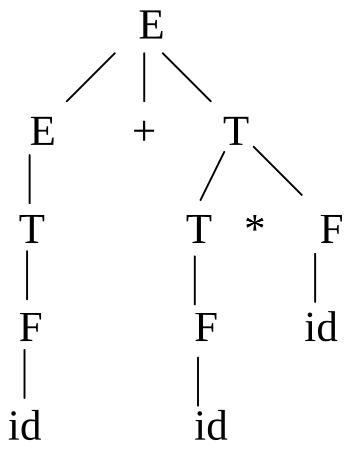
two parse tree for **id+id*id**

E	E
E+E	E*E
id+E	E+E*E
id+E*E	id+E*E
id*id*E	id+id*E
id+id*id	id+id*id
(1)	**(2)**

Two *left-most derivation*s for **id+id*id**

Sometimes am ambiguous grammar can be rewritten to eliminate the ambiguity. Such as:

$$\begin{array}{l} E \rightarrow E+E \\ E \rightarrow E*E \\ E \rightarrow (E) \\ E \rightarrow -E \\ E \rightarrow id \end{array} \quad \Rightarrow \quad \begin{array}{l} E \rightarrow E+T \mid T \\ E \rightarrow (E) \\ E \rightarrow -E \\ T \rightarrow T*F \mid F \\ F \rightarrow id \end{array}$$



parse tree for **id+id*id**

<u>Left-Recursion</u> :-

A grammar is left-recursion if has a *nonterminal* **A** such that there is a derivation $A \longrightarrow A\alpha$ for some string α . Top-down parser cannot handle *left-recursion* grammars, so a transformation that eliminates left-recursion is needed:

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

$$A \rightarrow \beta A'$$
$$A' \rightarrow \alpha A' \mid \varepsilon$$

OR:

$A \rightarrow A\alpha 1 \mid A\alpha 2 \mid \ldots\ A\alpha m \mid \ldots \mid\ ß1 \mid\ ß2 \mid \ldots \mid\ ßn$

$A \rightarrow\ ß1A' \mid ß2\ A' \mid \ldots \mid\ ßn\ A'$

$A' \rightarrow \alpha 1\ A' \mid \alpha 2\ A' \mid\ \ldots \mid \alpha m A' \mid\ \varepsilon$.

Example:

$E \rightarrow E+T \mid T$

$T \rightarrow T*F \mid F$

$F \rightarrow (E) \mid id$

$E \rightarrow T\ E'$

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

$T \rightarrow FT'$

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

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

<u>Left-Factoring</u> :-

The basic idea is that when is not clear which of two alternative production to use to expand a *nonterminal* A . We may be able to rewrite the A-productions to defer the decision until we have seen enough of the input to make the right choice.

$A \rightarrow \alpha\ ß1 \mid \alpha\ ß2$ where $\alpha \neq \epsilon$

$\mathbf{A \rightarrow \alpha\ A'}$

$\mathbf{A' \rightarrow ß1 \mid\ ß2}$

OR :-

$A \rightarrow \alpha ß1 \mid \alpha ß2 \mid .. \mid \alpha ßn \mid \gamma$ **where $\alpha \neq \epsilon$**

$A \rightarrow \alpha A' \mid \gamma$

$A' \rightarrow ß1 \mid ß2 \mid .. \mid ßn$

Example:

$S \rightarrow iEtS \mid iEtSeS \mid a$
$E \rightarrow b$

$S \rightarrow iEtSS' \mid a$

$S' \rightarrow eS \mid \epsilon$

$E \rightarrow b$

Easy Come , Easy Go