

Unit-IV (Phases of 2nd, 3rd & 4th) Syllabus

I-Syntax Analysis:

Introduction to types of parsing Techniques.

Top-down parsing.

Recursive descent Parsing.

Predictive Parsing : LL(1).

Bottom-up Parsing : SLR, CLR, LALR.

II Semantic Analysis (SA)

Introduction

Syntax directed definition (SDD).

Syntax directed Translation

Attributes , Types of attributes

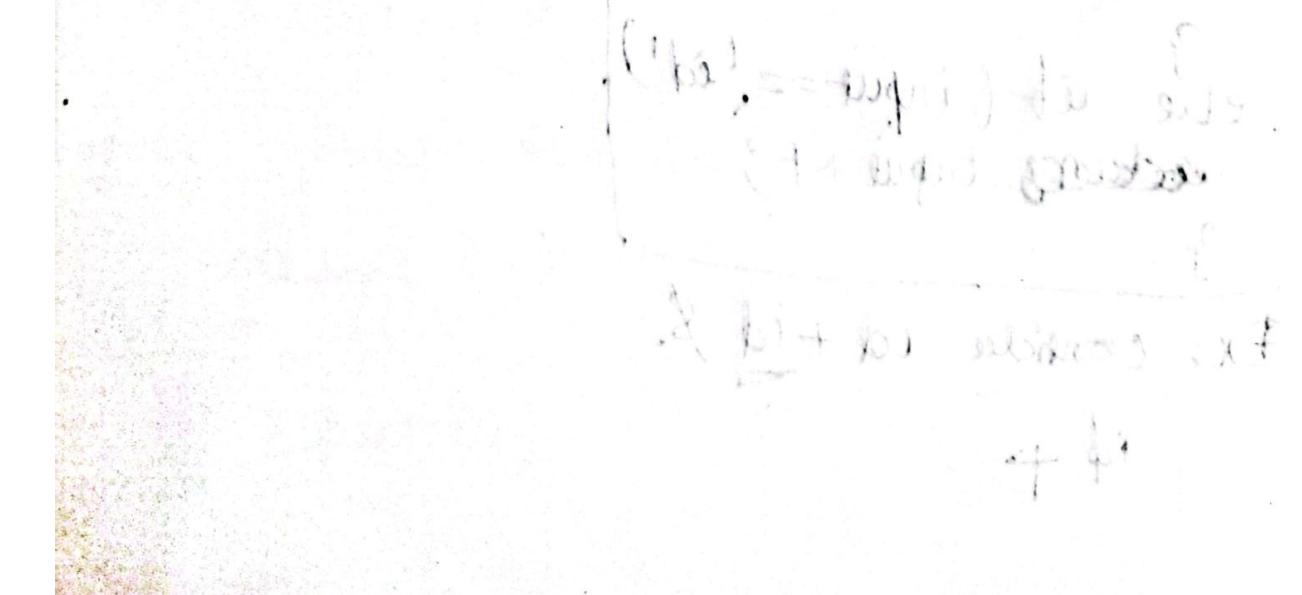
Bottom-up evaluation of attributes.

III Intermediate Code Generation (ICG)

Types of Intermediate code

Type of 3 address codes

Evaluation of 3-address code



Unit - IV

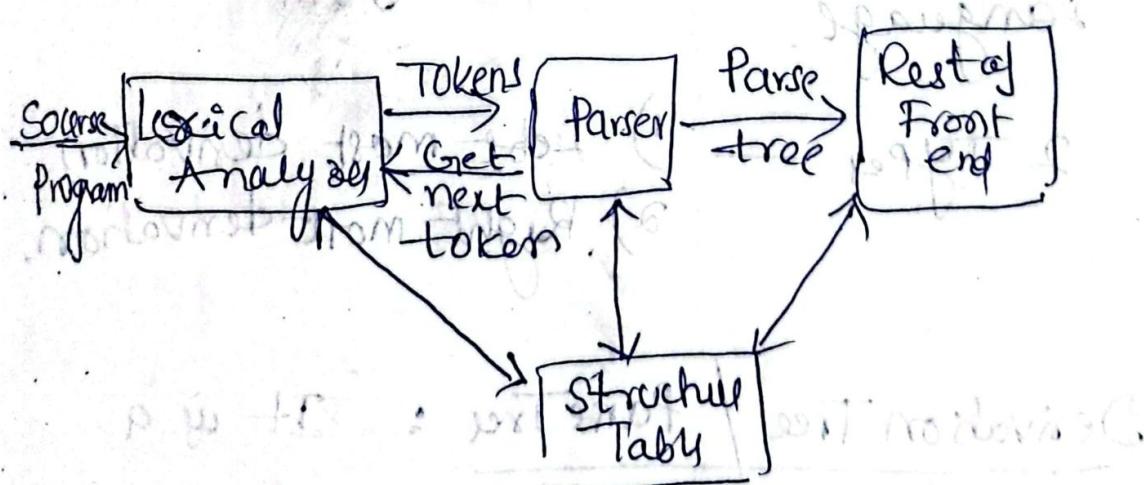
chapter - II . Syllabus : (II phase of compiler)
Syntax Analysis : Types of parsing,

Recursive Descent Parsing, Predictive Parsing,
Bottom up Parsing SLR, CLR, LALR.

Syntax Analysis is the second phase of
the compiler.

It can also be called Syntax verification
or parsing or Parse tree generator.

The parser constructs the parse tree
and passes to the rest of the compiler
for further processing.



Context Free Grammar

VI

CFG has 4 components $G = \{V, T, P, S\}$

V — set of Non-Terminals /
Variables in uppercase letters.

T — Set of Terminals ($V \cap T = \emptyset$)
Represented in lowercase / digits.

P — Production Rules

S — Start symbol.

Derivation : The process of deriving a

String / sentence from a grammar is
called derivation.

If generates strings / sentence of

Language

2 Types

- 1) Left most derivation
- 2) Right most derivation.

Derivation Tree / Parse Tree : It is a

Graphical representation of the derivation
of the production rules for a given
CFG.

Left most DT

Right most DT

Derivation from left to right

Derivation from right to left

Q: $E \rightarrow E+E$

$E \rightarrow E*E$

$E \rightarrow a/b/c$

derive the string $a+b*c$

Terminals = $(+, *, /, a, b, c)$

NonTerminal = E

Parse tree

$E \rightarrow E+E$

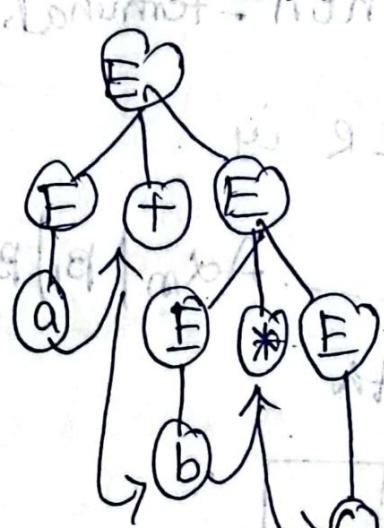
\downarrow

$a+E$

$a+E*E$

$a+b*E$

$a+b*c$



Read Left to Right

$a+b*c$

$E \rightarrow E+E$

\downarrow

$E \rightarrow E+E*E$

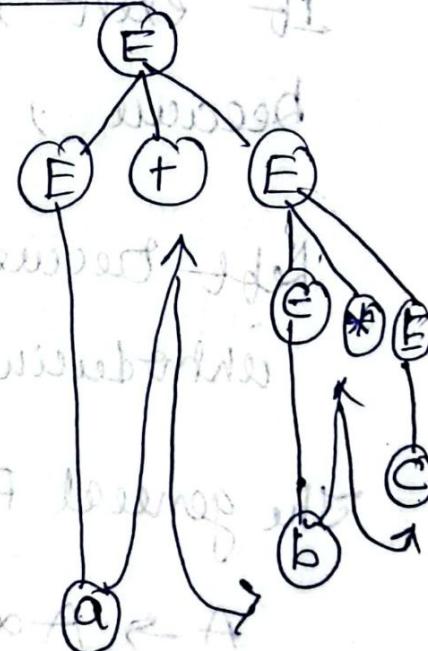
\downarrow

$E \rightarrow E+E*c$

\downarrow

$E \rightarrow a+b*c$

Parse tree



Read from Left to Right

$a+b*c$

Ambiguous Grammars:

A CFG is said to be ambiguous if there exist more than one derivation tree for given input string.
i.e More than one left-most derivation
or More than one right-most derivation

Removing Left Recursion (LR)

Left recursion occurs during parsing
If exist, remove from grammar
because ; it can create infinite loop

Left Recursion can be eliminated by introducing new non-terminal.

The general form of ELR is

$$A \rightarrow [A\alpha_1 | A\alpha_2 | \dots | A\alpha_m | B_1 | B_2 - \dots]$$

Can be replace with

$$\boxed{A \rightarrow B A' \\ A' \rightarrow \alpha A' / \epsilon}$$

$$E \rightarrow E + T / T \quad \text{then} \\ \underbrace{\quad}_{A} \quad \underbrace{\alpha}_{\alpha} \quad \underbrace{\quad}_{B} \quad \underbrace{\beta}_{\beta} \quad \leftarrow \quad \leftarrow$$

$$\text{Top: } A \rightarrow A \alpha / \beta \text{ then} \\ \begin{cases} A \rightarrow B A^1 \rightarrow E \rightarrow T E^1 \\ A^1 \rightarrow \alpha A^1 / \beta \rightarrow E^1 \rightarrow E + T / \epsilon \end{cases}$$

Left Factoring (LF)

Left Factoring transforms the grammar to make it useful for Top-down parsers or Predictive Parsers.

or
Left factoring is a process by which the grammar with common prefixes is transformed to make it useful for Top-down parsers.

Grammar

$$A \rightarrow \alpha B_1 + \alpha B_2 + \dots + \alpha r_1 r_2 \dots$$

The equivalence Left Factoring creates

is

$$A \rightarrow \alpha A^1 / \beta$$

$$A^1 \rightarrow B_1 | B_2 | \dots | B_n$$

Ex 1: $S \rightarrow \underline{iEts} / iEts \cdot es / a$

$E \rightarrow b$

Soln: $S \rightarrow iEts / iEts \cdot es / a$

common prefix $\rightarrow iEts$

$A \rightarrow \alpha A^1 / r$ $\xrightarrow{\text{Lett}} S \rightarrow iEts \underline{S} / a$

$A^1 \rightarrow \beta_1 / \beta_2 \dots \beta_n$ $\xrightarrow{\text{Factoring}} S \rightarrow e / es$

At 2nd step \rightarrow $\xrightarrow{\text{Lett}} E \rightarrow b$

Ex 2: $A \rightarrow aAB / aA$

$B \rightarrow bB / b$

Soln:

$A \rightarrow aAB / aA \xrightarrow{\text{Lett}} A \rightarrow aA^1$

$\xrightarrow{\text{Factoring}} A^1 \rightarrow AB / A^1$

$\xrightarrow{\text{Lett}}$

$B \rightarrow bB / b \xrightarrow{\text{Lett}} B \rightarrow bB^1$

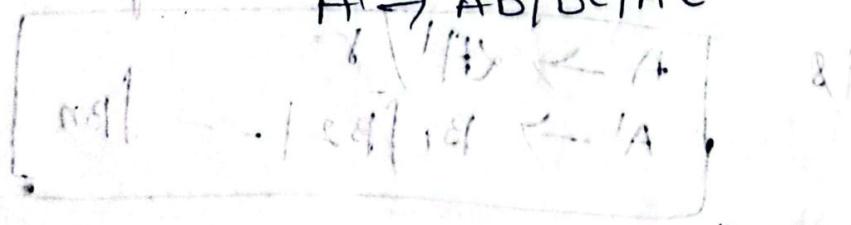
$\xrightarrow{\text{Factoring}} B \rightarrow B / e$

Ex 3: $A \rightarrow aAB / aBc / aAC$

Soln:

$A \rightarrow \alpha A^1$

$A^1 \rightarrow AB / BC / AC$



D.R.K. Bhargavi

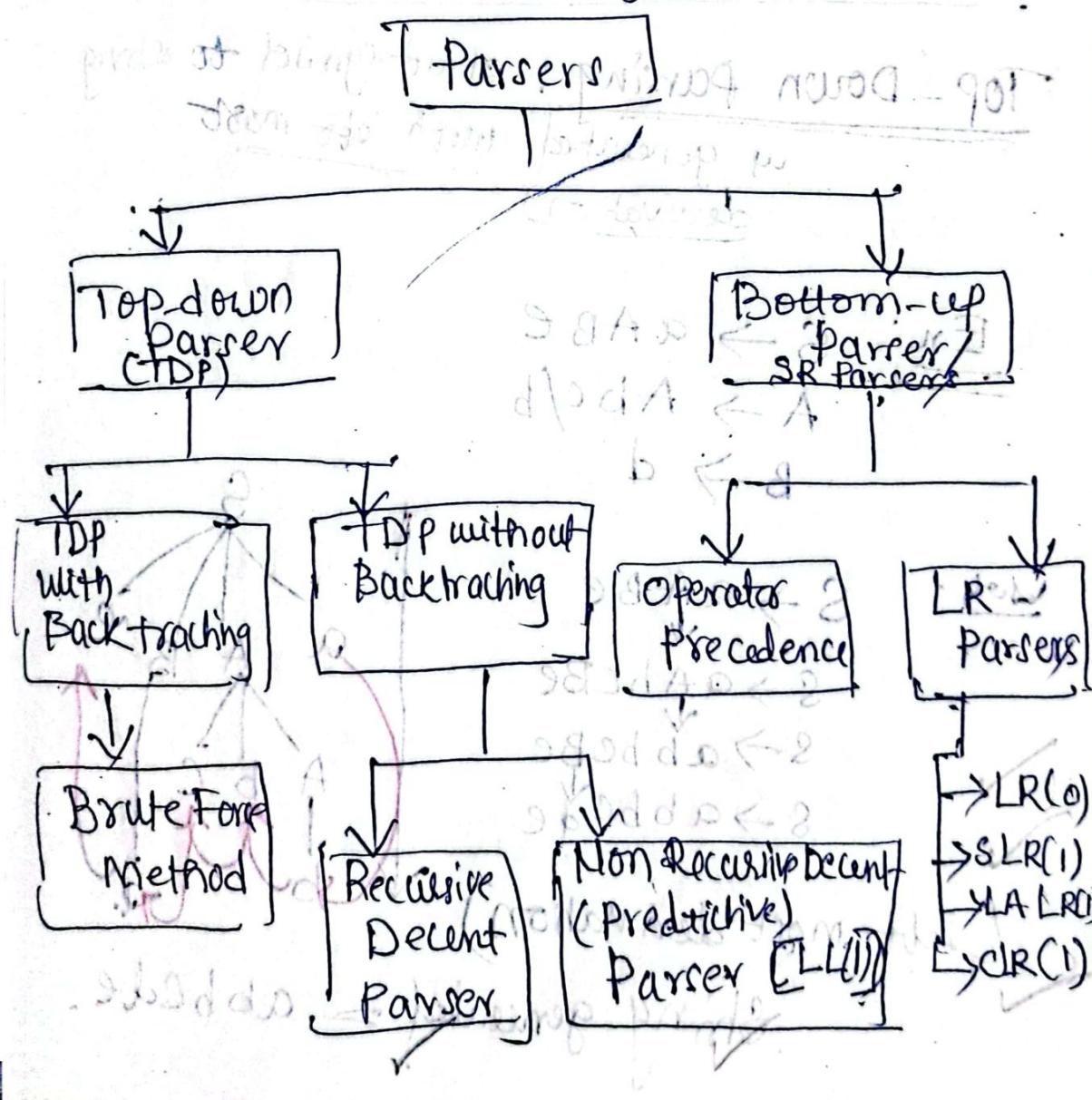
Types of Parsing:

The task of Parser is to determine if and how the input can be derived from the start symbol within the rules of a Formal Grammar.

2 types of Parsers:

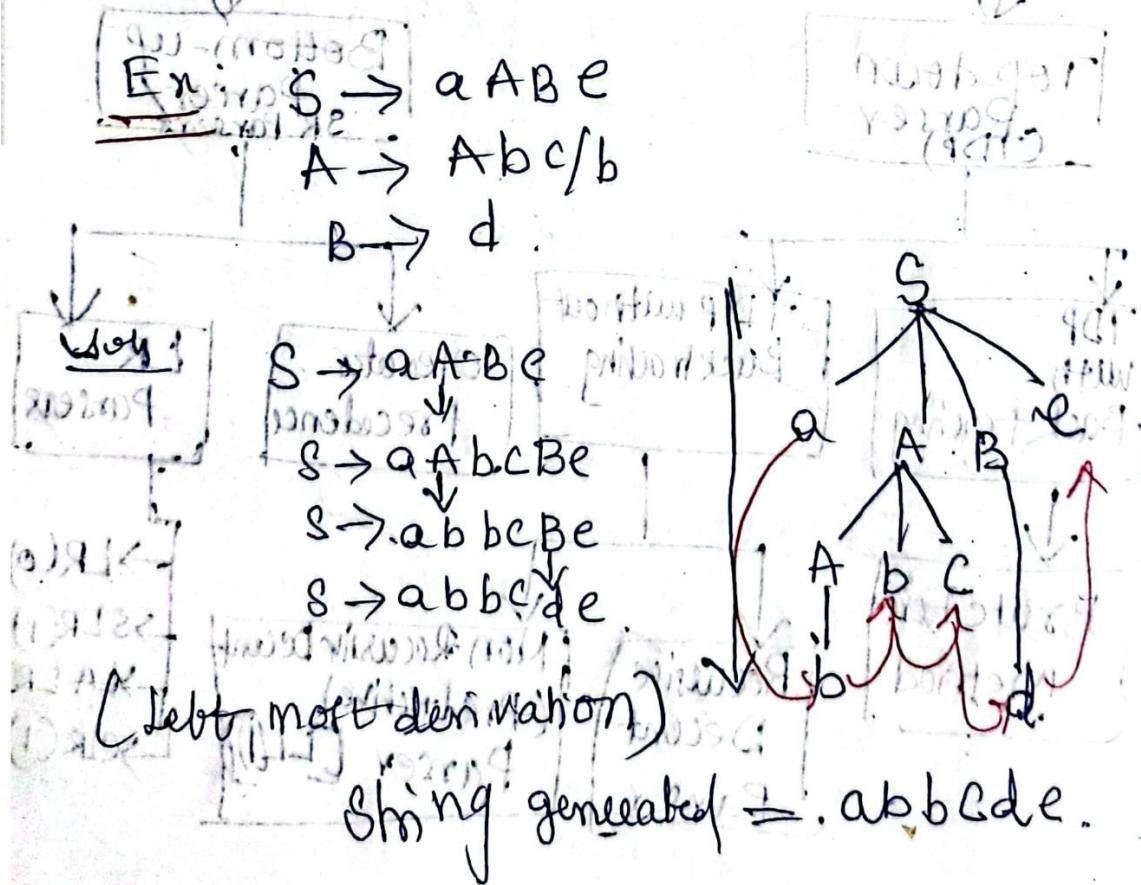
① Top-down parsing

② Bottom-up parsing



Grammar	Topdown Parser	Bottom up Parser
Ambiguous.	X	X
Unambiguous	✓	✓
LR	X	✓
RR	✓	✓
Non-Deterministic	X	✓
Deterministic	✓	✓

Top - Down Parsing: start symbol to string is generated with left most derivation



Bottom up parse tree: Bottom-up parsing starts from the leaf nodes of a tree and works in upward direction till it reaches the root node. String considered and start symbol generated by using Right most derivation.

$$\text{Ex.: } S \rightarrow a A B e$$

$$A \rightarrow A b c / b$$

$$B \rightarrow d$$

Right most derivation of given string

Grammar is

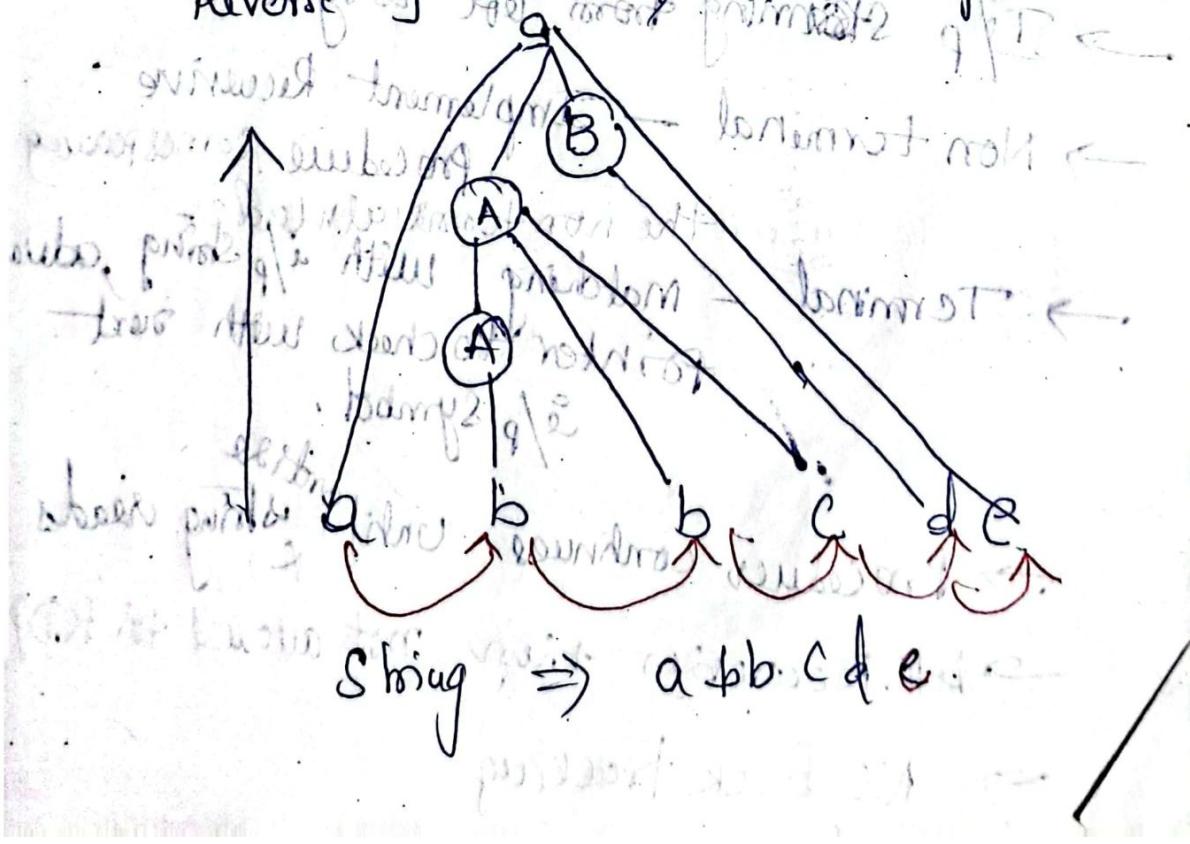
$$S \rightarrow a A B e$$

$$S \rightarrow a A d e$$

$$S \rightarrow a A b c d e$$

$$S \rightarrow a b b c d e$$

Reverse of RMD is Bottom up parser



Recursive Descent parser (RDP) : (Topdown Parser)

A Recursive descent parser is a topdown parser.

This parser is simple to implement and is suitable for LL(1) grammars, where decision can be made based on a single lookahead token.

- RDP uses recursion to analyse the syntax of a programming language based on a predefined grammar.
- Starts from the start symbol of the grammar & recursively calls function for each Non-terminal symbol, matching it against the production rule.
- RDP is used in lexical analysis (scanning) & syntax analysis (parsing) phases of compiler.
i.e 1 & 2 phases of compiler.

Recursive Descent Parser (RDP)

The parser uses collection of recursive procedures for parsing the given input string is called RDP.

- CFG is used to build recursive routine.
- RHS of Production rule is converted into program.
- Parsing-table is not constructed.
- First & Follow is not used.

-
- Procedure is associated with each non-terminal of the grammar
 - Input scanning from left to right
 - Non terminal → Implement Recursive procedure corresponding to the non-terminal made
 - Terminal → matching with i/p string, advance pointer to check with next i/p symbol.
 - Procedure continues until string reads entire.
 - Left Recursion then not allowed for RDP.
 - NO back tracking

Recursive decent parser (RDP).

Q1). construct the Recursive decent Parser
for the following grammar.

$$E \rightarrow iE' \\ E' \rightarrow +iE'/c$$

Soln: Given grammar is

$$E \rightarrow iE' \\ E' \rightarrow +iE'/c$$

Main()

{
 E()
 if input == '\$')

Prints ("Parsing successfully done");

y.

E()

```
{ if (input == i)
    input++;
    E'();
```

}

E'()

```
{ if (input == '+') .
```

```
{     input++;
    }
```

```
if (input == 't')
```

```
    input++;
    
```

```
E'();
```

}

else

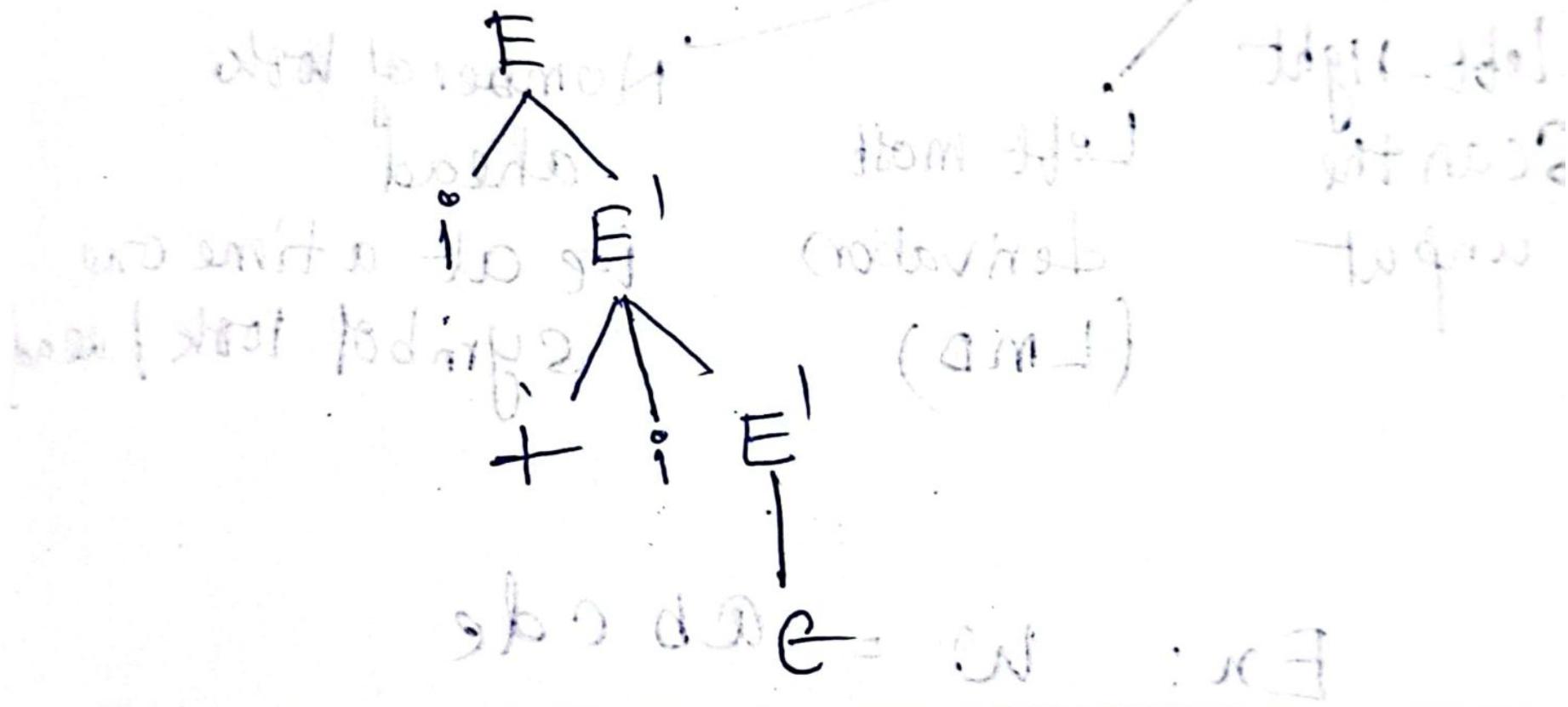
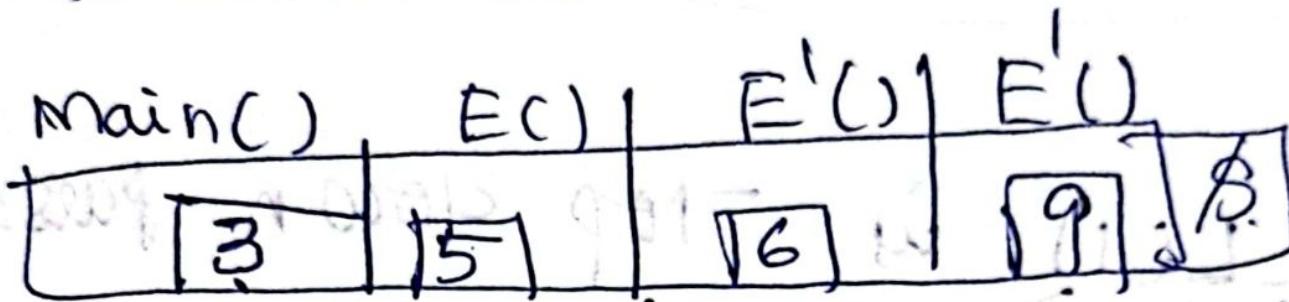
```
    return;
```

}

L

Consider the input string w. construct RDP

$$w = 1 + 1\%.$$



RDP (Recursive decent Parser)

construct the recursive decent parser for the grammar

$$S \rightarrow 01S'$$

$$S' \rightarrow 0S1SS' / \epsilon$$
 & generate the string
or
$$S \rightarrow S0S1S | 0$$

solution: Left recursion having
removing it

① main()

```
{  
    S();  
}
```

```
if (input == '$')  
{ prints ("Parsing successful"); }
```

③ S'()

```
{  
    if (input == '0')  
        input++;
```

```
    if (input == '1')  
        input++;
```

```
    if (input == '0')  
        S();
```

```
    if (input == '1')  
        input++;
```

```
    S();
```

else
return; // fail

```
}
```

```
}
```

```
}
```

```
}
```

```
}
```

```
}
```

```
}
```

```
}
```

```
}
```

```
}
```

```
}
```

```
}
```

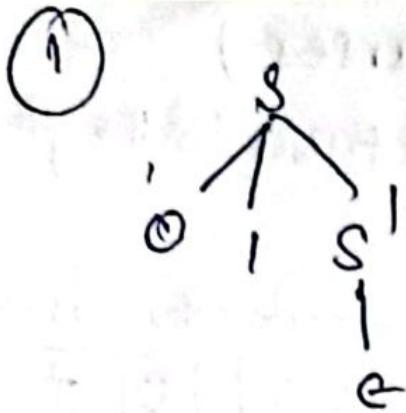
```
}
```

```
}
```

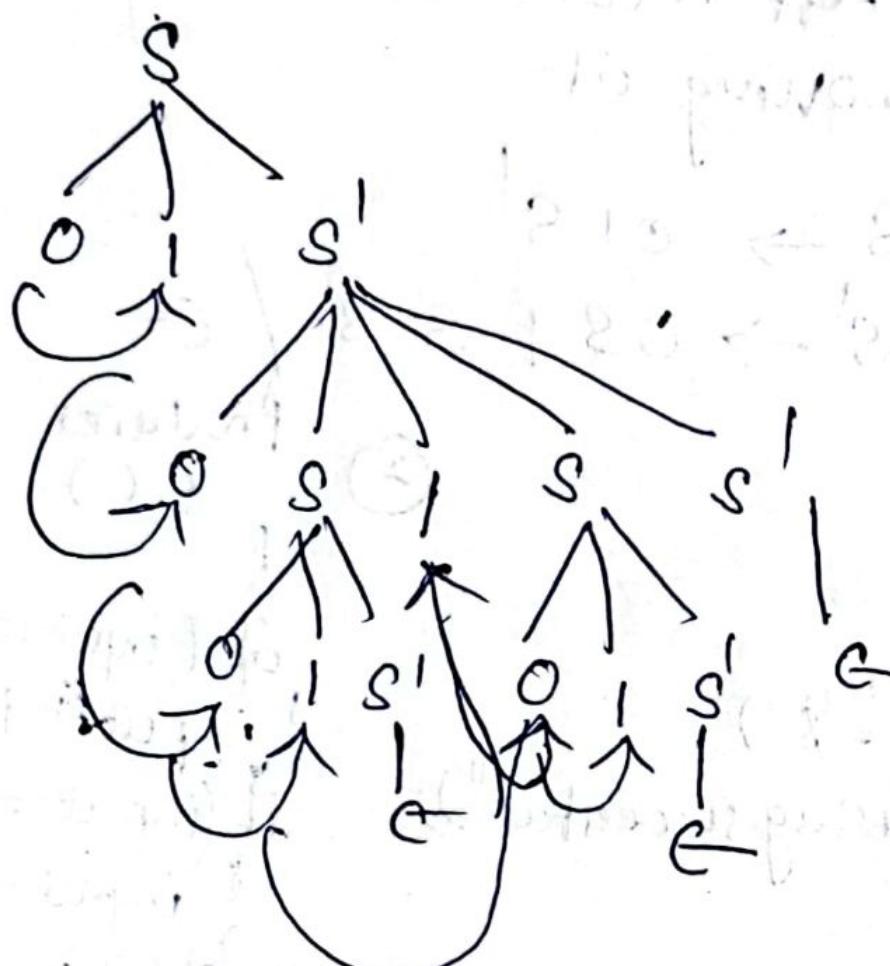
```
}
```

```
}
```

```
}
```



01 generated.



$01001101 \Rightarrow \text{no. of } 0's = \text{No. of } \leq$

LL(1) Parser or Predictive Parser

LL(1) is top down parser.

Left-right
Scan the
input

Left-most
derivation
(LMD)

Number of look
ahead
i.e. at a time one
symbol look/ahead

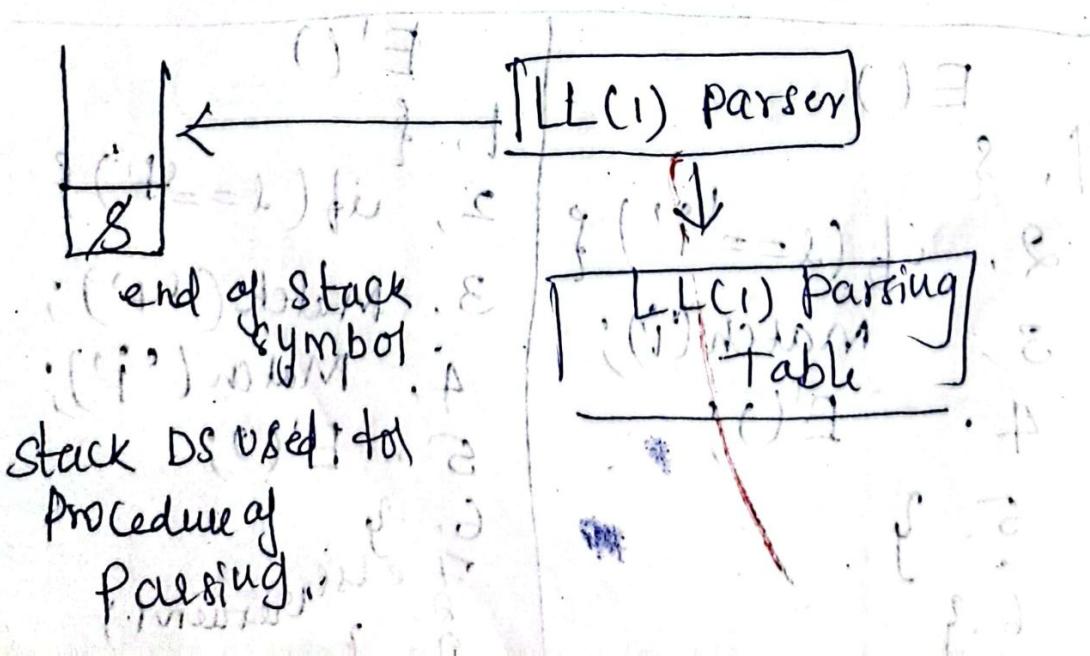
Ex: $w = ab cde$

String scan from left to right

a b c d e t

↑ end of the string always represented with

(i) (0 1 2 3 4 5 6 7 8 9)



First() & Follow()

Ex: 1

$$S \rightarrow A B C D$$

$$A \rightarrow b$$

$$B \rightarrow c$$

$$C \rightarrow d$$

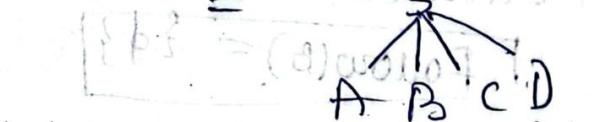
$$D \rightarrow e$$

Find first & follow functions for above production rules.

Soln:

Find out first & follow for LHS

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



$$\{S\} = \{b\} + \{c\} + \{d\} + \{e\}$$

if it is terminal \rightarrow consider it

if it is non-terminal \rightarrow check for

concern PR write the first of it.

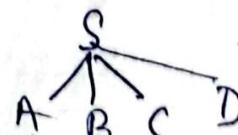
$$\therefore \text{First}(S) = b.$$

(Ex 2) $S \rightarrow A B C D$

$$A \rightarrow b/e$$

$$B \rightarrow c$$

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



if $b \rightarrow \text{First}(S) = b$

$$e \rightarrow S \rightarrow \underline{B C D}$$

$$\text{In this case } \text{First}(S) = C$$

First(A) = {b}

First(B) = C

First(C) = {d}

First(D) = {e}

Follow(S) : $S \rightarrow A B C D$ $\underline{A B C D}$

Follow(S) = { \$ } }

Follow(A) = First(B, C, D)

Follow(A) = {C, D}

Follow(B) = First(C, D)

Follow(B) = {D}

Follow(C) = First(D) = {e}

Follow(D) = Follow(S) = { \$ }

Always i/p symbol followed by \$.

Follow(C) never contain null or G

* * *

Follow Rules

Terminal — write as it is

Non-Terminal — write its First element.

Left element — write follow of LHS.

First()	Follow()
First(S) = b	Follow(S) = { \$ }
First(A) = b	Follow(A) = { c }
First(B) = c	Follow(B) = { d }
First(C) = d	Follow(C) = { e }
First(D) = e	Follow(D) = { \$ }

Construct the LL(1) table / predictive parse table

	b	c	d	e	\$
S	$S \rightarrow ABCD$				
A	$A \rightarrow b$				
B		$B \rightarrow c$			
C			$c \rightarrow d$		
D				$d \rightarrow e$	$e \rightarrow \$$

{ columns — Terminal
Rows — Non Terminal.

Predictive Parser (LL(1))

Top down Parser

No Backtracking

LL(1) Parser

~~Top down~~ → CFG

* Steps for LL(1) parser

① Elimination of Left Recursion

② Elimination of Left Factoring

③ Find First() and Follow()

④ Construct parse Table

⑤ Parsing input string after

⑥ construct Parse tree

	S	b	a	d	f	e	c	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
S																											
b																											
a																											
d																											
f																											
e																											
c																											
g																											
h																											
i																											
j																											
k																											
l																											
m																											
n																											
o																											
p																											
q																											
r																											
s																											
t																											
u																											
v																											
w																											
x																											
y																											
z																											

Q 4.

10

$$S \rightarrow iEts^* / a$$

$$S \rightarrow eS / e$$

$$E \rightarrow b.$$

$$S \rightarrow iEtS / iEb$$

$$E \rightarrow b.$$

Remove left

Factoring

construct first(), & follow also
construct Pause table of LLC(1).

Why: (i) No left recursion & left factoring

$$(ii) \text{First}(S) = \{i, a\}$$

$$\text{First}(S^*) = \{e, e\}$$

$$\text{First}(E) = \{b\}.$$

$\text{Follow}(S)$ = Follow of starting symbol always \$.

e. Check on RHS of S followed by First of S

$$\text{First}(S^*) = \{e, e\}$$

Now if \$ not allowed in Follow()
Substitute \$ in PR.

$$S \rightarrow iEtS^* \\ \downarrow \text{last element}$$

$\Rightarrow \text{Follow}(S)$

$$\text{Follow}(S) = \{\$\}, \{e\}$$

working pointer : 0101
bit no 2

$$\text{follow}(S) = \text{follow}(S) = \{e, \$\}$$

$$\text{follow}(E) = \{t\}$$

	First	Follow
S	{i, a}	{e, \\$}
S'	{e, C}	{\\$, e}
E	{b}	{t}

② LL(1) Parse table :

	i	a	e	b	t	\$
S	$S \rightarrow iETSS$	$S \rightarrow a$				
S'				$S \rightarrow eS$ $S \rightarrow e$		$S \rightarrow e$
E					$E \rightarrow b$	

Column \rightarrow Terminals

Row \rightarrow NonTerminals

**

If any entry on the consisting more than one entry, then we can say that, it is not in

Note: LL(1) Parser Grammar
The above one having multiple entries.

Q5: Check LL(1) parser for the
following Grammar.

$$S \rightarrow a A B b$$

$$A \rightarrow c / e$$

$$B \rightarrow d / e$$

and also check validity of string acd.

Soln:

(i) No. Left Recursion having in the given grammar.

(ii) Calculation of First().

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

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

$$\text{First}(B) = \{d, e\}$$

Follow(C):

$$\text{Follow}(C) = \{\$\}$$

$$\text{Follow}(A) = \text{First}(B)$$

= {d, e} but e not allowed

substitute e in

$$S \rightarrow a A B b \Rightarrow (6)$$

in 2nd position, Follow(A) = {d, b}

$$(2) \text{Follow}(C, B) = \{b\}$$

in 3rd position, Follow(B) = {e}

in 4th position, Follow(B) = {e}

in 5th position, Follow(B) = {e}

11(i) Parse Table

	Q $S \geq aA B b$	$b.$	c	d	\emptyset
S					
A		$A \rightarrow \emptyset$	$A \rightarrow c,$	$A \rightarrow \emptyset$	
B		$B \rightarrow \emptyset$		$B \rightarrow d.$	

Result table does not contain multiple entries thus given grammar LL(1).

(4) String validation

Stack	Input	Action
\$ \$	a c d b \$	$S \rightarrow a A B b$.
\$ b B A d.	a c d b \$	pop('a')
\$ b A	a c d b \$	$A \rightarrow C b$
\$ b B d	c d b \$	pop('c')
\$ b B	d b \$	$B \rightarrow d$
\$ b d	d b \$	pop('d')
\$ b	b \$	pop('b')
\$	\$	Accepted.

LLC Pulse Train

LR Parsers (Bottom-up P)

Bottom up parsing starts from left nodes of tree & works in upward direction till it reaches root node.
* LR is one type of Bottom up parser.

LR parsers are non-Recursive, Shift-reduce bottom-up parser.

LR parsers are also known as LR(k) parser

where

L \rightarrow left-to-right scanning of i/p Strng

R \rightarrow construction of right-most derivation in reverse

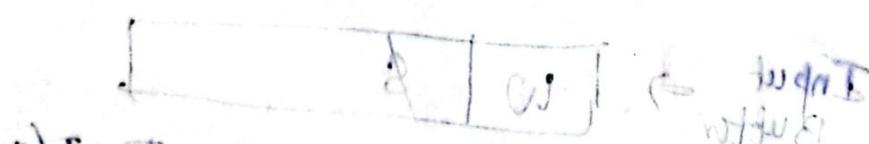
k \rightarrow No. of look heads needed for decision.

LR(0) < SLR(1) < LALR(1) < CLR(1)

simple LR lookahead LR canonical

CLR(1) is most LR powerful parser.

\rightarrow LR parsing is non-Backtracking Shift-reducer parser & make use of DPA



LR(0) parser converts i/p to states (S, A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S)

Steps for constructing the LR Parsing Table

*Check the ambiguity of the Grammar

1. write the augmented grammar
2. LR(0) collection of items to be found.
3. Defining two (2) functions:
 - a) Goto (list of non-terminals)
 - b) Action (list of terminals)
4. construct parse table with Step-3
(goto and actions)

Augment Grammar helps when to stop the parser and acceptance of grammar.
We use dot (.) to scan or proceed symbol after dot.

Terminal (T)

NonTerminal (V)

$F \rightarrow xy_2 \rightarrow$ Ready to scan to x
 $F \rightarrow x_1 y_2 \rightarrow x_1$ is scanned & Ready to scan y_2
 $F \rightarrow x_1 y_2 \rightarrow$ All symbols scanned.

At state tree of state does not have only one of symbol 'x'

SLR(1) Parser:

Simple LR(1) Parser.

Scan from left to right Right most derivation

Pre requisite : LR parser.

SLR is simple LR parser. It is the smallest class of grammar having few no. of states.

SLR(1) is similar to LR(0) parsing.

In LR(0)	SLR(1)
1. There is a chance of conflicts shift-reduce	1. conflict are solved here.
2. Entering "reduce" corresponding to all the terminals (i.e Action part)	2. In SLR(1) "reduce" corresponding to FOLLOW of LHS of the reduced non-terminal.
The conflicts in LR(0) are solved with SLR(1).	
3. zero lookahead	3. one lookahead.

Steps for SLR(1) Parser :

1. writing Augmented grammar

2. LR(0) collection of items to be found.

3. Find FOLLOW of LHS of production.

4. Define 2 functions

① Goto \rightarrow list of terminals

② Action \rightarrow list of terminals

In compiler design, SLR(1) stands for simple left-to-right rightmost derivable.

Parser with 1 symbol looks ahead,

a type of bottom-up parsing method.

It is simple and efficient parser.

SLR(1)
Simple Scans i/p Rightmost derivation in reverse
Left-to-right consider the next input symbol (looked to determine the parsing action)
Shift Reduce

Shift — Push the current i/p symbol to stack & move to the next state

Reduce — If current stack symbol match in prod replace with non-terminal symbol of production & push int stack, then move into corresponding state

Eg2: construct the SLR(0) parsing table for the given grammar

$$S \rightarrow AA$$

$$A \rightarrow a A / b$$

Soln:

1. constructing augmented grammar.

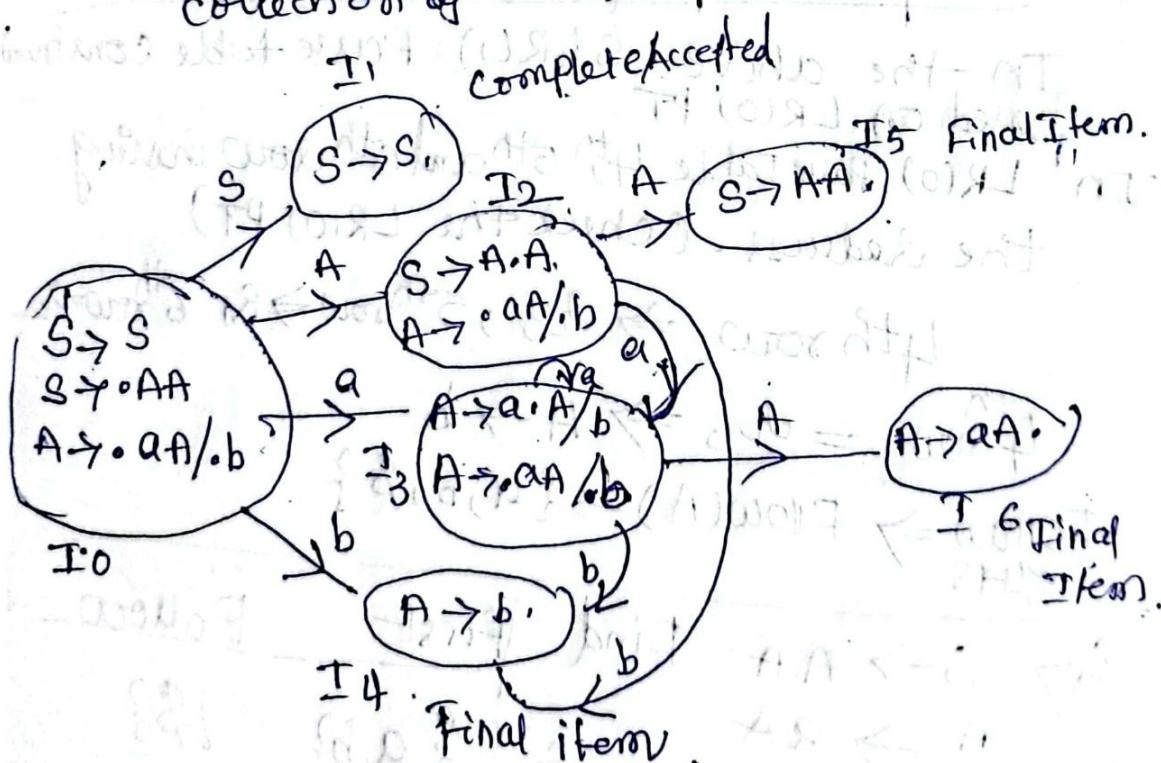
$$S' \rightarrow S$$

$$S \rightarrow \cdot AA - I$$

$$A \rightarrow \cdot a A - 2$$

$$A \rightarrow \cdot b - 3$$

2. computing LR(0) items / canonical collection of LR(0) items



Step 9 : Construct SLR(1) Parse Table

States	Action Part			Goto part	
	a	b	\$	S	A
s ₀					
s ₃	s ₄			1	2
			Accept		
2	s ₃	s ₄			5
3	s ₃	s ₄		1	2
4	R ₃	R ₃	R ₃	A	
5			R ₁		A
6	R ₂	R ₂	R ₂		

In the above SLR(1) Parse table constructed on LR(0) PT.

In LR(0) Parse Table 4th, 5th and 6th row having the reduce. (Check the LR(0) PT)

4th row \Rightarrow S₃, 5th row \Rightarrow R₁, 6th row

4th row = S₃ \Rightarrow A \rightarrow b

FOLLOW \Rightarrow FOLLOW(A) = {a, b, \$}

\therefore S \rightarrow AA: Find FIRST FOLLOW

A \rightarrow aA

A \rightarrow b

S	{a, b}	{\$}
A	{a, b}	{a, b, \$}

Place r_2 in a, b, & columns.

5th row = r_1 and r_2 in a, b, & columns
 $S \rightarrow AA$
Follow LHS = Follow(S) = $\emptyset \rightarrow$ place it in only in $\$$ column.

6th row = $r_2 \Rightarrow A \rightarrow aA$
Follow RHS = Follow(A) = {a, b, $\$$ }

place r_2 in a, b & $\$$ columns.

2Q: Create or construct the SLR(0) parse table for the following grammar.

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

$$T \rightarrow id$$

or Check the given grammar is in LR(0) or not. (Answer till step 2)

Solu:

$$E \rightarrow T + E$$

$$E \rightarrow T$$

$$T \rightarrow id$$

Identify the Non-terminal of LHS from production rule

i.e E & write the
1) Augmented Grammar

$$E \rightarrow .E \quad \textcircled{0}$$

$$E \rightarrow .T + E \quad \textcircled{1}$$

$$E \rightarrow .T \quad \textcircled{2}$$

$$T \rightarrow .id \quad \textcircled{3}$$

$$E' \rightarrow .E$$

After dot(.)

$$E \rightarrow .T + E$$

if Terminal \rightarrow write all

$$E \rightarrow .T$$

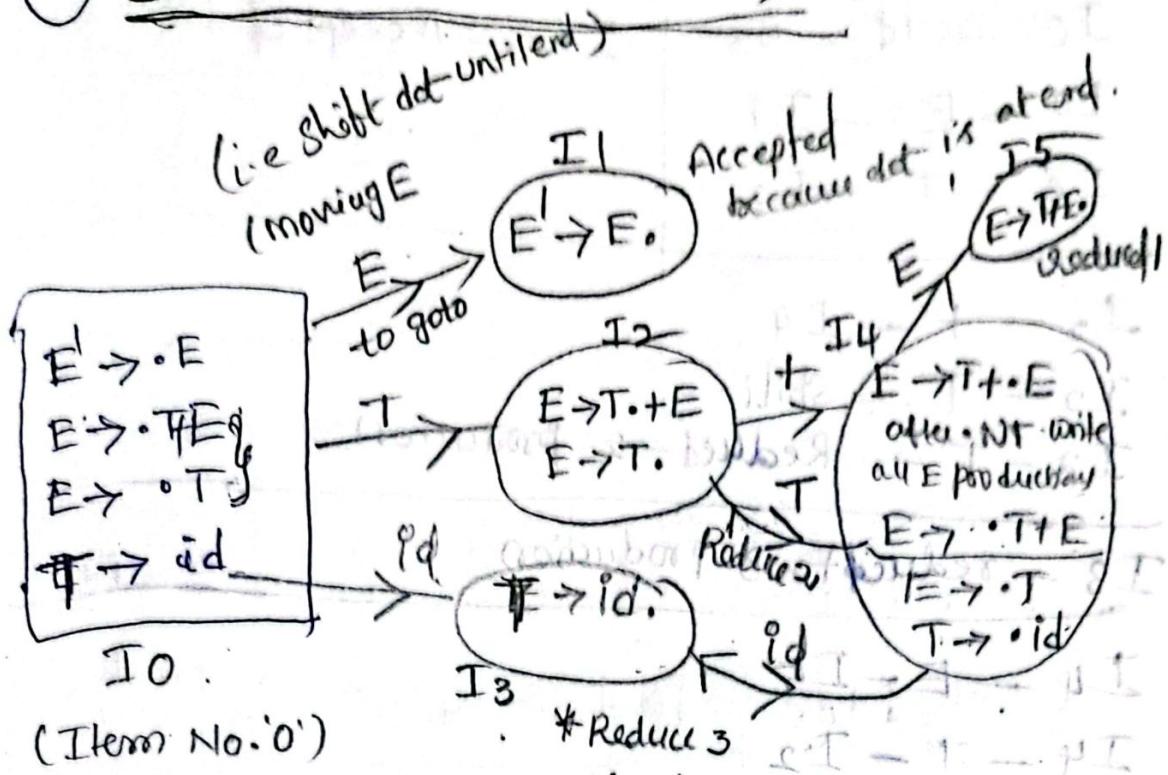
Non-Terminal \rightarrow open the

$$T \rightarrow .id$$

concern production

Rule.

2) Canonical items of LR(0) :



③ -

LR(0) Parse Table :

State	Action Part		Goto Part		*
	id	the	E	T	
0 (I0)	S3	-	-	1	2
1 (I1)	-	-	Accepted	-	-
2 (I2)	S2	S4	v2	-	-
3 (I3)	S3	S3	v3	-	-
4 (I4)	S3	(0)	-	5	2
5 (I5)	S2	w2	v2	-	-

(0) $\rightarrow L$ not ...

... not L^* S_1

$I_0 - id - S_3$	$I_1 - Accepted$
$I_0 - E - I_1$	
$I_0 - T - I_2$	

$I_2 - + - I_4$

$I_2 - T - \text{Shift}$

$I_2 - T - \text{Reduced} \rightarrow 2 \text{ production}$

$I_3 - \text{reduced to } 3 \text{ production}$

$I_4 - E - I_5$

$I_4 - T - I_2$

$I_4 - id - \text{Shift } I_3$

$I_5 - \text{Reduced to } 1 \text{ Production}$

* Note

If $LR(0)$ Parse Table having the following then it is not in $LR(0)$ Parse

1) Shift - Reduce Conflict

2) Reduce - Reduce Conflict

In this example $LR(0)$ Parse table

consisting Shift - Reduce Conflict

i.e $S_4/S_2 \therefore \text{Not in } LR(0)$
Parser.

Solv:

Canonical Items of SLR(1) = LR(0)

only the difference is Parse Table.

States	id	+	\$	E	T
0 (I ₀)	S ₃	-	-	1	2
1 (I ₁)		-	-	Accept	
2 (I ₂)	S ₄	S ₂	S ₂		
3 (I ₃)	S ₃	S ₂	S ₂		
4 (I ₄)	S ₃	-	-	5	2
5 (I ₅)	-	-	S ₂		

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

$$\text{Follow}(T) = \text{Follow}(E) \cup \{ + \}$$

$$= \{ +, \$ \}$$

consider "Reduce" from LR(0).

2nd, 3rd & 5th Rows containing Reduce.

2nd Row $\Rightarrow \text{v}_2$

i.e $E \rightarrow T$

LHS w/ E, $\therefore \text{Follow}(E) = \{ +, \$ \}$

\therefore Write v_2 only under $\$$ column

3rd row = v_3 .

$T \Rightarrow id$

LHS $\Rightarrow T \Rightarrow \text{Follow}(T) = \{ +, \$ \}$

Write v_3 under $+, \$$.

5th row = v_5

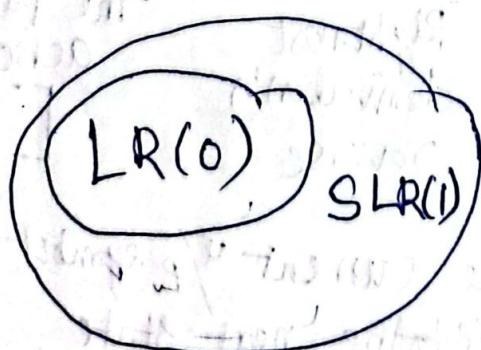
$E \rightarrow T+E \Rightarrow \text{LHS } E \Rightarrow \text{Follow}(E) = \$$

i.e. write v_5 under $\$$

Remaining Goto part w/ same as LR(0).

SLR(1) Not going to have conflicts

i.e



CLR(1) \Leftrightarrow Canonical Left Right Parser.

CLR refers to canonical lookahead.
→ Type of Bottom up Parser
→ CLR Parsing uses canonical collection of LR(1) items to build the CLR(1) parsing table.

→ CLR(1) parsing table produces the more number of states compare to SLR(1).

→ In CLR(1) we place the reduce mode only in the lookahead symbols.

CLR(1) \Rightarrow Using canonical Items of LR(1).
Canonical scanning from left to Right

$$\boxed{\text{LR}(1) \text{ Items} = \text{LR}(0) \text{ items} + \text{a lookahead head.}}$$

Note: In CLR(1) & LALR(1) Using Single lookahead.

Steps for CLR(1):

1) Write the argument grammar & add the lookahead.

2) Find canonical collection of LR(1) items.

3) Construct the CLR(1) Parse table by considering lookahead symbols.

(Place Shift/Reduce in lookahead symbol).

Steps

1. For given input string write a context free grammar.
2. Check the ambiguity of grammar.
3. Add augment production in the given grammar & add lookahead.
4. Create canonical collection of LR(0) items.
5. Draw a state transition diagram.
6. Construct CLR(1) parsing table.

⑧ Example: construct the CLR(1)

* parse-table for the following grammar

$$S \rightarrow AA$$

$$A \rightarrow aA/b$$

You:

$$S \rightarrow AA$$

$$A \rightarrow aA$$

$$A \rightarrow b$$

① Add Argument production.

Insert dot (.) symbol at the 1st position
for every production in G.

And also add the lookahead i.e (\$)

$S' \rightarrow \cdot S, \$$ - (Starting symbol of G)

$S \rightarrow \cdot A A, \$$ - (Production S)

$A \rightarrow \cdot a A ; a/b$ - (Note: $S \rightarrow \cdot A (A, \$)$)

$A \rightarrow \cdot b ; a/b$ - (After A, it is

First(A))

$= \{a, b\}$.

$S \rightarrow \cdot A A, \$$ - (After dot) : - (Add in A's products)

② Note: $LR(1) = LR(0) + \text{all lookahead}$

Initial symbol / a lookahead is \$.

1. $S' \rightarrow \cdot S, \$$ (Initial state lookahead is \$)

2. $S \rightarrow \cdot A A, \$ \rightarrow L$ (look head fit previous one).

3. $A \rightarrow \cdot a A, a/b$

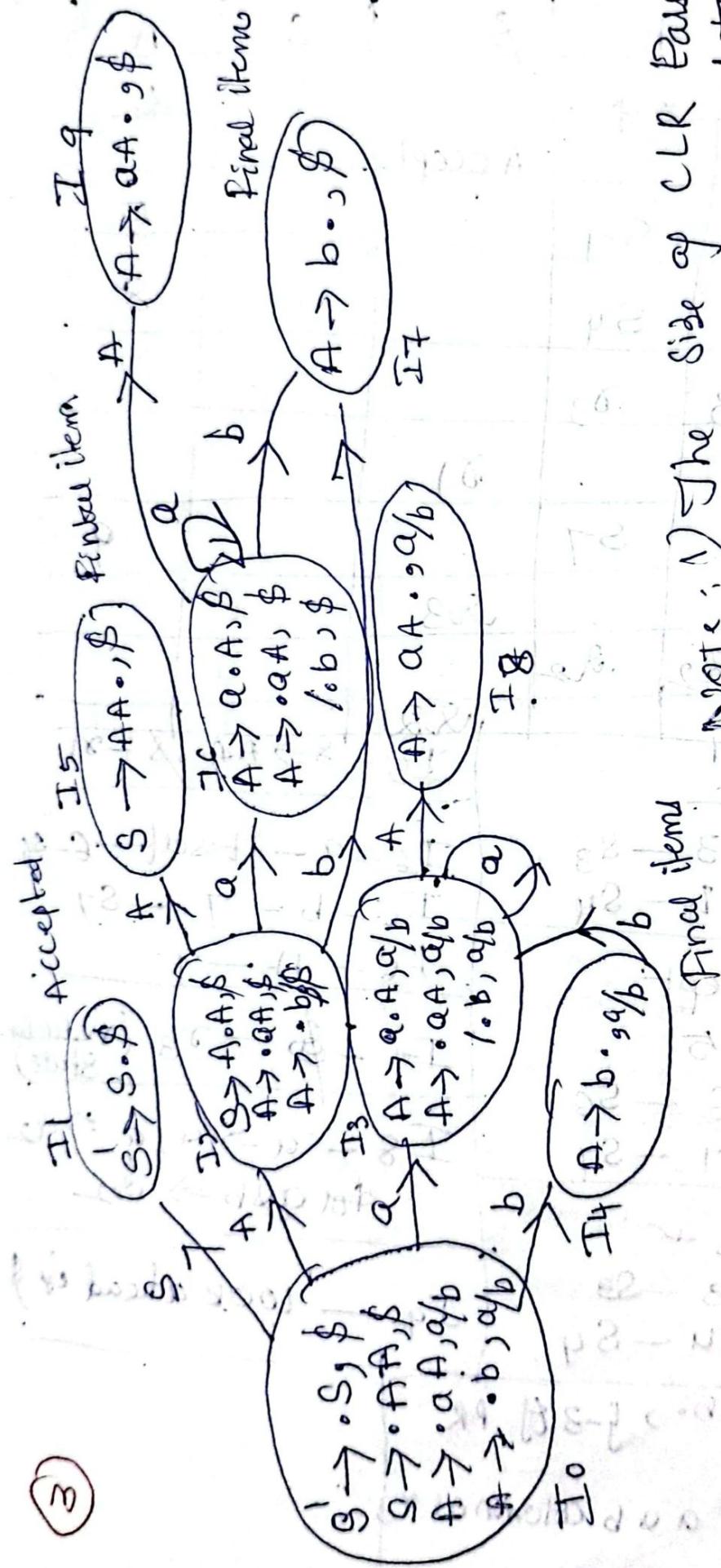
4. $A \rightarrow \cdot b ; a/b \rightarrow L$ (First(A, \$))

here: same symbol of

LHS either (S)

same lookahead

for 4 production



Note: The size of CLR Parse table is quite large compared to other so, it occupies more space in memory. CLR avoids conflicts in parsing table.

Note:

States	Action - Terminal					Goto -
	a	b	\$	s	A	
0	s3	s4		1	2	
1				Accept		
2	s6	s7				5
3	s2	s4				8
4	s3	s3				
5				s1		
6	s6	s7				9
7				s3		
8	s2	s2				
9				s2		
I0 - s - 1				I5 - \$ → AA, s - s1		
I0 - A - 2						
I0 - a - 3 - s3				I6 - a - it self → 6 - s		
I0 - b - 4 - s4				I6 - b - 7 - s7		
I1 - Accept				I6 - A - 9		
I2 - A - 5						
I2 - a - I6 - s6				I7 - \$ + s3 (reducing state)		
I2 - b - I7 - s7						
I3 - A - 8 ←				I8 - a → q, q → d2 for a & b ⇒ s12		
I3 - a - I3 - s3						
I3 - b - I4 - s4				I9 ← look ahead if { } }		
I4 → A → b, f - 3 of PR look head → a/b are place a & b column at r3						

LALR(1)

⇒ LALR(1) parser is lookahead LR parser,
It is the most powerful parser.
which can handle large classes of grammar.

⇒ LALR(1) uses canonical items of LR(1)

LALR(1) parsing is same as the CLR(1)
parsing, only difference is the parsing
table.

In LALR(1), combines the similar
states of CLR parsing table into one
single state.

construction of LALR(1) Parser

- ① Construct all canonical LR(1) states.
- ② Merge those states that are identical.
if the lookaheads are ignored.
i.e 2 states being merged must have
the same number of items & items have
the same cone.
- ③ The goto function for the new LALR(1)
state is the union of merged states.
- ④ Action & goto entries are constructed
from the LALR(1) states as for the
canonical LR(1) parser.

Q1) construct the LALR(1) Parse to
for the given grammar.

$$S \rightarrow AA$$

$$A \rightarrow aA/b$$

Solution: $S \rightarrow AA$

$$A \rightarrow aA$$

$$A \rightarrow b$$

→ Follow the
Same answer given for CLR(1).
Write the CLR(1) Parse Table.

→ Check CLR(1) parse table, having
Same production & different
look ahead.

→ Then Write the LALR(1)
Parse table.

Check if there are any two(2) states which have the same production but different look-a-head, from CLR(0)'s parse table.

Identify that

$$\Rightarrow I_3, I_6 \Rightarrow I^{36}$$

$$\Rightarrow I_4, I_7 \Rightarrow I^{47}$$

$$\Rightarrow I_8, I_9 \Rightarrow I^{89}$$

Consider $I_3 \times I_6$.

Then column 3 becomes 36

& remove 6th column.

Check Shifits related to 36

add these also

$$Ex: S_3 \times S_6$$

then becomes $S_{36} \cancel{S_6}$

$I_4 \times I_7$ 4th column become 47

\Rightarrow do similarly for shifits

$I_8 \times I_9 \Rightarrow$ 8th column become 89.

& do similarly for shifits

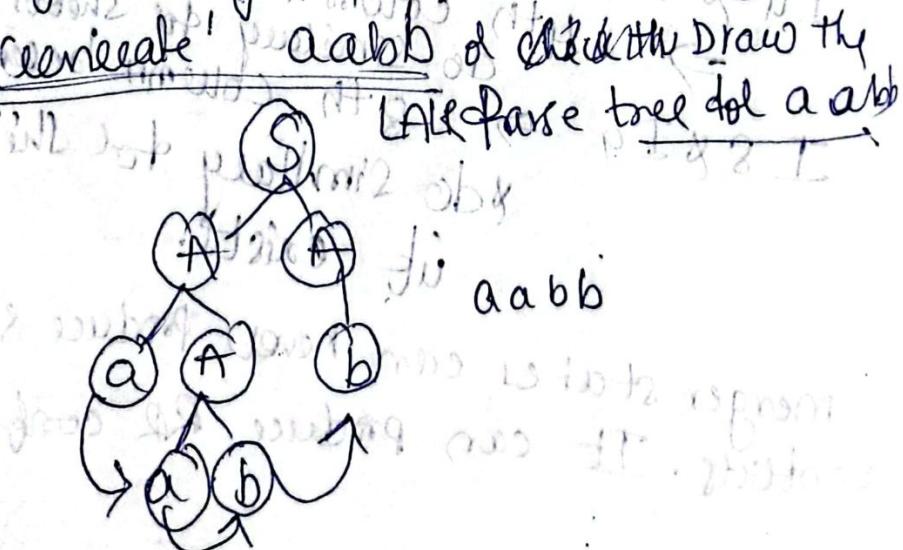
if exists.

merger states can never produce shift-reduce conflicts. It can produce RR conflict.

LALR(1) Parsing Table

States	Action Part			Goto Part	
	a	b	\$	S	A
0	S36	S47	.	1	2
1			Accept		
2	S36	S47	.	5	
36	S36	S47	.	89	
47	R3	R3	R3		
5			R1		
89	R2	R2	R2		

- * The no. of rows has been reduced from 10 to 7 rows.
- * Row 36, row 47 have the same data twice, so delete 1 row.
- * We combine two 47 rows into one by combining each value in the single 47 row.



UNIT-4(Chapter 2 & 3)

Semantic Analysis: Introduction to Syntax Directed Definition, Syntax Directed Translation, Attributes, Types of Attributes, Bottom-up evaluation of attributes.

Intermediate code generation: Types of Intermediate codes, Types of Three address codes

Introduction to Syntax-Directed Definitions and Translation Schemes

1. When we associate semantic rules with productions, we use two notations:
 1. **Syntax directed definition (SDD)**
 2. **Translation Scheme (SDT)**
-
- A. Syntax-Directed Definitions:**
- give high-level specifications for translations
 - hide many implementation details such as order of evaluation of semantic actions.
 - We associate a production rule with a set of semantic actions, and we do not say when they will be evaluated.
-
- B. Translation Schemes:**
- indicate the order of evaluation of semantic actions associated with a production rule.
 - In other words, translation schemes give a little bit information about implementation details.

Syntax-Directed Definition

A **syntax-directed definition (SDD)** is a context-free grammar with attributes and rules.

Syntax Directed Definition = CFG + Semantic rules

- Attributes are associated with grammar symbols and rules are associated with productions.
- These rules are set by the grammar of the language and evaluated in semantic analysis.

Example:

- If X is a symbol and a is one of its attributes, we write $X.a$ to denote the value of a at a particular parse-tree node labeled X . Attributes may be numbers, types, table references or strings

$$E \rightarrow E + T$$

$$E.value = E.value + T.value$$

- Example: $E \rightarrow E + T$
- The above CFG production has **no semantic rule** associated with it and it cannot help in making any sense of the production
- A parser **builds parse trees in the syntax analysis phase.**
- The **plain parse-tree is of no use for a compiler**, as it does **not carry any information** of how to evaluate the tree.
- The productions of context-free grammar that makes the rule so the languages do not accommodate how to interpret them

Syntax-Directed Definition -- Example

Production

$L \rightarrow E \ n$

$E \rightarrow E_1 + T$

$E \rightarrow T$

$T \rightarrow T_1 * F$

$T \rightarrow F$

$F \rightarrow (E)$

$F \rightarrow \text{digit}$

Semantic Rules

$\text{print}(E.\text{val})$

$E.\text{val} = E_1.\text{val} + T.\text{val}$

$E.\text{val} = T.\text{val}$

$T.\text{val} = T_1.\text{val} * F.\text{val}$

$T.\text{val} = F.\text{val}$

$F.\text{val} = E.\text{val}$

$F.\text{val} = \text{digit}.\text{lexval}$

Semantic Errors

We have mentioned some of the semantics errors that the semantic analyser is expected to recognize:

- Type mismatch
- Undeclared variable
- Reserved identifier misuse.
- Multiple declaration of variable in a scope.
- Accessing an out of scope variable.
- Actual and formal parameter mismatch.

Attributes

- Each Grammar symbol is associated with set of attributes
 - Each attribute has well-defined domain of values, such as integer, float, character, string, and expressions
 - Attributes can be
 - String
 - Number
 - Type
 - Memory Location
- The value of the attribute at a parse tree node is defined by a semantic rule associated with the production used at the node

Types of Attributes

- Semantic attributes may be assigned to their values from their domain at the time of parsing and evaluated at the time of assignment or conditions.
- Based on the way the attributes get their values.
- They divided into two(2) categories:
 1. Synthesized attributes
 2. Inherited attributes

Synthesized attributes

These attributes get values from the attribute values of their child nodes. To illustrate, assume the following production:

$$S \rightarrow ABC$$

If S is taking values from its child nodes (A,B,C), then it is said to be a synthesized attribute, as the values of ABC are synthesized to S.

- 1. It can be evaluated during a single bottom-up traversal of parse tree.**
- 2. Synthesized attributes can be contained by both the terminals or non-terminals.**

Synthesized attributes

CFG

$E \rightarrow E + T$

$E \rightarrow T$

$T \rightarrow T * F$

$T \rightarrow F$

$F \rightarrow \text{num}$

Semantic Actions

$\{ E.\text{value} = E.\text{value} + T.\text{value} \}$

$\{ E.\text{value} = T.\text{value} \}$

$\{ T.\text{value} = T.\text{value} * F.\text{value} \}$

$\{ T.\text{value} = F.\text{value} \}$

$\{ F.\text{value} = \text{num}. \text{lexicalvalue} \}$

Synthesized attribute is used by both S-attributed SDT and L-attributed SDT.

Inherited attributes

In contrast to synthesized attributes, **inherited attributes can take values from parent and/or siblings**. As in the following production,

$$S \rightarrow ABC$$

A can get values from S, B and C. B can take values from S, A, and C. Likewise, C can take values from S, A, and B.

- 1. It can be evaluated during a single top-down and sideways traversal of parse tree.**
- 2. Inherited attributes can't be contained by both, It is only contained by non-terminals.**
- 3. Inherited attribute is used by only L-attributed SDT.**

Inherited attributes

Unlike synthesized attributes, inherited attributes can take values from parent and/or siblings. As in the following production,

Annotate the parse tree for the computation of inherited attributes for the given string int a, b, c.

Solution

Step1. The Syntax Directed Definition for the given grammar.

Productions	Semantic Rules
$S \rightarrow T \ L$	$L.in \ .T.type$
$T \rightarrow int$	$T.type = int$
$T \rightarrow float$	$T.type = float$
$T \rightarrow char$	$T.type = char$
$T \rightarrow double$	$T.type = double$
$L \rightarrow L1, id$	$L1.in = L.in$
$L \rightarrow id$	$Id.entry = L.inh$

Example Attribute Grammar with Synthesized & Inherited Attributes

Simple Type Declaration

Production	Semantic Rule	
$D \rightarrow TL$	$L.inh = T.type$	treated as dummy synthesized attribute
$T \rightarrow \text{int}$	$T.type = \text{'integer'}$	
$T \rightarrow \text{float}$	$T.type = \text{'float'}$	
$L \rightarrow L_1, id$	$L_1.inh = L.inh;$ $\text{addtype}(id.entry, L.inh)$	
$L \rightarrow id$	$\text{addtype}(id.entry, L.inh)$	

Synthesized: $T.type$, $id.entry$

Inherited: $L.inh$

Parse Tree

Production Rules

$D \rightarrow T \ L$

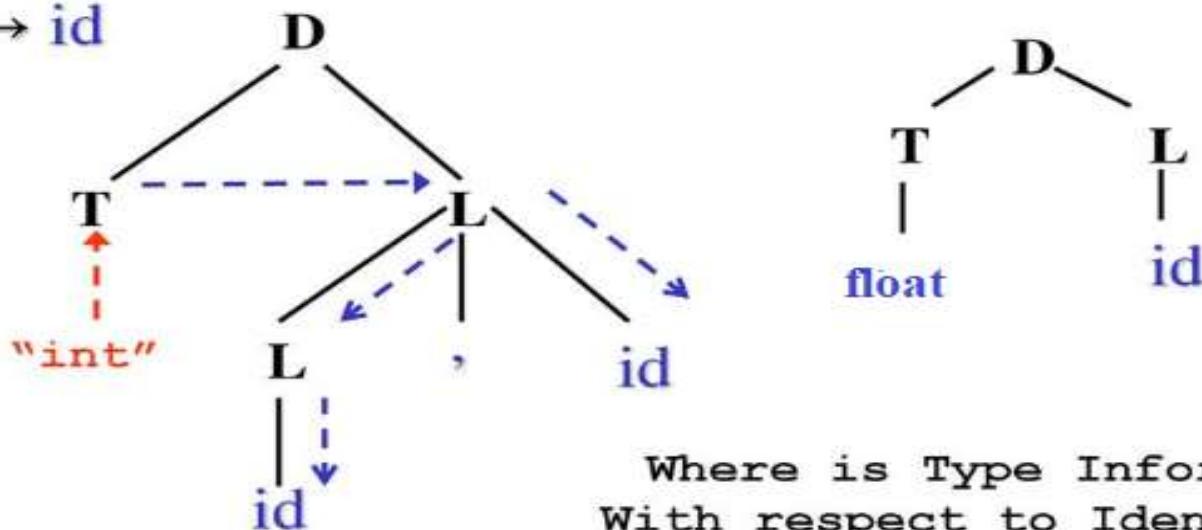
$T \rightarrow \text{int}$

$T \rightarrow \text{float}$

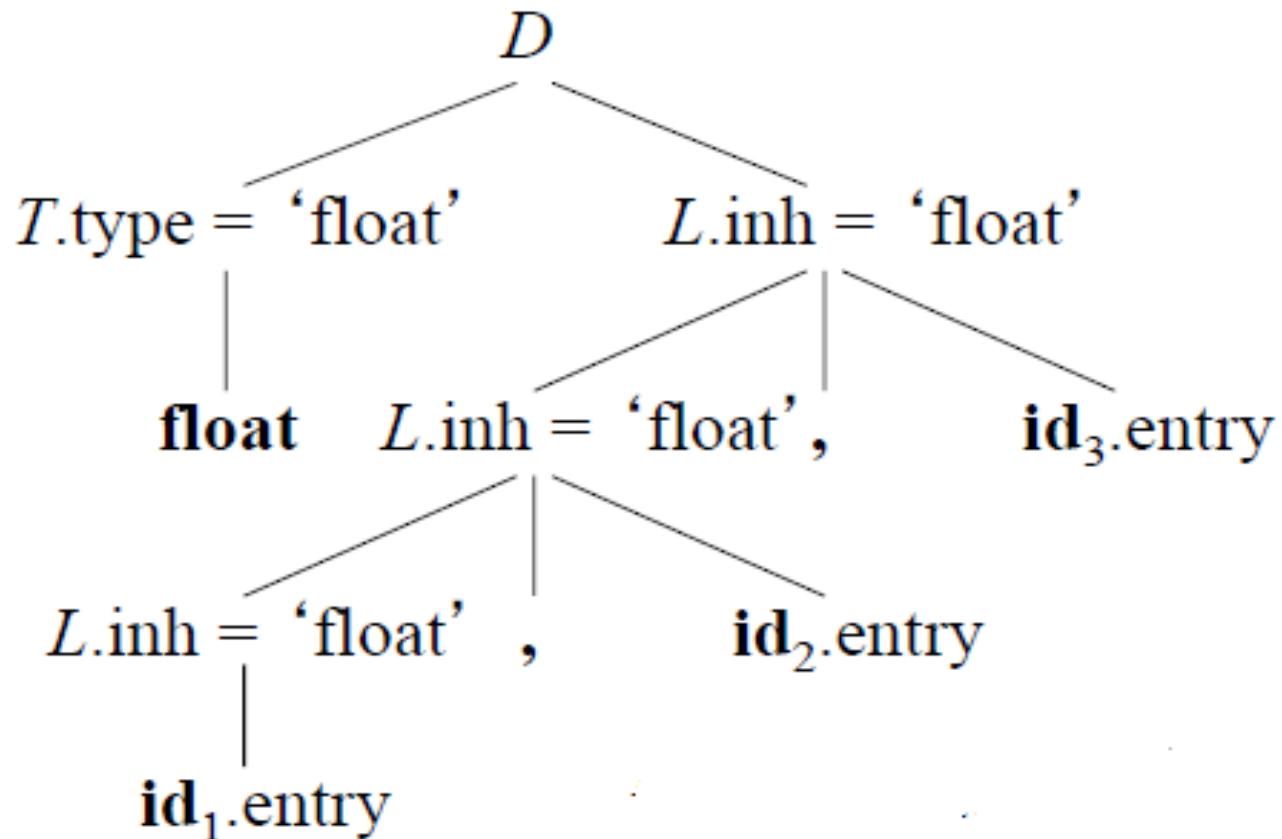
$L \rightarrow L , id$

$L \rightarrow id$

$D \Rightarrow TL \Rightarrow \text{int } L$
 $\Rightarrow \text{int } L , id$
 $\Rightarrow \text{int } L , id , id$
 $\Rightarrow \text{int id , id , id}$

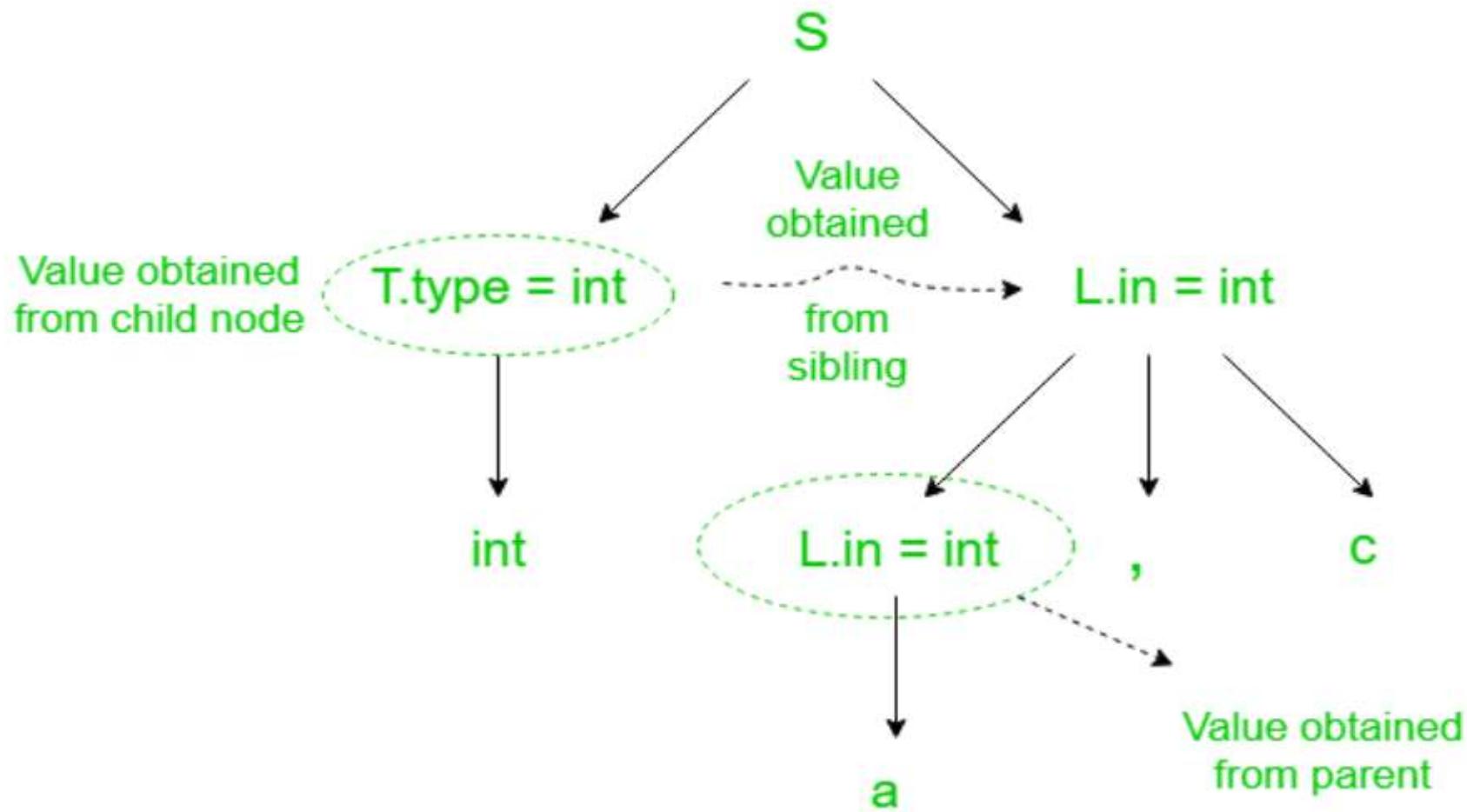


Construct the annotated parse tree for
float id1,id2,id3 using previous production



Annotated Parse Tree

1. A parse tree showing the values of attributes at each node is called an **annotated parse tree**.
2. Values of Attributes in nodes of annotated parse-tree are either,
 - initialized to constant values or by the lexical analyzer.
 - determined by the semantic-rules.
3. The process of computing the attributes values at the nodes is called **annotating** (or **decorating**) of the parse tree.
4. Of course, the order of these computations depends on the dependency graph induced by the semantic rules.



Annotated Parse Tree

Annotated Parse Tree -- Example

Production

$L \rightarrow E \ n$

$E \rightarrow E_1 + T$

$E \rightarrow T$

$T \rightarrow T_1 * F$

$T \rightarrow F$

$F \rightarrow (E)$

$F \rightarrow \text{digit}$

Semantic Rules

$\text{print}(E.\text{val})$

$E.\text{val} = E_1.\text{val} + T.\text{val}$

$E.\text{val} = T.\text{val}$

$T.\text{val} = T_1.\text{val} * F.\text{val}$

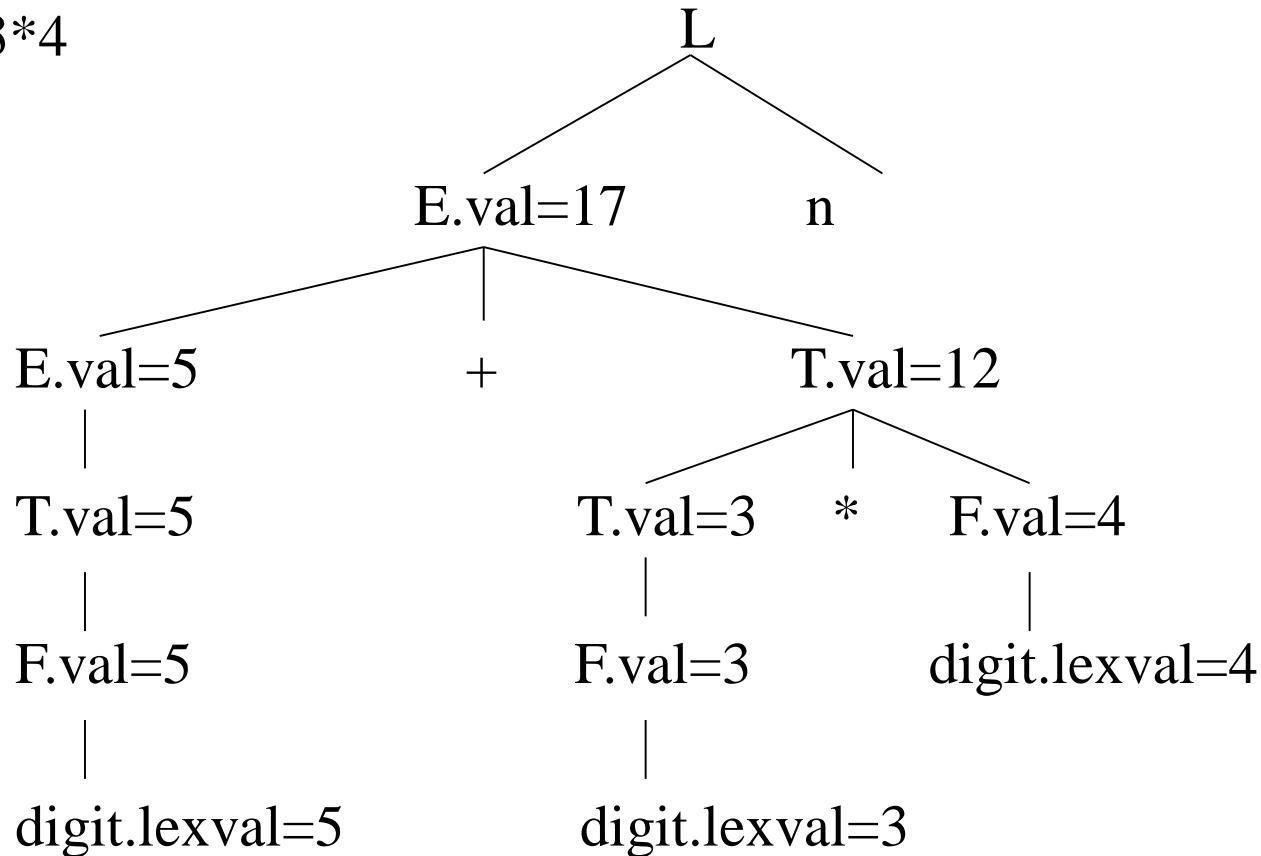
$T.\text{val} = F.\text{val}$

$F.\text{val} = E.\text{val}$

$F.\text{val} = \text{digit}.lexval$

Annotated Parse Tree -- Example

Input: $5+3*4$



Evaluation order/ Implementation of SDD's

- Evaluation order for SDD includes how the SDD(Syntax Directed Definition) is evaluated with the help of attributes, dependency graphs, semantic rules, and S and L attributed definitions

Evaluation order/ Implementation of SDD's :

- Dependency graphs are a useful tool for determining an evaluation order for the attribute instances in a given parse tree.
- While an annotated parse tree shows the values of attributes, a dependency graph helps us determine how those values can be computed.
- Dependency graphs define two important classes of SDD's:
 - » S-attributed SDD's
 - » L-attributed SDD's

Dependency Graph

- The basic idea behind dependency graphs is for compiler to look for various kinds of dependence among statements to prevent their execution in wrong order i.e. the order that changes the meaning of the program.
- This helps it to identify various parallel executions of the components in the program.
- For rule $X \rightarrow YZ$
- the semantic action is given by $X.x = f(Y.y, Z.z)$
- then synthesized attribute is $X.x$ and $X.x$ depends upon attributes $Y.y$ and $Z.z$

PRODUCTION

$$E \rightarrow E_1 + T$$

SEMANTIC RULE

$$E.val = E_1.val + T.val$$

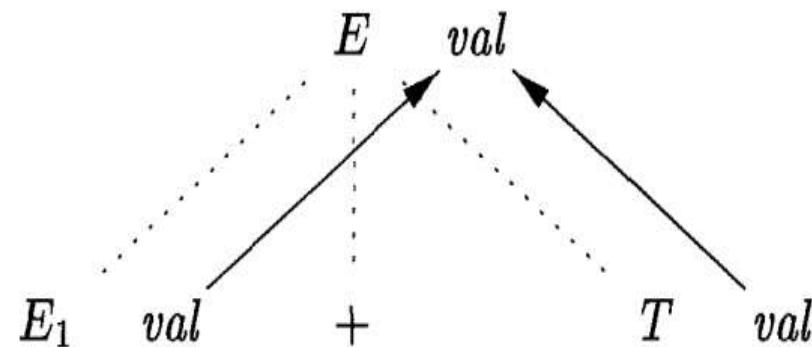


Figure 5.6: $E.val$ is synthesized from $E_1.val$ and $E_2.val$

Syntax Directed Translations

- Syntax Directed Translation has augmented rules to the grammar that facilitate semantic analysis.
- In stream is the result obtained by evaluating the semantic rules.
- Semantic analyzer receives AST (Abstract Syntax Tree) from its previous stage (syntax analysis).
- Semantic analyzer attaches attribute information with AST, which are called Attributed AST.

Attributes are two tuple value, <attribute name, attribute value>

For example:

int value = 5;

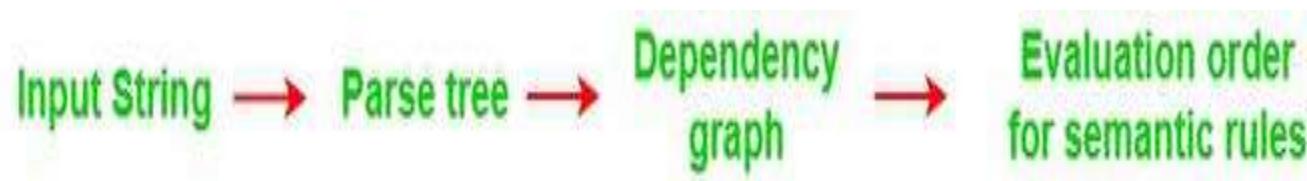
<type, “integer”>

<present value, “5”>

For every production, we attach a semantic rule.

Syntax Directed Translation

- syntax-directed definition and translation schemes, we parse the input token stream, build the parse tree, and then traverse the tree as needed to evaluate the semantic rules at the parse tree nodes.



The conceptual view of syntax directed translation

In Syntax Directed Translations the Productions with semantic actions embedded with production bodies.(Actions are specified)

Eg: $E \rightarrow E_1 + T \quad \{ \text{print} '+' \}$

- The general approach to Syntax-Directed Translation is **to construct a parse tree or syntax tree** and compute the values of attributes at the nodes of the tree by visiting them in some order.
- SDT involves passing information bottom-up and/or top-down to the parse tree in form of attributes attached to the nodes.
- Syntax-directed translation rules use
 1. lexical values of nodes,
 2. constants &
 3. attributes associated with the non-terminals in their definitions.

Construct Syntax Directed Translation for the given grammar (for 2^*3+4)

$E \rightarrow E + T$

$E \rightarrow T$

$T \rightarrow T * F$

$T \rightarrow F$

$F \rightarrow \text{num}$

Solution:

1. Context free grammar + semantic actions are said to be syntax Definition Translation
2. Abstract Syntax Tree for $2 * 3 + 4$
3. Constructing Parse Tree for the above grammar
4. Annotated Parse Tree

Step 1: 1. Syntax-Directed Translation (SDT)

This is a context-free grammar + semantic actions. SDT is a way to attach meaning (semantics) to the syntactic structure of a program.

Example with semantic actions:

Semantic actions are to be written in between curly braces.

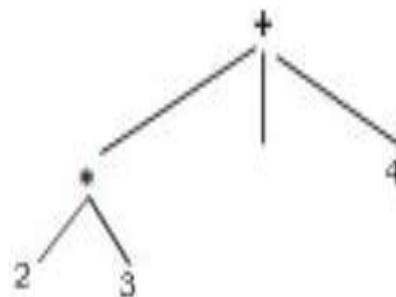
Context Free Grammar	Semantic Actions
$E \rightarrow E + T$	{ E.value = E.value + T.value }
$E \rightarrow T$	{E.value = T.value}
$T \rightarrow T * F$	{T.value = T.value * F.value}
$T \rightarrow F$	{T.value = F.value}
$F \rightarrow \text{num}$	{F.value = num.lexical value}

Step 2 -Abstract Syntax Tree (AST) for $2 * 3 + 4$

An AST omits the intermediate grammar rules and focuses only on operations and operands. It shows the essential computation.

Expression: $2 * 3 + 4$

Operator Precedence: * has higher precedence than +

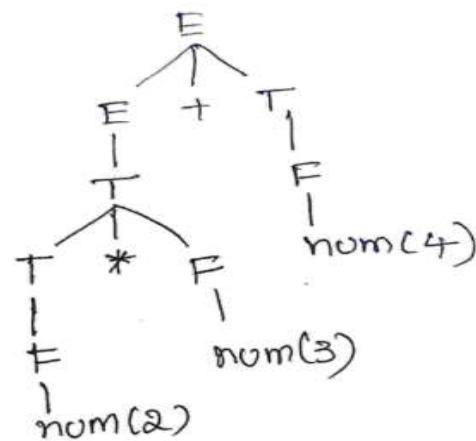


Abstract Syntax Tree

Step 3: Constructing Parse Tree for the above grammar.

This uses the actual grammar rules (productions) to derive the expression $2 * 3 + 4$.

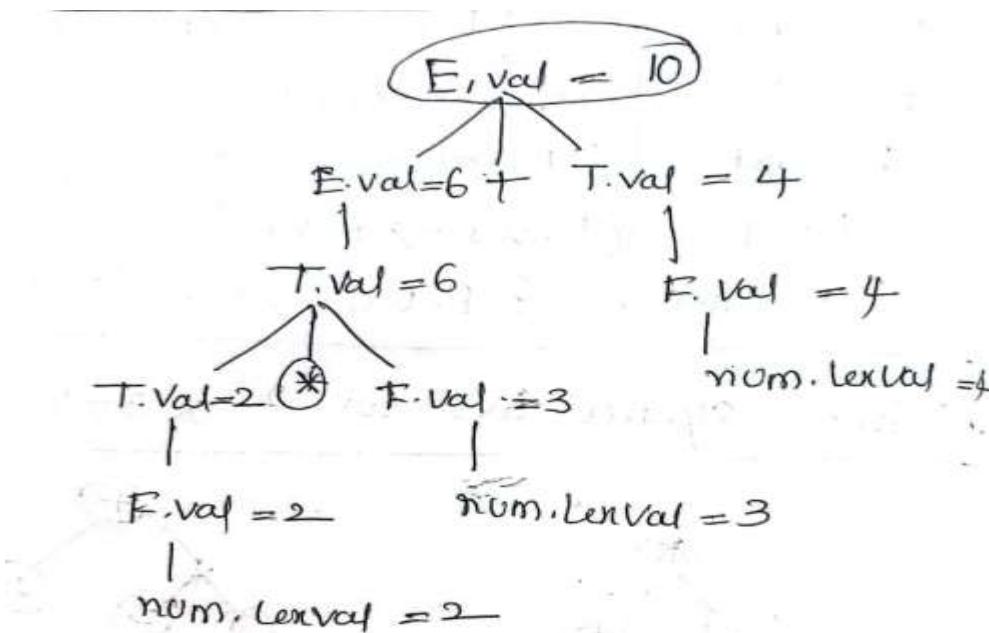
Parse Tree for $2 * 3 + 4$:



Step 4: Annotated Parse Tree / Decorated Parse Tree

An annotated parse tree is a parse tree where each node is decorated with the result of semantic actions (like computed values).

Let's compute values bottom-up for $2 * 3 + 4$.



- The use of SDT's to implement two important classes of SDD's:
 1. The underlying grammar is **LR-parsable**, and the SDD is **S-attributed**.
 2. The underlying grammar is **LL-parsable**, and the SDD is **L-attributed**.

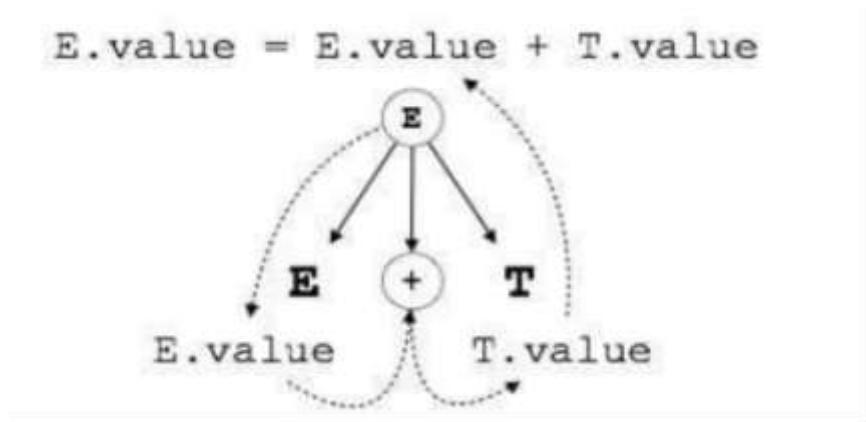
Two SDD's considered for constructing syntax trees for expressions.

The first, an **S-attributed** definition, is suitable for use **during bottom-up parsing**.

second, **L-attributed**, is suitable for use during **top-down parsing**.

S-attributed SDT

- If an SDT uses only synthesized attributes, it is called as S-attributed SDT.
- S-attributed SDTs are evaluated in bottom-up parsing, as the values of the parent nodes depend upon the values of the child nodes.
- Semantic actions are placed in rightmost place of RHS.

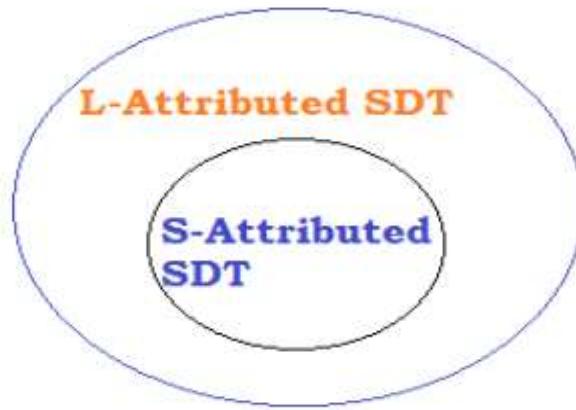


L-Attributed SDT

- If an SDT uses both synthesized attributes and inherited attributes with a restriction that inherited attribute can inherit values from left siblings only, it is called as L-attributed SDT.
- Attributes in L-attributed SDTs are evaluated by **depth-first and left-to-right parsing manner**.
- Semantic actions are placed anywhere in RHS.

Example : S->ABC,

- Here attribute B can only obtain its value either from the parent – S or its left sibling A but It can't inherit from its right sibling C. Same goes for A & C –
- A can only get its value from its parent & C can get its value from S, A, & B as well because C is the rightmost attribute in the given production.
- Attributes in L-attributed SDTs are evaluated by **depth-first and left-to-right parsing manner**



- If a definition is S-attributed, then it is also L-attributed
- L-attributed definition encloses S-attributed definitions.

S-Attributed SDT Vs L-Attributed ADT

- Based on Synthesized attributes(i.e get values from the attribute values of their child nodes)
- Use Bottom up parsing.
- Sematic rule always written at rightmost position in RHS.
- Based on Synthesized and Inherited attributes (I.A- can take values from parent and/or siblings)
- Use Top-Down parsing.
- Sematic rules anywhere in RHS.

Applications of Syntax-Directed Translation

- The main application in this section is the construction of **syntax tree**.
- A syntax tree is condensed form of a Parse Tree.
- Useful for representing the expressions and statements
- The child node represents the meaningful component of the construct.

Consider the grammar that is used for Simple desk calculator. Obtain the Semantic action and also the annotated parse tree for the string $3*5+4n$.

$L \rightarrow E_n$

$E \rightarrow E_1 + T$

$E \rightarrow T$

$T \rightarrow T_1 * F$

$T \rightarrow F$

$F \rightarrow (E)$

$F \rightarrow \text{digit}$

Solution : Syntax-directed Definition for A simple desk calculator

Production	Semantic Rule
$L \rightarrow E \text{ n}$	$L.\text{val} = E.\text{val}$
$E \rightarrow E_1 + T$	$E.\text{val} = E_1.\text{val} + T.\text{val}$
$E \rightarrow T$	$E.\text{val} = T.\text{val}$
$T \rightarrow T_1 * F$	$T.\text{val} = T_1.\text{val} * F.\text{val}$
$T \rightarrow F$	$T.\text{val} = F.\text{val}$
$F \rightarrow (E)$	$F.\text{val} = E.\text{val}$
$F \rightarrow \text{digit}$	$F.\text{val} = \text{digit}.\text{lexval}$

An SDD with only synthesized attributes is called S-attributed.

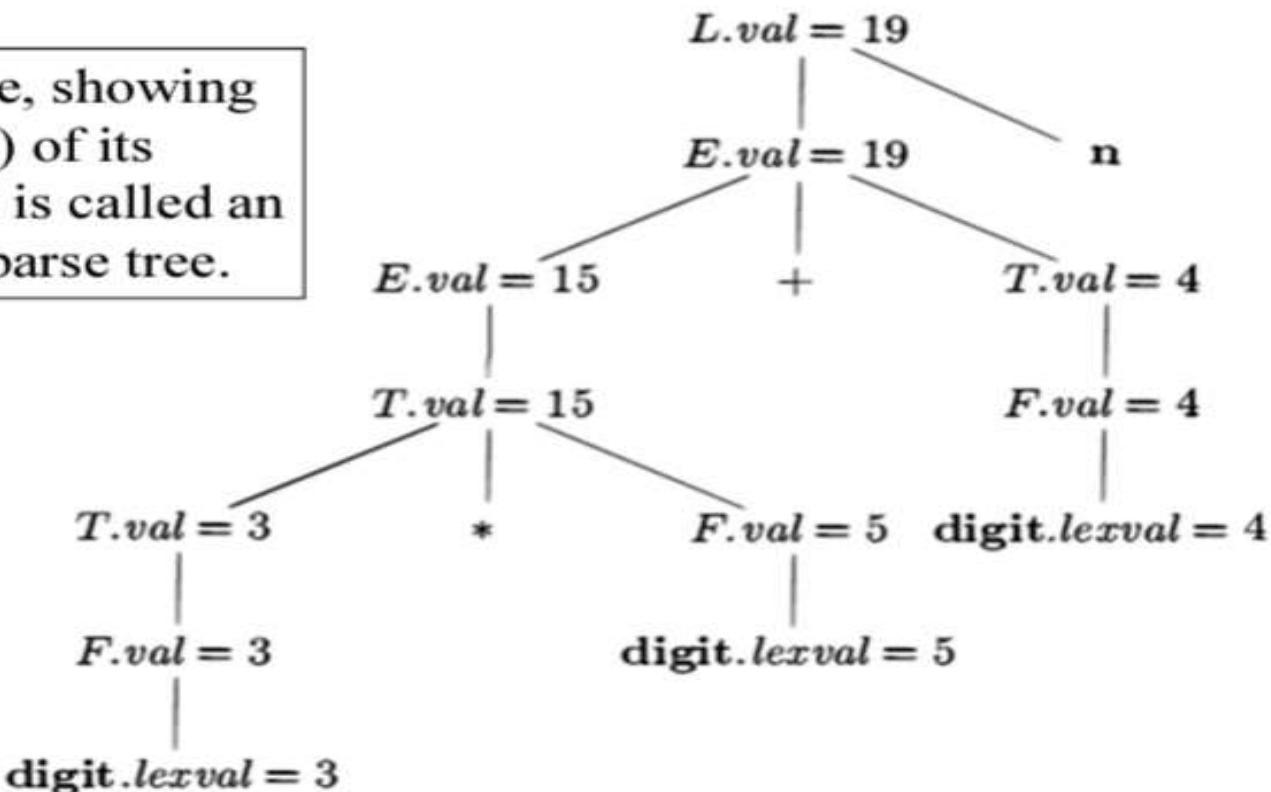
An SDD without side effects is called an attribute grammar

Note: all attributes in this example are of the synthesized type

Solution : Syntax-directed Definition for A simple desk calculator

Annotated Parse Tree for $3 * 5 + 4 \text{ n}$

A parse tree, showing the value(s) of its attribute(s) is called an annotated parse tree.



Bottom up Evaluation of Attributes

Explain in detail about Bottom-Up Evaluation of S-Attributed Definitions.

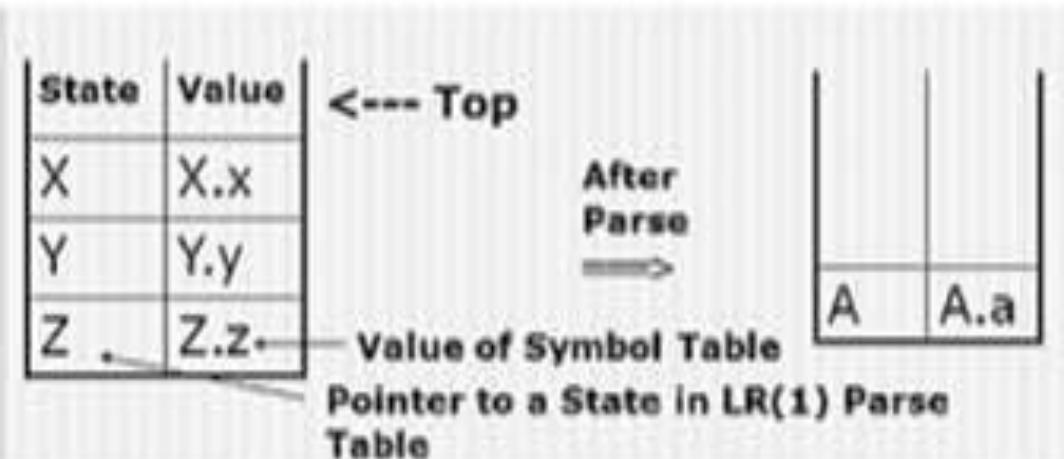
- Syntax-directed definitions with only synthesized attributes(S-attribute) can be evaluated by a **bottom up parser** (BUP) as input is parsed
- In this approach, the parser will keep the values of synthesized attributes associated with the grammar symbol on its stack.
- The stack is implemented as a **pair of state and value**.
- When a reduction is made ,the values of the synthesized attributes are computed from the attribute appearing on the stack for the grammar symbols.
- Implementation is by using an LR parser (e.g. YACC)

Bottom up Evaluation of Attributes

- e.g. $A \Rightarrow XYZ$ and $A.a := f(X.x, Y.y, Z.z)$

TOS

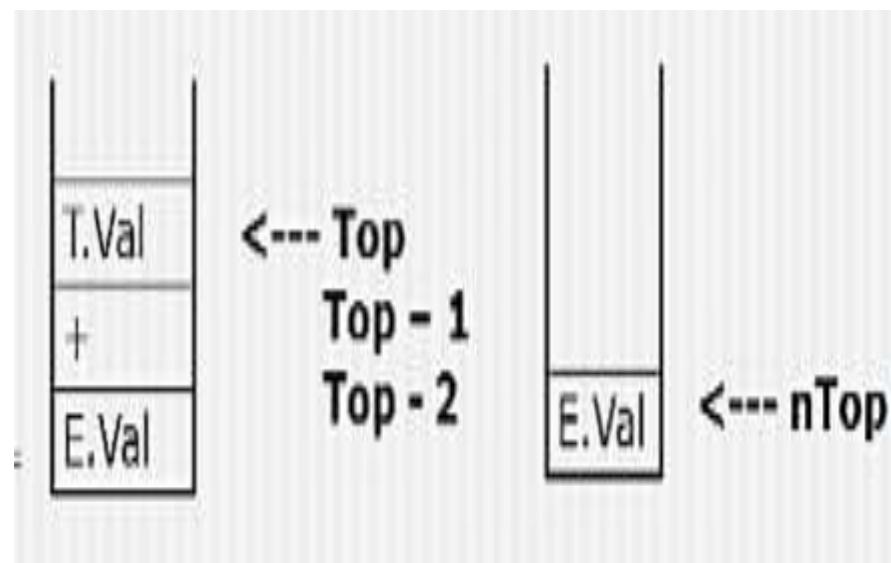
state	value
X	X.x
Y	Y.y
Z	Z.z



Bottom up Evaluation of Attributes

Production	Semantic Rules
$L \rightarrow E \text{ n}$	$print(E.val)$
$E \rightarrow E_1 + T$	$E.val := E_1.val + T.val$
$E \rightarrow T$	$E.val := T.val$
$T \rightarrow T_1 * F$	$T.val := T_1.val + F.val$
$T \rightarrow F$	$T.val := F.val$
$F \rightarrow (E)$	$F.val := E.val$
$F \rightarrow \text{digit}$	$F.val := \text{digit.lexval}$

- Implementation of Desk Calculator.



Bottom up Evaluation of Attributes

- **Synthesized Attributes on the Parser Stack**

- If attribute values are placed on the stack as described, it is implement the semantic rules for the desk calculator.
- Example – Desk Calculator

Production	Semantic Rule	Code
$L \rightarrow E \text{ newline}$	$\text{print}(E.\text{val})$	$\text{print val}[\text{top}-1]$
$E \rightarrow E_1 + T$	$E.\text{val} := E_1.\text{val} + T.\text{val}$	$\text{val}[\text{newtop}] = \text{val}[\text{top-2}] + \text{val}[\text{top}]$
$E \rightarrow T$	$E.\text{val} := T.\text{val}$	$/*\text{newtop}=\text{top}, \text{so nothing to do}*/$
$T \rightarrow T_1 * F$	$T.\text{val} := T_1.\text{val} \times F.\text{val}$	$\text{val}[\text{newtop}] = \text{val}[\text{top-2}] * \text{val}[\text{top}]$
$T \rightarrow F$	$T.\text{val} := F.\text{val}$	$/*\text{newtop}=\text{top}, \text{so nothing to do}*/$
$F \rightarrow (E)$	$F.\text{val} := E.\text{val}$	$\text{val}[\text{newtop}] = \text{val}[\text{top-1}]$
$F \rightarrow \text{digit}$	$F.\text{val} := \text{number.lexval}$	$/*\text{newtop}=\text{top}, \text{so nothing to do}*/$

Production	Value Stack Implementation
$L \rightarrow E\ n$	print(E.val)
$E \rightarrow E + T$	$\text{val}(n\text{top}) = \text{val}(\text{top} - 2) + \text{val}(\text{top})$
$E \rightarrow T$	no action
$T \rightarrow T * F$	$\text{val}(n\text{top}) = \text{val}(\text{top} - 2) * \text{val}(\text{top})$
$T \rightarrow F$	no action
$F \rightarrow (E)$	$\text{val}(n\text{top}) = \text{val}(\text{top} - 1)$
$F \rightarrow \text{digit}$	no action

Bottom up Evaluation of Attributes

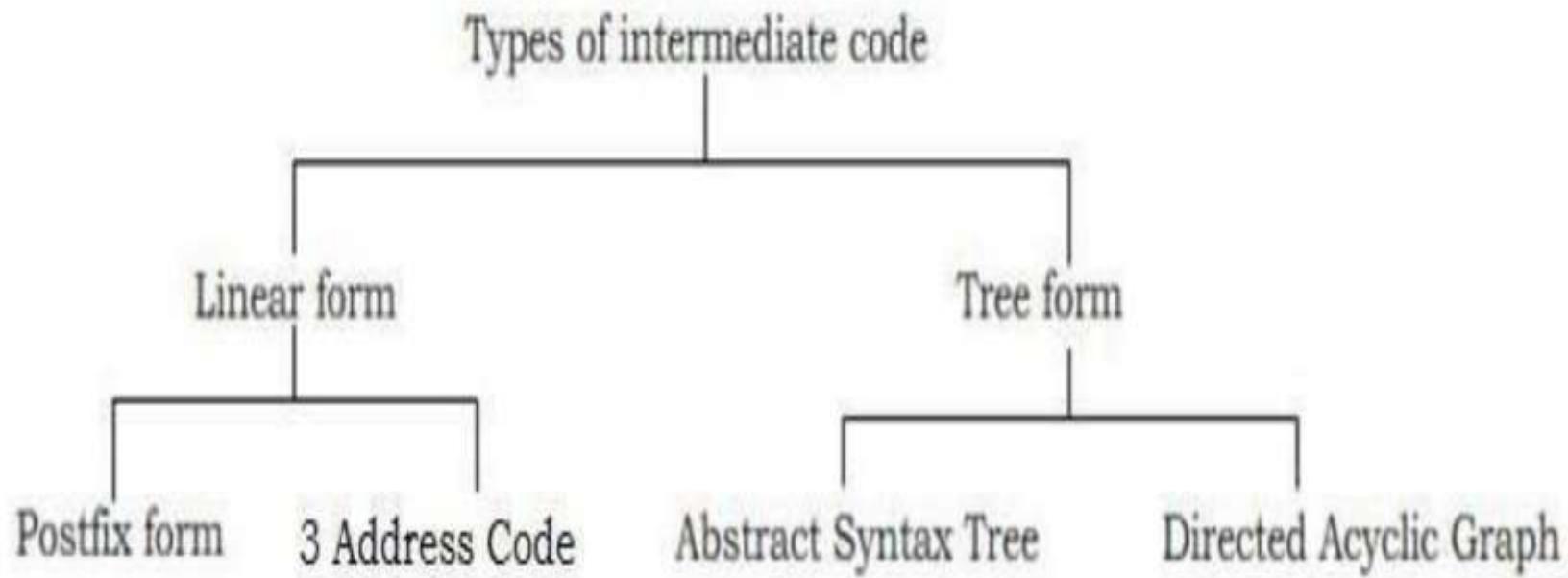
INPUT	STATE	VAL	PRODUCTION USED
3 * 5 + 4 n	-	-	
* 5 + 4 n	3	3	
* 5 + 4 n	F	3	$F \rightarrow \text{digit}$
* 5 + 4 n	T	3	$T \rightarrow F$
5 + 4 n	T *	3 -	
+ 4 n	T * 5	3 - 5	
+ 4 n	T * F	3 - 5	$F \rightarrow \text{digit}$
+ 4 n	T	15	$T \rightarrow T * F$
+ 4 n	E	15	$E \rightarrow T$
4 n	E *	15 -	
n	E * 4	15 - 4	
n	E * F	15 - 4	$F \rightarrow \text{digit}$
n	E * T	15 - 4	$T \rightarrow F$
n	E	19	$E \rightarrow E * T$
	E n	19 -	
	L	19	$L \rightarrow E n$

Intermediate code generation:
Types of Intermediate codes,
Types of Three address codes.

Intermediate code generation

- In the analysis-synthesis model of a compiler,
 - the front end of a compiler translates a source program into an independent intermediate code,
 - then the back end of the compiler uses this intermediate code to generate the target code.
- A three-address statement is an abstract form of intermediate code.
- Intermediate code is used to translate the source code into the machine code. Intermediate code lies between the high-level language and the machine language.
- Intermediate code can be either language specific (e.g., Byte Code for Java) or language independent (three-address code).

Types of Intermediate code



Intermediate Code Representations

The following are commonly used intermediate code representations:

- 1) Postfix Notation
- 2) Syntax Tree
- 3) Three-Address Code

Postfix Notation: Also known as reverse Polish notation or suffix notation.

- In this notation, the operator is placed after operands,
e.g., $a + b$. Postfix notation positions the operator at the right end,
as in $ab +$.

Example 1: The postfix representation of the expression

$(a + b) * c$ is :

$ab + c *$

$(a - b) * (c + d) + (a - b)$ is :

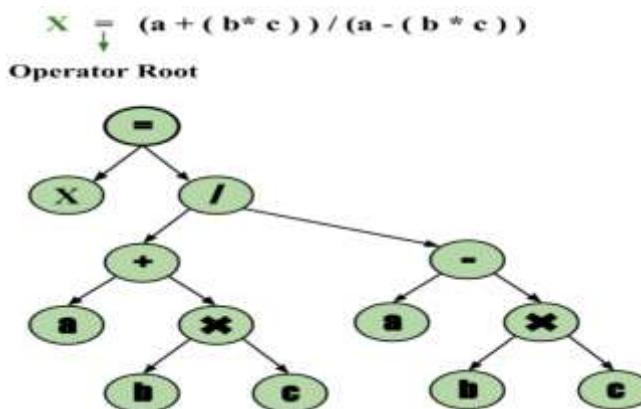
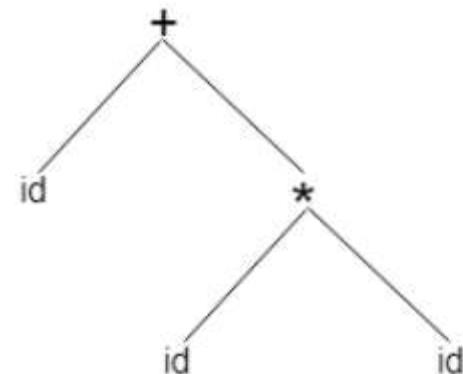
$ab - cd + *ab - +$

2. Syntax Tree: A syntax tree serves as a condensed representation of a parse tree.

syntax tree is usually used when represent a program in a tree structure.

- The operators are represented as parent and intermediate nodes, identifiers and numbers are represented as leaf nodes.
- Creating a syntax tree involves strategically placing parentheses within the expression.
- The syntax tree not only condenses the parse tree but also offers an improved visual representation of the program's syntactic structure
- Example:** $x = (a + b * c) / (a - b * c)$

sentence $id + id * id$
would have the following
syntax tree:



3. Three-Address Code: A three address statement involves a maximum of three references, consisting of two for operands and one for the result.

- A sequence of three address statements collectively forms a three address code.
- The typical form of a three address statement is expressed as

$x = y \ op \ z,$ where $x, y,$ and z represent memory addresses.

- Each variable (x, y, z) in a three address statement is associated with a specific memory location.

Example: The three address code for the expression $a = b + c * d:$

First, convert this statement into Three Address code

∴ Three Address code will be

$t1 = c * d$

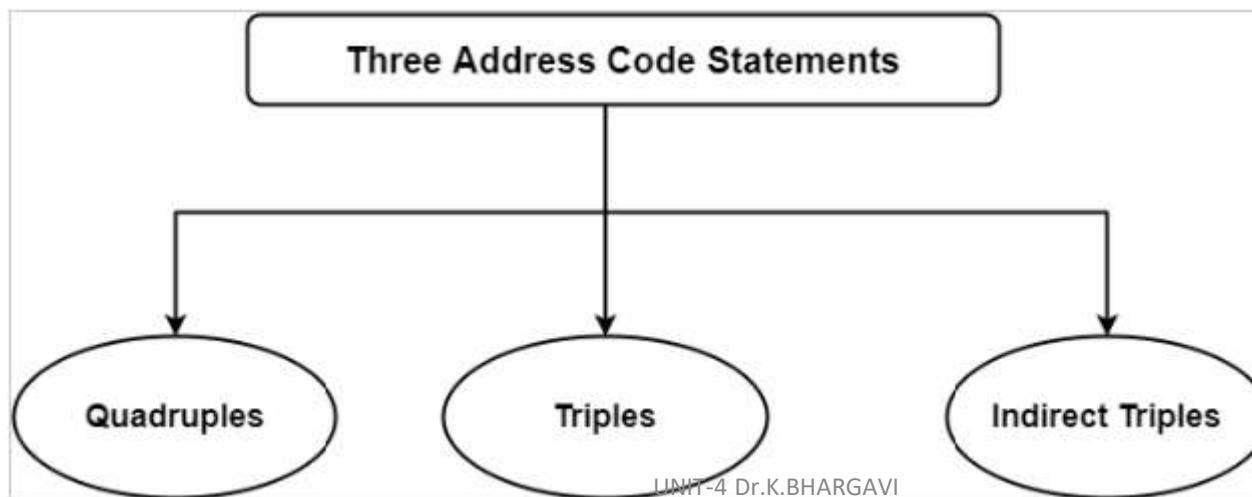
$t2 = b + t1$

$a = t2$

. $t1, t2$ are temporary variables.

Implementation of Three Address Code

- There are 3 representations of three address code.
 1. Quadruple
 2. Triples
 3. Indirect Triples



Quadruple

- It is structure with consist of 4 fields namely op, arg1, arg2 and result. op denotes the operator and arg1 and arg2 denotes the two operands and result is used to store the result of the expression.
- **Advantages**
 - Easy to rearrange code for global optimization.
 - One can quickly access value of temporary variables using symbol table.
- **Disadvantage**
 - Contain lot of temporaries.
 - Temporary variable creation increases time and space complexity.

Example $a = b + c * d$



Example $a = b + c * d$

Quadruple

Location	Operator	arg 1	arg 2	Result
(0)	*	c	d	t1
(1)	+	b	t1	t2
(2)	=	t2		a

The content of fields arg 1, arg 2 and Result are pointers to symbol table entries for names represented by these entries.

Triples

- This representation doesn't make use of extra temporary variable to represent a single operation instead **when a reference to another triple's value is needed, a pointer to that triple is used.**
- So, it consist of only three fields namely **op, arg1 and arg2.**
- **Disadvantage**
 - Temporaries are implicit and difficult to rearrange code.
 - It is difficult to optimize because optimization involves moving intermediate code. When a triple is moved, any other triple referring to it must be updated also.
 - With help of pointer one can directly access symbol table entry.

Triples

Triple

Location	Operator	arg 1	arg 2
(0)	*	c	d
(1)	+	b	(0)
(2)	=	a	(1)

Here (0) represents a pointer that refers to the result $c * d$, which can be used in further statements, i.e., when $c * d$ is added with b. This result will be saved at the position pointer by (1). Pointer (1) will be used further when it is assigned to a.

Indirect Triples

- The indirect triple representation uses an extra array to list the pointers to the triples in the desired order. This is known as indirect triple representation.
- Its similar in utility as compared to quadruple representation but requires less space than it.
- Temporaries **are implicit and easier to rearrange code.**

Indirect Triples

Indirect Triples

Statement	
(0)	(11)
(1)	(12)
(2)	(13)

Location	Operator	arg 1	arg 2
(11)	*	c	d
(12)	+	b	(11)
(13)	=	a	(12)

In this, it can only need to refer to a pointer (0), (1), (2) which will further refer pointers (11), (12), (13) respectively & then pointers (11), (12), (13) point to triples that is why this representation is called indirect triple representation.

6.2.5 Exercises for Section 6.2

Exercise 6.2.1: Translate the arithmetic expression $a + -(b + c)$ into:

- a) A syntax tree.
- b) Quadruples.
- c) Triples.
- d) Indirect triples.

Exercise 6.2.2: Repeat Exercise 6.2.1 for the following assignment statements:

i. $a = b[i] + c[j].$

ii. $a[i] = b*c - b*d.$

iii. $x = f(y+1) + 2.$

iv. $x = *p + &y.$

Example:

- Construct Quadruples ,Triples, Indirect triples for expression

$a = b * - c + b * - c.$