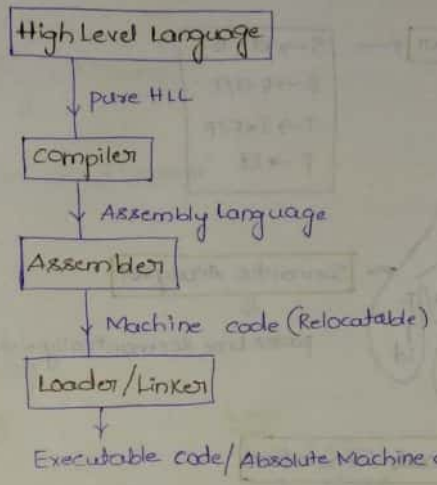
COMPILER DESIGN COMPLETE NOTES

HTTPS://WWW.LINKEDIN.COM/MWLITE/IN/ADARSH-C-942702A5

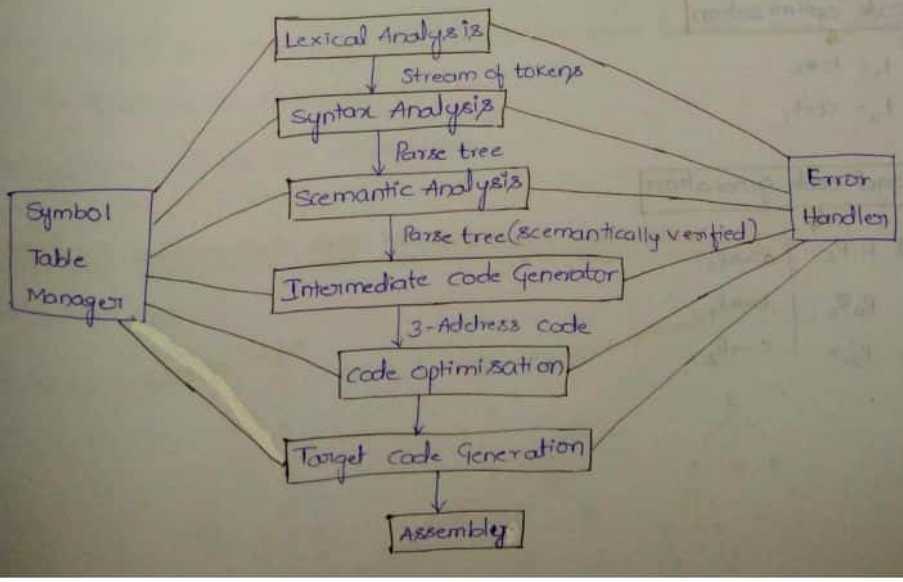
INTRODUCTION AND VARIOUS PHASES OF COMPILER (L-1)

⇒ The main aim of the compiler is to convert a High Level Language to Low level language.



⇒ Removing #<include> by including a specific file is called "File Inclusion"

⇒ "Assembler" is dependent on the platform [Hardware + OS].



Example :

1) $x = a + b * c$

⇓

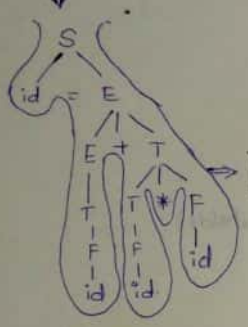
Lexical Analysis ⇒ The Lexical Analyzer identifies the "identifiers" tokens" is using some Regular expressions called patterns $l(l+d)^*$

$id = id + id * id$

⇓

Syntax Analyser ←

$S \rightarrow id = E;$
$E \rightarrow E+T/T$
$T \rightarrow T*F/F$
$F \rightarrow id$



⇒ **Semantic Analyser**

⇓

parse tree semantically verified

Intermediate code Generation

⇓

$t_1 = b * c$
$t_2 = a + t_1$
$x = t_2$

Code optimization

⇓

$t_1 = b * c$
$t_2 = a + t_1$

Target code generation

⇓

Mul R_1, R_2	$a \rightarrow R_0$
Add R_0, R_2	$b \rightarrow R_1$
Mov R_2, x	$c \rightarrow R_2$

INTRODUCTION TO LEXICAL ANALYSER (L-2)

⇒ This is the only phase that Reads the Input character by character

⇒
```
int max(x,y)
  int x,y;
  /* find max of x and y */
  {
     return (x>y ? x : y);
  }
```
} 35 Tokens are present

int = 1 token
max = 1 token
return = 1 token

⇒ printf ("%d Hai", &x); ⇒ 8 Tokens

(1 Token for "%d Hai", ; 1 2 3 4 5 6 7 8)

⇒ Syntax Analyser is also called parser

Grammar:

E → E+E / E*E / id

⇒ id+id*id = Given string

LMD	RMD	LMD	RMD
E → E+E	E → E+E	E → E*E	E → E*E
→ id + E*E	→ E + E*E	→ E+E*E	→ E* id
→ id + id*E	→ E + E*id	→ id+E*E	→ E+E*id
→ id + id*id	→ E + id*id	→ id+id*E	→ E+id*id
	→ id+ id*id	→ id+id*id	→ id+id*id

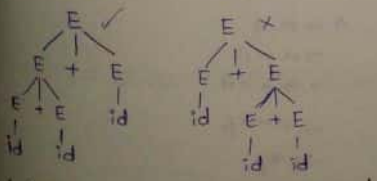
⇒ If a Gramar has more than one derivation trees for the same string then the Gramar is Ambigious

⇒ Ambiguity problems are undecidable

AMBIGIOUS GRAMMARS AND MAKING THEM UNAMBIGIOUS (L-3)

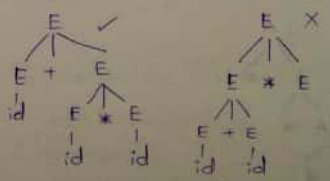
E → E+E / E*E / id

id+id+id ⇒ Associativity ×



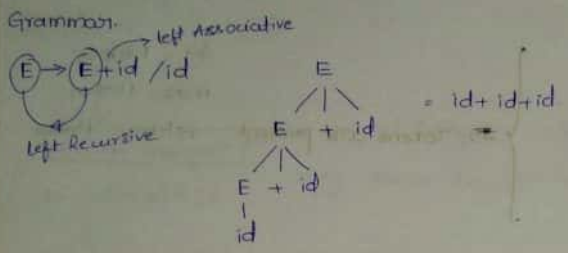
Leftmost plus should be evaluated first.

id+id*id ⇒ precedence (×)



Highest precedence one should be evaluated first (* should be evaluated first)

To avoid the above Ambiguity we have to restrict the growth of the Grammar.



⇒ The Grammar is said to be left recursive if the left most Symbol in the RHS = LHS

⇒ Inorder to overcome the Associativity we need to define the Grammar to be Left Recursive

⇒ E → E+T/T
T → T*F/F
F → id

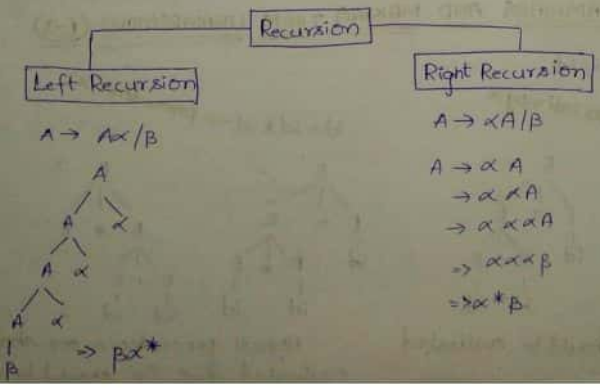
Associativity = Recursion
Precedence = Levels

⇒ 2↑3↑2 = 2^{3^2} = $(2)^{(3^2)}$ = 2^9

A → A $B/B
B → B#C/C
C → C@D/D
D → d

$, #, @ = Left Associative
$ > $, # > #, @ > @
$ < # < @

LEFT RECURSION ELIMINATION AND LEFT FACTORING OF GRAMMARS (L-4)



$\Rightarrow \beta\alpha^*\ (A \to \beta\alpha^*)$

$A \to \beta A'$, $A' \to \alpha A'/\epsilon$ $\Leftrightarrow$ $A \to A\alpha/\beta$

If the grammar is of the form $A \to A\alpha/\beta$ then eliminate left Recursion by $A \to \beta A'$, $A' \to \alpha A'/\epsilon$

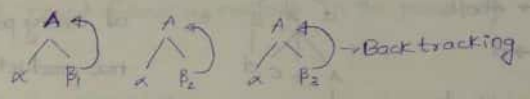
1) $E \to E+T/T$, α/β $\quad E \to TE'$, $E' \to \epsilon/+TE'$ = Left Recursion Eliminated

2) $A \to B_1A'/B_2A'/B_3A'$, $A' \to \alpha_1A'/\alpha_2A'/\alpha_3A'$ } is the eliminated LR of the Grammar $A \to A\alpha_1/A\alpha_2/A\alpha_3/\ldots/\beta_1/\beta_2/\beta_3\ldots$

3) Grammars can be classified into various categories

G

Ambigious	unambigious
Left Recursive	Right Recursive
Deterministic	Non-Deteiministic $(A \to \alpha\beta_1/\alpha\beta_2/\alpha\beta_3)$



$\Rightarrow A \to \alpha A'$, $A' \to \beta_1/\beta_2/\beta_3$

① $S \to iEtS/iEtSeS/a$, $E \to b$ | $S \to iEtSS'/a$, $S' \to eS/\epsilon$, $E \to b$

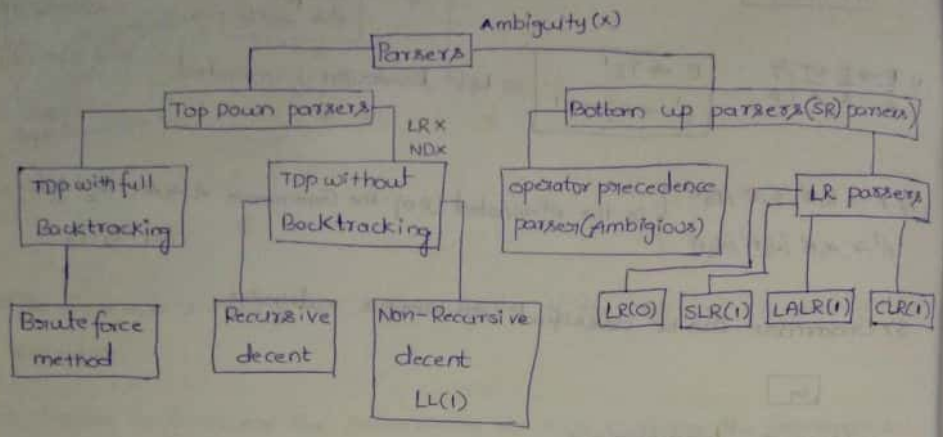
⇒ The elimination of Non-determinism doesnot guarentee the elimination of Ambiguity.

② $S \to aSSbS/aSaSb/abb/b$ ⇒ $S \to aS'/b$, $S' \to SSbS/SaSb/bb$, $S' \to SS''$, $S'' \to SbS/aSb$

③ $S \to bSSaaS/bSSaSb/bSb/a$ ⇒ $S \to bSS'/a$, $S' \to SaaS/SaSb/b$, $S' \to SaS''/b$, $S'' \to aS/Sb/$

PARSERS

⇒ parsers are nothing but syntax Analysers. (L-5)



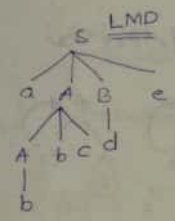
⇒

S → aABe

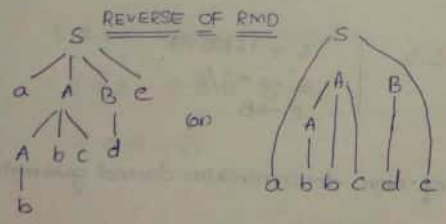
A → Abc/b

B → d

w = abbcde



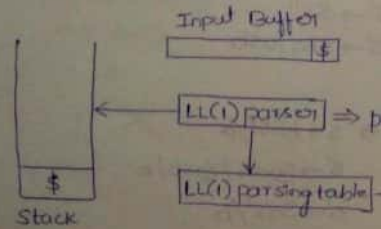
The main thing that we must assure is at every point when we have more than two productions to be choosen "which one to choose"



⇒ The main decision that should be made is if I see a terminal when to Reduce.

LL(1) PARSER / NON RECURSIVE DECENT PARSER

Left to Right, Left most Derivation. (1) = No. of look aheads



Now, Befo

functions

FIRST():

S → a ABC

A → b

B → c

C → d

D → e

S → ABCD

A → b

B → c

C → d

D → e

FOLLOW():

⇒ what is

deri vati

⇒ follow o

S → ABCD

A → b/ε

B → c/ε

C → d

D → e

⇒ Now, if n

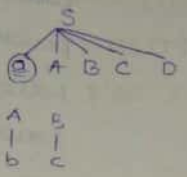
Variable

Now, Before constructing the LL(1) parsing table we should know two functions they are FIRST AND FOLLOW

(7)

FIRST():

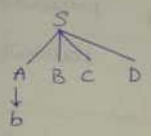
S → a ABCD
A → b
B → c
C → d
D → e



FIRST(S) = a
FIRST(A) = b
FIRST(B) = c
FIRST(C) = d
FIRST(D) = e

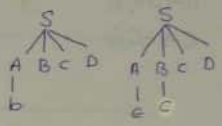
S → ABCD
A → b
B → c
C → d
D → e

FIRST(S) = b



S → ABCD
A → b/ϵ
B → c (small c)
C → d
D → e

FIRST(S) = {b, c}



FOLLOW():

⇒ what is the terminal which could follow a variable in the process of derivation.

⇒ follow of start symbol always contain '$'

S → ABCD
A → b/ϵ
B → c/ϵ
C → d
D → e

FOLLOW(S) = {$}
FOLLOW(A) = FIRST(B,C,D) = {c,d,e}
FOLLOW(B) = FIRST(C,D) = {d}
FOLLOW(C) = FIRST(D) = {e}

FOLLOW NEVER CONTAIN "EPSILON"

* Now, If nothing is following a variable i.e If nothing is at the RHS of a variable then the follow of that variable is the follow of LHS

∴ FOLLOW(D) = FOLLOW(S) = {$}

EXAMPLES ON HOW TO FIND FIRST AND FOLLOW IN LL(1) PARSER (L-6)

1) S → ABCDE
A → a/ϵ
B → b/ϵ
C → c
D → d/ϵ
E → e/ϵ

FIRST(S) = {a,b,c}
FIRST(A) = {a,ϵ}
FIRST(B) = {b,ϵ}
FIRST(C) = {c}
FIRST(D) = {d,ϵ}
FIRST(E) = {e,ϵ}

FOLLOW(S) = {$}
FOLLOW(A) = {b,c}
FOLLOW(B) = {c}
FOLLOW(C) = {d,e,$}
FOLLOW(D) = {e,$}
FOLLOW(E) = {$}

2) S → Bb/cd
B → aB/ϵ
C → cC/ϵ

FIRST(S) = {a,b,c,d}
FIRST(B) = {a,ϵ}
FIRST(C) = {c,ϵ}

FOLLOW(S) = {$}
FOLLOW(B) = {b}
FOLLOW(C) = {d}

3) E → TE'
E' → +TE'/ϵ
T → FT'
T' → *FT'/ϵ
F → id/(E)

FIRST(E) = {id, (}
FIRST(E') = {+,ϵ}
FIRST(T) = {id, (}
FIRST(T') = {*,ϵ}
FIRST(F) = {id, (}

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

4) S → ACB/CbB/Ba
A → da/BC
B → g/ϵ
C → h/ϵ

FIRST(S) = {ϵ,d,g,h,b,a}
FIRST(A) = {d,g,h,ϵ}
FIRST(B) = {g,ϵ}
FIRST(C) = {h,ϵ}

FOLLOW(S) = {$}
FOLLOW(A) = {h,g,$}
FOLLOW(B) = {$,a,h,g}
FOLLOW(C) = {g,$,b,h}

5) S → aABb
A → C/ϵ
B → d/ϵ

FIRST(S) = {a}
FIRST(A) = {C,ϵ}
FIRST(B) = {d,ϵ}

FOLLOW(S) = {$}
FOLLOW(A) = {d,b}
FOLLOW(B) = {b}

6) S → aBDh
B → cC
C → bC/ϵ
D → EF
E → g/ϵ
F → f/ϵ

FIRST(S) = {a}
FIRST(B) = {c}
FIRST(C) = {b,ϵ}
FIRST(D) = {g,f,ϵ}
FIRST(E) = {g,ϵ}
FIRST(F) = {f,ϵ}

FOLLOW(S) = {$}
FOLLOW(B) = {g,f,h}
FOLLOW(C) = {g,f,h}
FOLLOW(D) = {h}
FOLLOW(E) = {f,h}
FOLLOW(F) = {h}

CONSTRUC

7
E → TE'
E' → +TE'/ϵ
T → FT'
T' → *FT'/ϵ
F → id/(E)

NOW THE

	id
E	E→T
E'	
T	T→F
T'	
F	F→

2) S → (S)

	S
S	S

Now, w =

CONSTRUCTION OF LL(1) PARSING TABLE (1-7)

$E \to TE'$	FIRST(E) = {id, (}	FOLLOW(E) = {$,)}
$E' \to +TE' / \epsilon$	FIRST(E') = {+, ε}	FOLLOW(E') = {$,)}
$T \to FT'$	FIRST(T) = {id, (}	FOLLOW(T) = {+, $,)}
$T' \to *FT' / \epsilon$	FIRST(T') = {*, ε}	FOLLOW(T') = {+, $,)}
$F \to id / (E)$	FIRST(F) = {id, (}	FOLLOW(F) = {*, +, $,)}

NOW THE LL(1) PARSING TABLE LOOKS LIKE

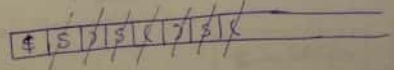
	id	+	*	(	)	$
E	E→TE'			E→TE'		
E'		E'→+TE'			E'→ε	E'→ε
T	T→FT'			T→FT'		
T'		T'→ε	T'→*FT'		T'→ε	T'→ε
F	F→id			F→(E)		

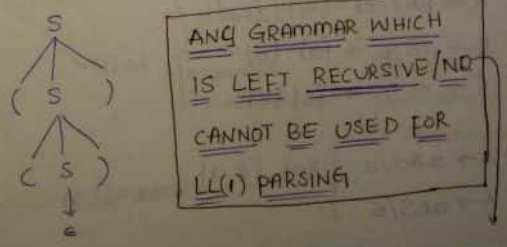
⇒ $S \to (S) / \epsilon$

	(	)	$
S	S→(S)	S→ε	S→ε

All Grammars are not feasable for LL(1) parsing.

Now, ω = (())$





ANY GRAMMAR WHICH IS LEFT RECURSIVE/NO CANNOT BE USED FOR LL(1) PARSING

Non-deterministic

CHECK WHETHER THE GRAMMARS ARE LL(1) OR NOT

1) S→aSbS / bSaS/ϵ
{a} {b} {a,b,$} ⇒ Not LL(1)
↳ which means the production must be placed under 'S' row and 'a' column.
↳ Since the S→ϵ production should be placed under the follow(s) which are {a,b}, we have already placed production in S' row and 'a' column which means a single CELL has more than one entry, so the grammar is not LL(1) ✗

2) S→aABb ⇒ {a}
A→c/ϵ ⇒ {c} and follow(A) = {d,b}
B→d/ϵ ⇒ {d} and {b} ⇒ IS LL(1) ✓
(A→c, A→ϵ)

3) S→A/a {a}∩{a} = Not LL(1) ✗ (conflict)
A→a {a}

4) S→aB/ϵ {a} {$} ⇒ Not x^n
B→bC/ϵ {b} {$} ⇒ NO x^n
C→cS/ϵ {c} {$} ⇒ No x^n
} LL(1) ✓

5) S→AB
A→a/ϵ ⇒ {a} {b,$} ⇒ No x^n
B→b/ϵ ⇒ {b} {$} ⇒ No x^n
} LL(1) ✓

6) S→aSA/ϵ ⇒ {a} {c,$} ⇒ No x^n
A→c/ϵ ⇒ {c} {c} ⇒ x^n is there
} ✗

7) S→A
A→Bb/cd ⇒ {a,b} {c,d}
B→aB/ϵ ⇒ {a} {b}
C→cC/ϵ ⇒ {c} {d}
} LL(1) ✓

8) S→aAa/ϵ {a} {$,a}
A→abS/ϵ
✗ Not LL(1)

9) S→iEtSS'/a {i} {a}
S'→ es/ϵ {e} {e}
E→b
} Not LL(1) ✗

RECURSIVE

E→iE'
E'→+iE'/ϵ

⇒ The parser

we are goin

```
E()
{
  1 if (l=
  2{
        m
    }3 E'
}4
l= getchar()
```

```
main()
{
  1) E()
  2) if(l==
  3) printf(
}
```

OPERATO

OPERATOR

A Gramma
operator G

⇒ NO Two
⇒ No eps

⇒ This p

RECURSIVE DECENT PARSER (L-8)

$E \rightarrow iE'$

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

⇒ The parser is called Recursive Decent parser because for every variable we are going to write Recursive functions.

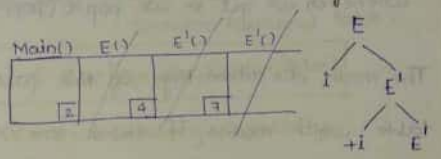
```
E()
{
1 if (l==='i')
2 {
      match('i');
  3 E'();
  }
}4
l=getchar();
```

```
E'()
{
1) if(l=='+')
  {
    2) match('+');
    3) match('i');
      4) E'();
  }
5)else 6)return;
}7)
```

```
match(char t)
{
  if(l==t)
    l=getchar();
  else
    printf("error");
}
```

{(i+i $)} is generated from above grammar

```
main()
{
  1) E()
  2) if(l=='$')
  3) printf("parsing success");
}
```



⇒ This parser uses Recursion stack for parsing.

⇒ Here l= look Ahead

OPERATOR GRAMMAR AND OPERATOR PRECEDENCE PARSER (L-9)

OPERATOR GRAMMAR

A Grammar that is used to define the mathematical operations is called operator Grammar (with some Restrictions)

⇒ No Two variables must be Adjacent

⇒ No epsilon productions

$E \rightarrow E+E/E*E/id$ (✓)

$E \rightarrow EAE/id$
$A \rightarrow +/*$
} Not operator Grammar

Not LL(1) ×

$S \rightarrow SAS/a$ $A \rightarrow bSb/b$	⇒	$S \rightarrow SbSbS/SbS/a$ $A \rightarrow bSb/b$
Not operator precedence		Operator precedence Grammar

⇒ This parser passes the Ambigious grammars by creating operator Relation Table

The operator Relational table looks like

	id	+	*	$
id	–	⋗	⋗	⋗
+	⋖	⋗	⋖	⋖
*	⋖	⋗	⋗	⋗
$	⋖	⋖	⋖	=

The given grammar is

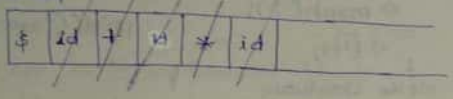
$E \to E+E / E*E / id$

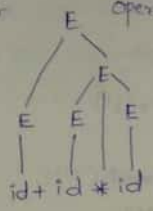
⇒ id will have higher precedence than any other operator

⇒ '$' will have least precedence than any other operator

W = id + id * id $

look head





⇒ Top of the stack will be $

Now The Algorithm goes like this

1) when the top of the Stack is ⋖ than the lookAhead then push it and whenever we get ⋗ we pop it (popping means actually we Reduced it)

The main disadvantage of this parser is the size of the operator Relational table which means if there are 4 operators then Size of the table is $16(4^2)$ and if there are 5 operators then there would be 25 entries(5^2). So Generally if there are 'n' operators, the size of the table is $O(n^2)$

Now, To reduce the size of the table we use operator function table

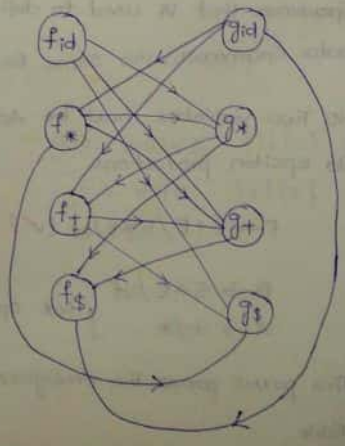
	id	+	*	$ → function 'G'
id	–	⋗	⋗	⋗
+	⋖	⋗	⋖	⋗
*	⋖	⋗	⋗	⋗
$	⋖	⋖	⋖	=

↓

Function (F)

⋗ = $F_{id} \to g_{id}$ (F → G)

⋖ = $g_{id} \to f_{id}$ (G → F)



Now, the Longest path from f_{id} is $f_{id} \to g_* \to f_+ \to g_+ \to f_\$$

Now the Longest path from g_{id} is $g_{id} \to f_* \to g_* \to f_+ \to g_+ \to f_\$$

Similarly find the longest paths from each node the function table looks like

	id	+	*	$
f	4	2	4	0
g	5	1	3	0

$f_{id} \xrightarrow{1} g_* \xrightarrow{2} f_+ \xrightarrow{3} g_+ \xrightarrow{4} f_\$ = 4$

$f_+ \xrightarrow{1} g_+ \xrightarrow{2} f_\$ = 2$

Now if i need to compare $(+,+) \Rightarrow f_+ \quad g_+$

$\Rightarrow \quad 2 \quad 1 \Rightarrow 2 > 1 \Rightarrow (+ \gtrdot +)$

∴ The Size of the table = $\boxed{O(2n)}$ n = no. of operators.

⇒ In the functional table we dont have Blank entries (Blank Entries are nothing but errors) so, the error detecting capability of the functional table is less than that of operator Relation table (we have Blank entries in operator Relational table)

$$\boxed{EDC\,[\text{Operator Functional Table}] < EDC\,[\text{Operator Relation Table}]}$$

EDC = Error Detecting Capability.

2) P → SR/S
R → bSR/bS
S → wbS/w
w → L*w/L
L → id

⇒

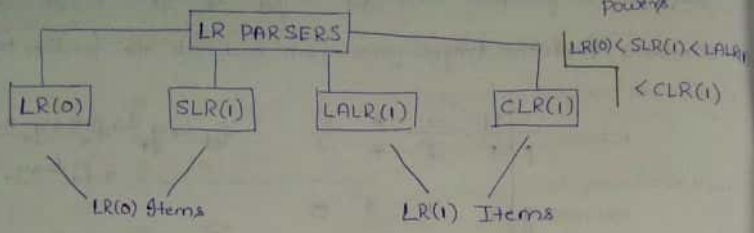
P → SbP/SbS/S
S → wbS/w
w → L*w/L
L → id

	id	*	b	$
id	−	⋗	⋗	⋗
*	⋖	⋖	⋗	⋗
b	⋖	⋖	⋖	⋗
$	⋖	⋖	⋖	−

⇒ Here * is defined as Right Associative (w → L*w) so the Right side Star has highest precedence

∴ * ⋖ *

LR PARSING, LR(0) ITEMS AND LR(0) PARSING TABLE (L-10)



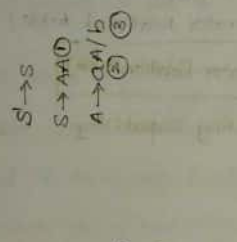
1) $S \to AA$
   $A \to aA/b$

} In LR parsers we have CLOSURE and GOTO Operations

The Augmented Grammar is $S' \to S$, $S \to AA$, $A \to aA/b$

Any production with a dot in the RHS is called an item

The LR(0) parsing Tree is,

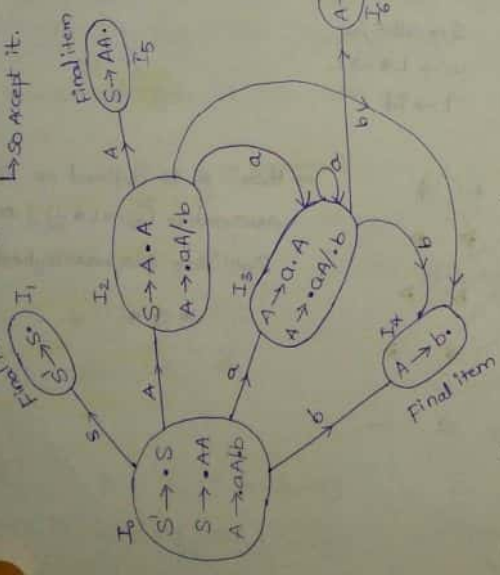


The parsing table is

	ACTION			GOTO	
	a	b	$	A	S
0	S_3	S_4		2	1
1			Accept		
2	S_3	S_4		5	
3	S_3	S_4		6	
4	R_3	R_3	R_3		
5	R_1	R_1	R_1		
6	R_2	R_2	R_2		

Same goes on...

CANONICAL COLLECTION OF LR(0) ITEMS



1. While writing the Reduce moves / while inspecting final items go to the productions and check

$S' \to S$
$S \to AA$ ①
$A \to aA$ ②
$A \to b$ ③

Now I_4 is $A \to b\cdot$ (Reduction) $\Rightarrow I_4 \to R_3$

LR(0) PARSING EXAMPLE AND SLR(1) TABLE (L-11)

$S' \to S$
$S \to AA$ ①
$A \to aA/b$
② ③

Let the given string be aabb$

Input pointer → a a b b $ (↑ ↑ ↑ ↑ ↑)

0	a	3	a	3	b	4	A	6	A	6	A	2	b	4	A	5	5	1	Accept

⇒ Always the top of the stack contains state (and first state will be zero)

⇒ Initially I_0 on 'a' is S_3 which means shift the input you are looking at and as well as the state no on to the stack ⇒ [Input = a, state = 3] = a|3 and increment the input pointer [continue]

⇒ Now top of the stack is '4' and input pointer is b which means R_3' which states that Reduce the production no 3. which means reduce the previous 'b' (previous symbol). ↳ (A → b)

⇒

⇒ Now in the stack how could we make Reduce move is look the RHS of the production that must be Reduced, and find the length of RHS in this example A → b is the production ⇒ length of RHS = 1 (∵ |b| = 1) which means pop 2 symbols (1×2) from stack. (If the length of RHS is 'x' then pop '2x' elements from stack) and push the LHS symbol onto the stack, and see the stack for the last used state number it turns out that it is '3' and look what '3' on 'A' is generating = 6, so push '6' onto the stack. When we see reduce moves we don't increment input pointer.

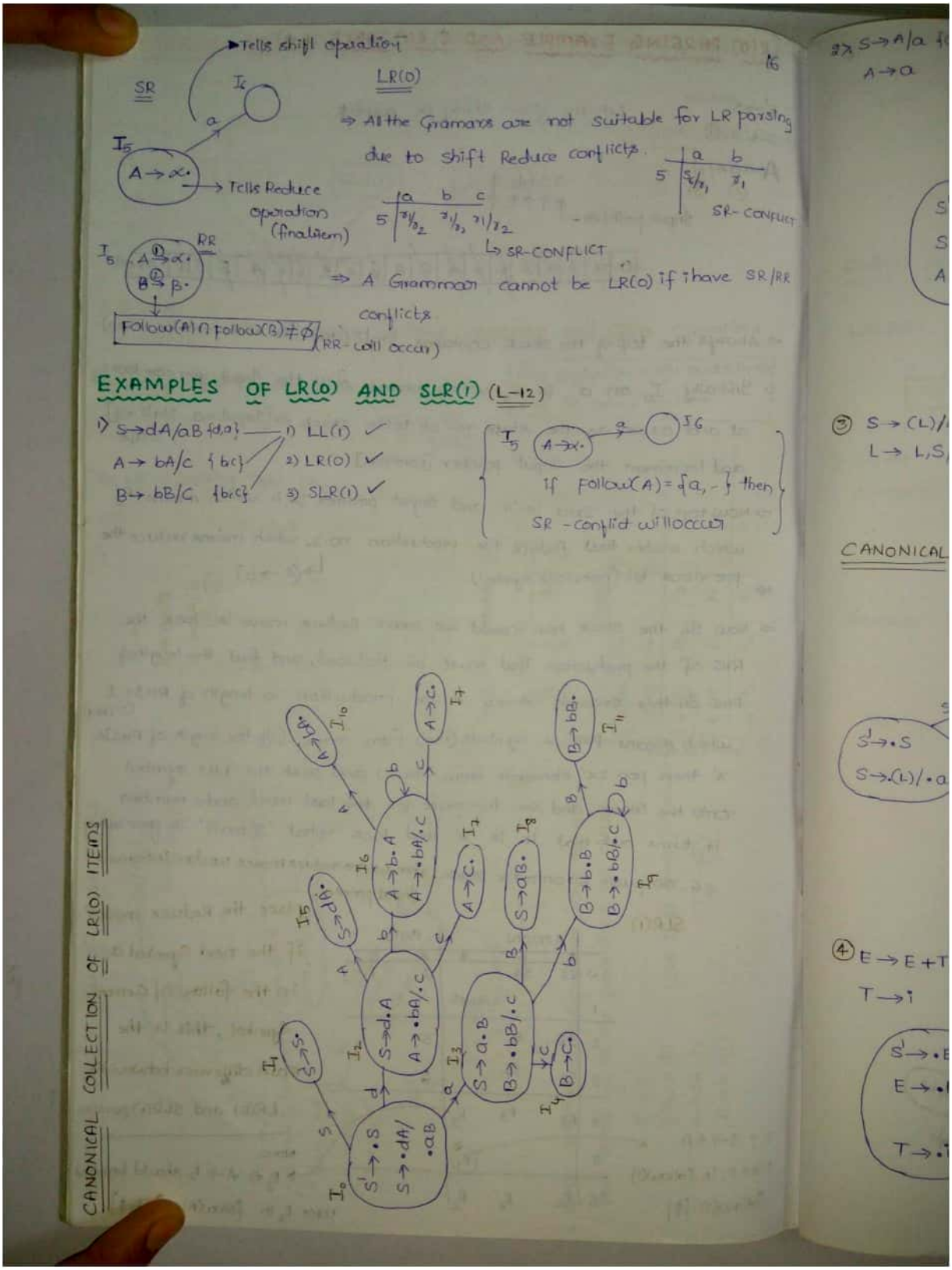
SLR(1)

	ACTION			GOTO	
	a	b	$	A	S
0	S3	S4		2	1
1			Accept		
2	S3	S4		5	
3	S3	S4		6	
4	R3	R3	R_3'		
5			R_1		
6	R_2	R_2	R_2		

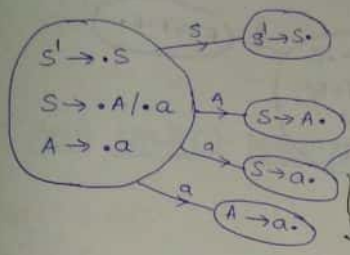
$R_1 : S \to AA$
Place R_1 in follow(S)
follow(S) = {$}

Place the Reduce moves If the next Symbol is in the follow of current Symbol, this is the main difference between the LR(0) and SLR(1) parsers

Now,
→ $R_3 \Rightarrow A \to b$ should be reduced
place R_3 in follow(A) = {a, b, $}



② $S \to A/a$ fafafa Not LL(1) ✗
$A \to a$ LR(0) ✗ } Ambiguous Grammar
SLR(1) ✗

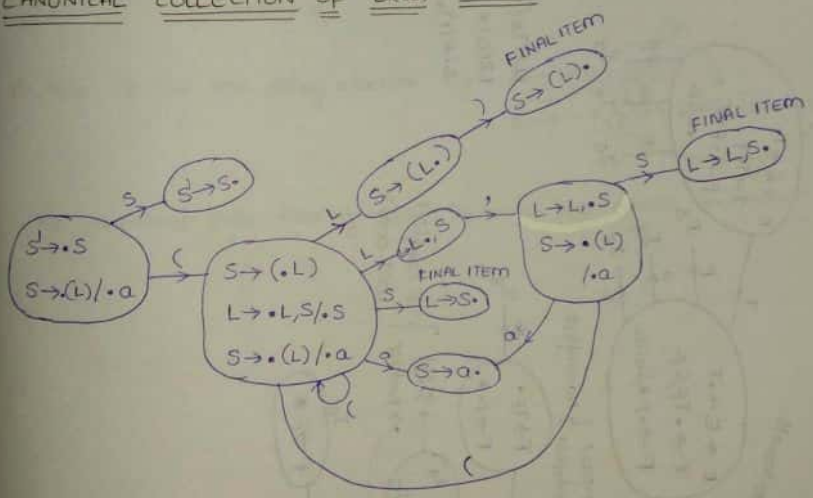


follow(S) = $ } Not SLR(1)
follow(A) = $

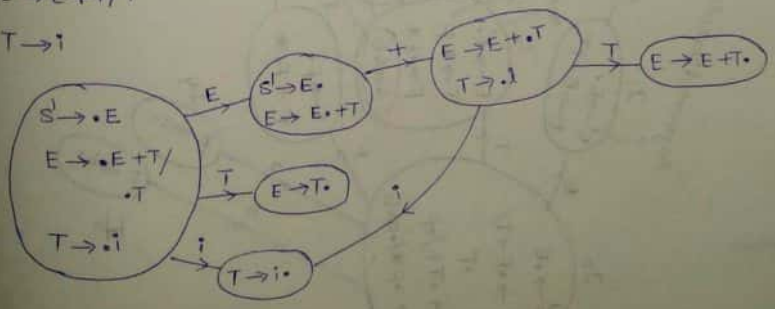
2 Reduce moves from the same state
RR-conflict so LR(0) ✗ (If we get 2 reduce moves from the same state then the grammar is not LR(0).

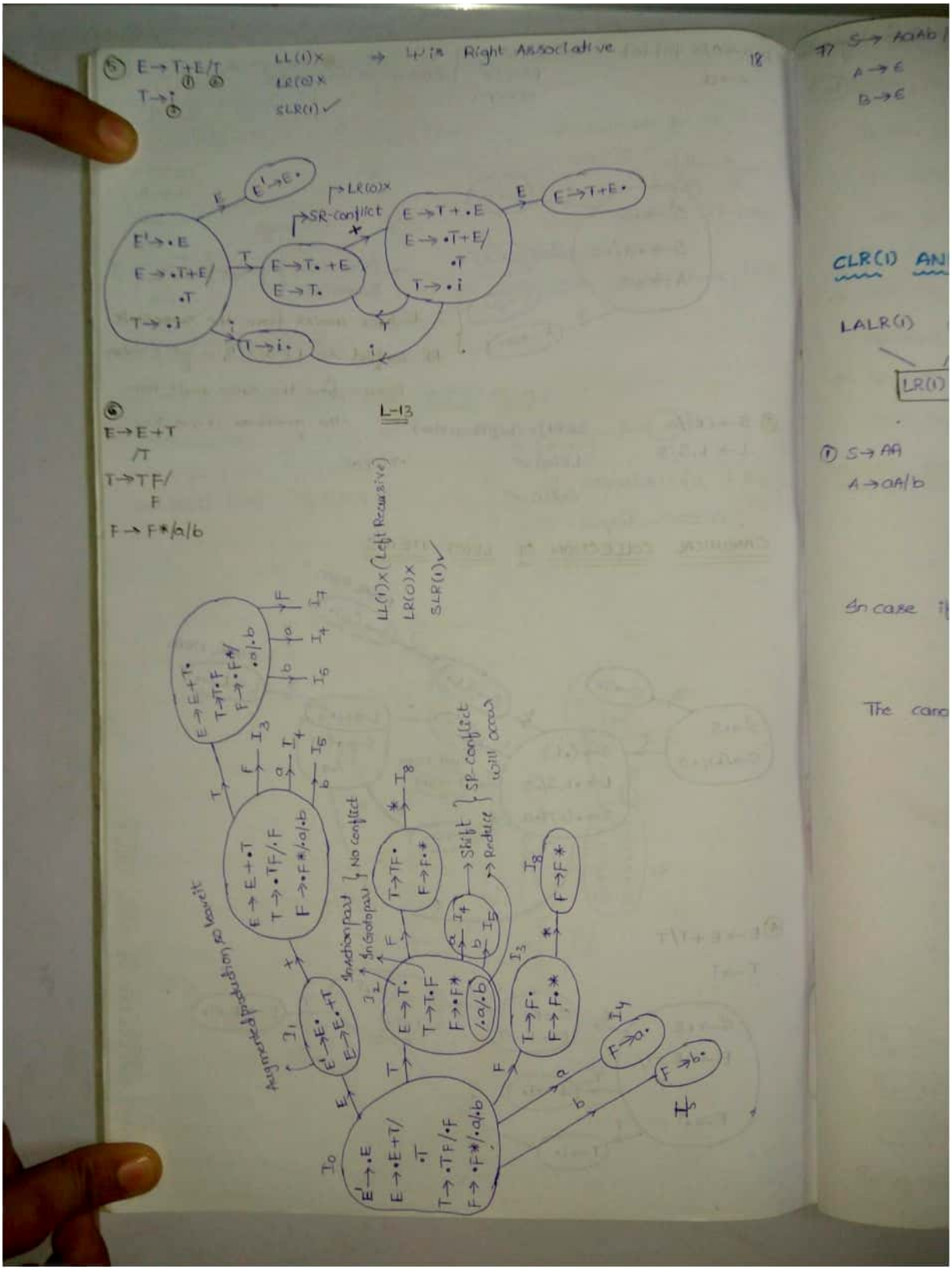
③ $S \to (L)/a$ LL(1) ✗ (Left Recursive)
$L \to L,S/S$ LR(0) ✓ ⇒ fail
SLR(1) ✓

CANONICAL COLLECTION OF LR(0) ITEMS



④ $E \to E+T/T$
$T \to i$

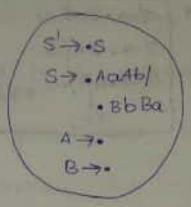




7) $S \to AaAb / BbBa \{a,b\}$ LL(1) ✓

$A \to \epsilon$ LR(0) ×

$B \to \epsilon$ SLR(1) ×



	a	b	$
0	r3/r4	r3/r4	

RR- conflict

CLR(1) AND LALR(1) PARSERS L-14

LALR(1) CLR(1)

LR(1) Items = LR(0) Items + Lookhead

1) $S \to AA$

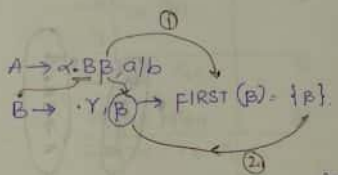
$A \to aA/b$

$S' \to .S$ is the Augmented Gramar

$S \to .AA$

$A \to .aA/.b$

In case if we are doing closure for $A \to \alpha.B\beta, a/b$



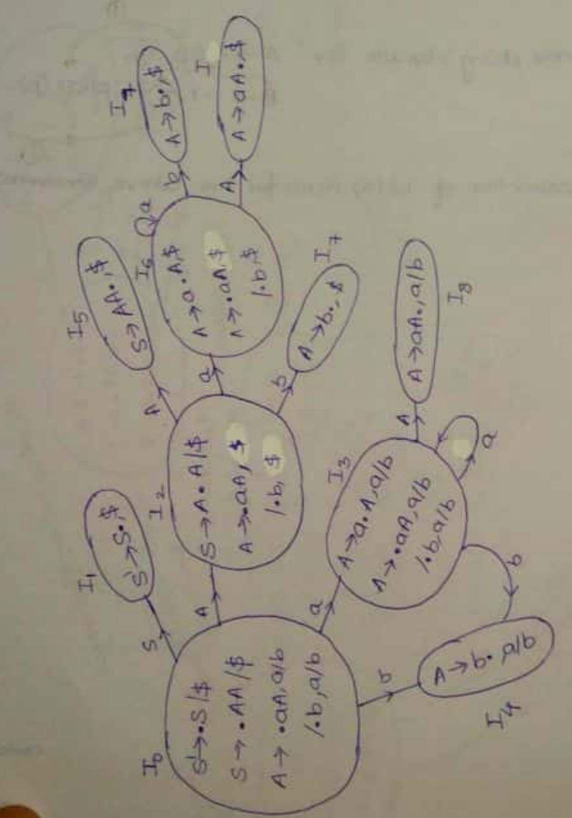
The canonical collection of LR(1) items for the above Grammar is

Augumented Grammar

$S' \to \cdot S, \$$
$S \to \cdot AA / \$$
$A \to \cdot aA / \cdot b \to a/b$

→ look ahead is always \$ for Augmented production.

CANONICAL COLLECTION OF LR(1) ITEMS



⇒ The Goto part and shift part will be the same as LR(0), SLR(1) parsing tables, the main difference arises in the placement of the final items In the LR(0) and SLR(1) we are going to place in the entire row and the follow of LHS (in SLR(1)) respectively. But in CLR and LALR(1) we are going to place the reduce moves only in look ahead symbols.

⇒ From the above diagrams $[I_3, I_6]$ have same LR(0) items but differ in look heads

⇒ Similarly $[I_4, I_7]$, $[I_8, I_9]$ also have same LR(0) items but differs in Look aheads.

CLR(1) PARSING TABLE

	a	b	$	S	A
0	S3	S4			2
1					
2	S6	S7			5
3	S3	S4			8
4	r3	r3			
5			r1		
6	S6	S7			9
7			r3		
8	r2	r2			
9			r2		

LALR(1) TABLE

	a	b	$	S	A
0	S_{36}	S_{47}			2
1					
2	S_{36}	S_{47}			5
36	S_{36}	S_{47}			89
47	r3	r3	r3		
5			r1		
89	r2	r2	r2		

$I_3, I_6 \Rightarrow I_{36}$

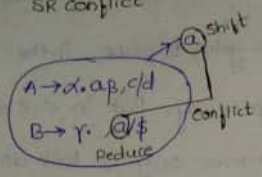
$I_4, I_7 \Rightarrow I_{47}$

$I_8, I_9 \Rightarrow I_{89}$

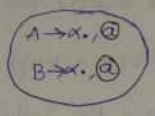
[No of states in CLR(1)] ≥ [[No of states in SLR(1)] = [No of states in LALR(1)] = [No of states in LR(0)]]

CONFLICTS AND EXAMPLES OF CLR(1) AND LALR(1) (L-15)

SR conflict

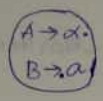


RR conflict



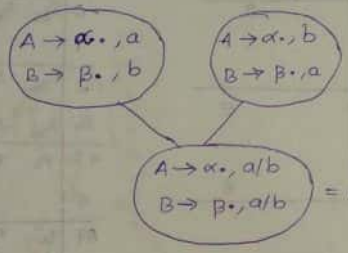
LR(1) Items

LR(0) Items



⇒ If the grammar is not CLR(1) then the Grammar is not LALR(1) because we reduce the size of the table but not the conflicts in LALR(1) parser

⇒ If the Grammar is CLR(1) then it may or maynot be LALR(1)



1) S → AaAb / BbBa

A → ∈

B → ∈

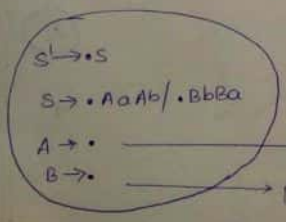
LL(1) ×

LR(0) ×

SLR(1) ×

CLR(1) ✓

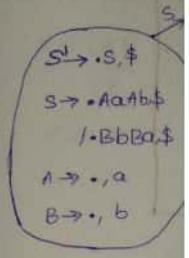
LALR(1) ✓



→ placed under follow of A = {a, b}

→ placed under follow of B = {a, b}

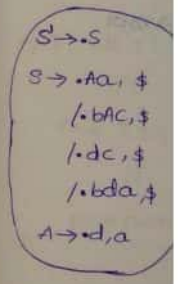
} ⇒ Not SLR(1)

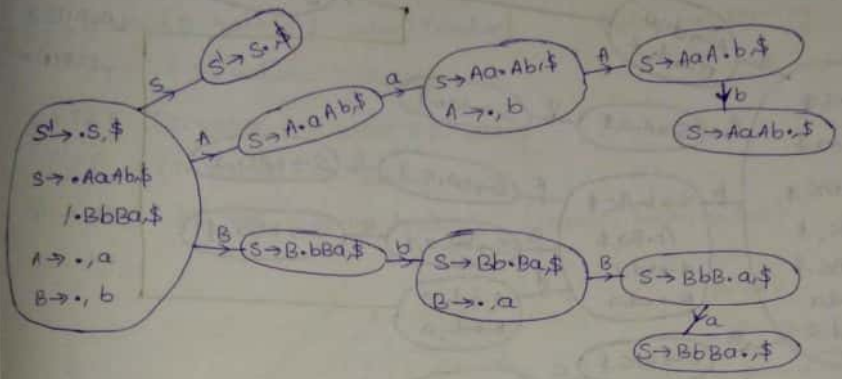


(2)

S → Aa / bAC /

A → d





because

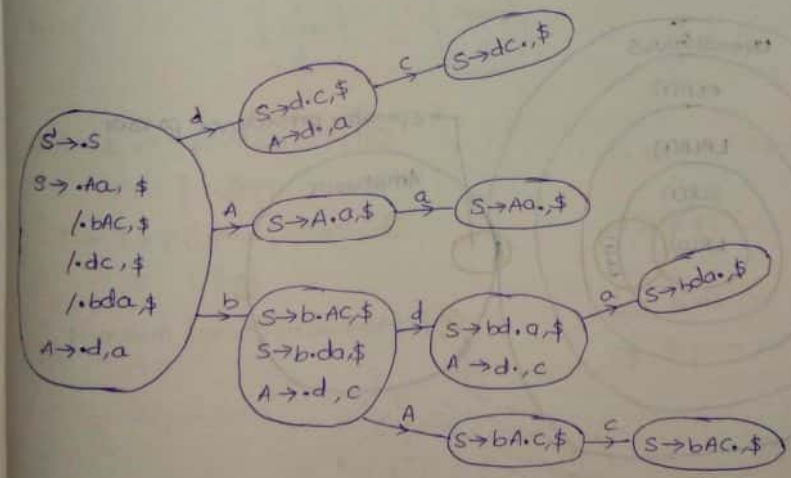
parser

(2)

S→Aa/bAc/dc/bda

A→d

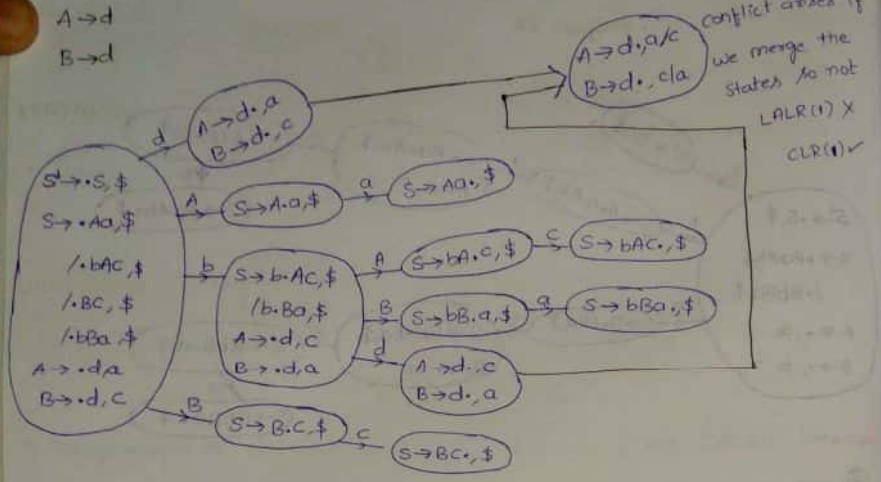
LL(1) ×
LR(0) ×
SLR(1) ×
CLR(1) ✓
LALR (1)



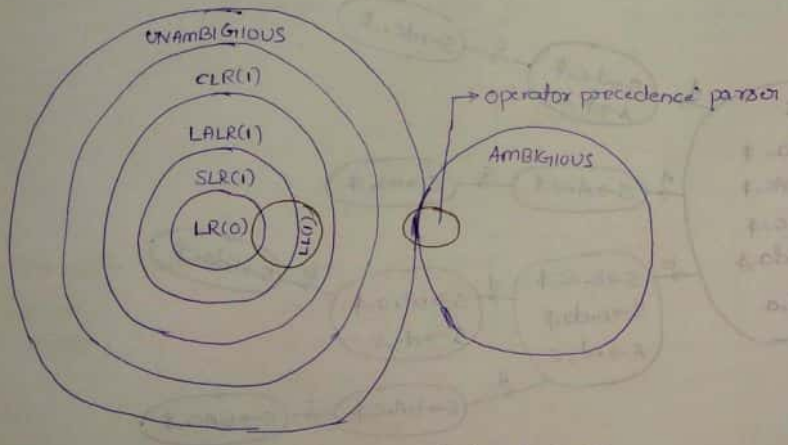
③ S→Aa/bAc/Bc/bBa

A→d

B→d



COMPARISION OF THE PARSERS (L-16)



⇒Every Grammar which is LL(1) is definitely LALR(1)

SYNTAX DIRE

⇒ Grammar +

1) E → E+T {E.
/T {E.
T → T*F {T.
/F {T.
F → num {F.

2) E → E+T {
/T {
T → T*F {
/F {
F → num {

num
2.

onflict arises if we merge the states so not

LALR(1) ✗

CLR(1) ✓

SYNTAX DIRECTED TRANSLATION (L-17)

⇒ Grammar + Semantic Rules = SDT

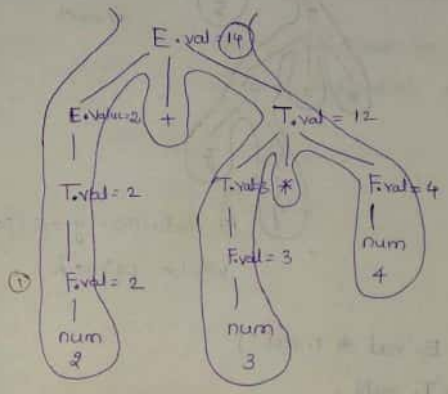
1) $E \to E+T$ {E.value = E.value + T.value} Bottom-up parsing.

/T { E.value = T.value }

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

/F {T.value = F.value}

$F \to num$ {F.val = num.lvalue} {lvalue = lexim value}



2+3*4
= 2+(3*4)
= 2+12 = 14

2) $E \to E+T$ { printf("+"); } ①

/T { } ②

$T \to T*F$ { printf("*"); } ③

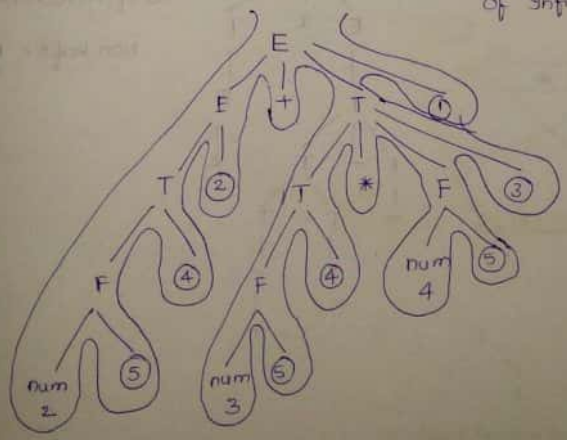
/F { } ④

$F \to num$ {printf(num.lval); } ⑤

2+3*4 (Top-down parser)

2 3 4 * +

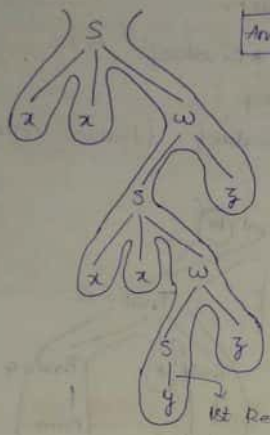
⇒ This is the SDT for convertion of infix to postfix Expression



Top down parsing

③ $S \to xxw$ {printf(1);}
/y {printf(2);}
$w \to Sz$ {printf(3);}

String xxxyzz



Ans: 2 3 1 3 1

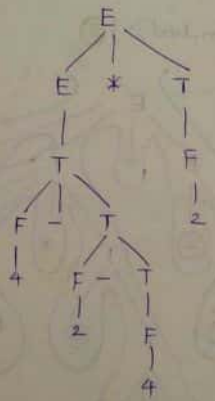
⇒ Convey the action on Reduce moves

1st Reduction $y \to S$ (printf(2)): '2' will be printed.

④

$E \to E * T$ {E.val = E.val * T.val;}
/T {E.val = T.val;}
$T \to F - T$ {T.val = F.val - T.val;}
/F {T.val = F.val;}
$F \to 2$ {F.val = 2;}
/4 {F.val = 4;}

w = 4 - 2 - (4 * 2)
w = ((4 - (2 - 4)) * 2)
= (4 - (-2)) * 2
= (4 + 2) * 2
= 12



No. of Reductions = No. of Non leafs = 10

⑤ $E \to E \# T$
/T
$T \to T \& F$
/F
$F \to num$

w = 2#3&5
= 2*3+5
= ((2*3)+5
= (6+30
= 40

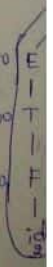
L-18

SDT TO BUIL

⑥ $E \to E_1 + T$ {
/T {
$T \to T_1 * F$ {
/F {
$F \to id$ {

W = 2+3*4

Tptr = 100
Tptr = 100
Tptr = 100



Conc

(5) E → E #T {E.val = E.val * T.val;}
/T {E.val = T.val}
T → T&F {T.val = T.val + F.val}
/F {T.val = F.val}
F → num {F.val = num.lvalue;}

ip = 2#3&5#6&4 what is the output

= 2*3+5*6+4
= ((2*3)+(5*6)+4)
= (6+30+4)
= 40

⇒ wrong because '+' is defined at highest level (Bottom level) and must be evaluated first and then multiplication must be Evaluated.

⇒ 2*(3+5)*(6+4)
= 2*(8)*(10)
= 160 ✓

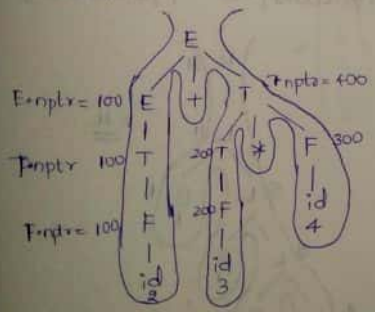
L-18

SDT TO BUILD SYNTAX TREE

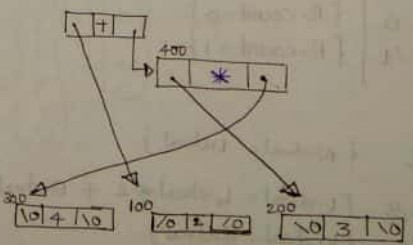
(6) E → E1+T {E.nptr = mknode(E1.nptr, '+', T.nptr)}
/T {E.nptr = T.nptr;}
T → T1*F {T.nptr = mknode(T1.nptr, *, F.nptr);}
/F {T.nptr = F.nptr}
F → id {F.nptr = mknode(null, id.name, null);}

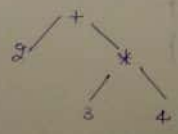
Returns Address
mknode = makenode
nptr = node pointer

W = 2+3*4



Concrete Syntax tree





Abstract Syntax tree

⑦ SDT FOR TYPE CHECKING

$E \rightarrow E_1 + E_2$ {if (E_1.type == E_2.type) && (E_1.type = int) then E.type = int else error}

/ $E_1 == E_2$ {if (E_1.type == E_2.type) && (E_1.type = int / boolean) then E.type = boolean else error}

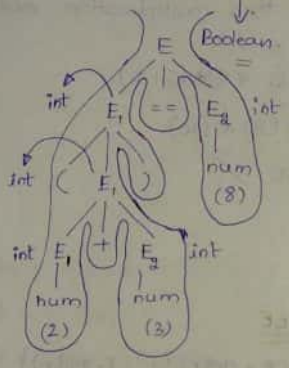
(E_1) { E.type = E_1.type }

/num { E.type = int; }

/True { E.type = boolean }

/False { E.type = boolean }

w = {(2+3) == 8} = Boolean Expression



⑧

	count all 1's	count all 0's	No. of Bits
$N \rightarrow L$	{ N.count = L.count }	⇒	⇒
$L \rightarrow L_1 B$	{ L.count = L_1.count + B.count }	⇒	⇒
/B	{ L.count = B.count }	⇒	⇒
$B \rightarrow 0$	{ B.count = 0 }	{B.count} = 1	{B.count = 1}
/1	{ B.count = 1 }	{B.count} = 0	{B.count = 1}

$N \rightarrow L$ { N.dval = L.dval }

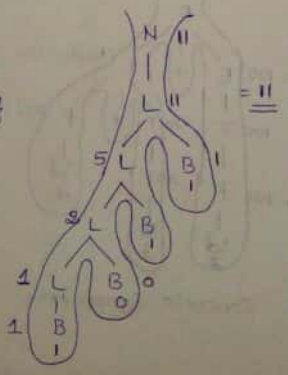
$L \rightarrow L_1 B$ { L.dval = L_1.dval * 2 + B.dval }

/B { L.dval = B.dval }

$B \rightarrow 0$ { B.dvalue = 0 }

/1 { B.dvalue = 1 }

w = 1011



L-19 S- ATT

d value

• (01) = $1/2^2$ =

decimal

• (11) ⇒ 3/4 =

⑨ $N \rightarrow L_1 . L_2$

$L \rightarrow L_1 B$ / B

$B \rightarrow 0$

/1

SDT TO GENE

⑩ $S \rightarrow id = E$

$E \rightarrow E_1 + T$ {

/T {

$T \rightarrow T_1 * F$ {

/F {

$F \rightarrow id$ {

L-19 S- ATTRIBUTED AND L- ATTRIBUTED DEFINITIONS

(01) → d value → $1/2^2 = 0.25$

(11) → decimal value → $3/4 = 0.75$

① $N \rightarrow L_1 \cdot L_2$	{N.dval = L_1.dval + (L2.dval / $2^{L_2.count}$)}
$L \rightarrow LB / B$	{L.dvalue = L_1.dval + B.dval} {L.count = L_1.count + B.count}
	{L.count = B.count} {L.dval = B.dvalue}
$B \rightarrow 0$	{B.count = 1, B.dvalue = 0}
/1	{B.count = 1, B.dvalue = 1}

SDT TO GENERATE THREE ADDRESS CODE

⑥ $S \rightarrow id = E$ {gen(id.name = E.place);}

$E \rightarrow E_1 + T$ {E.place = newTemp(); gen(E.place = E_1.place + T.place);}

/T {E.place = T.place;}

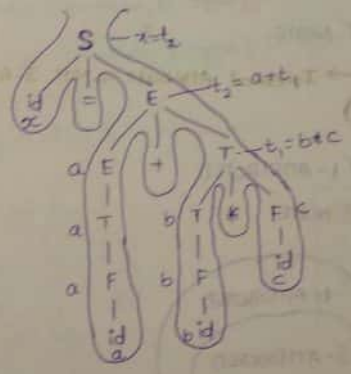
$T \rightarrow T_1 * F$ {T.place = newtemp(); gen(T.place = T_1.place * F.place);}

/F {T.place = F.place}

$F \rightarrow id$ {F.place = id.name}

w ⇒ x = a + b * c

$t_1 = b * c$
$t_2 = a + t_1$
$x = t_2$



DIFFERENCE BETWEEN S-ATTRIBUTED AND L-ATTRIBUTED SDT

S-ATTRIBUTED SDT

1) Uses only synthesized attributes

2) Semantic actions are placed at Right end of production

A → BCC { }

3) Attributes are evaluated during Bottom up parsing

L-ATTRIBUTED SDT

1) Uses Both inherited and Synthesized attributes. Each inherited attribute is restricted to inherit either from parent or Left siblings only

Ex: A → XYZ {Y.S = A.S ✓, Y.S = X.S ✗, Y.S = Z.S ✗}

2) Semantic Actions are placed anywhere on RHS

A → { } BC

/D { } E

/FG { }

3) Attributes are evaluated by traversing parse tree depth first left to Right

Inherited Attribute → Not S-ATTRIBUTED

① A → LM {L.i = f(A,i); M.i = f(L,S); A.S = f(M.S);}

A → QR {R.i = f(A,i), Q.i = f(R,i) ✗; A.S = f(Q,S);}

↳ NOT L-ATTRIBUTED

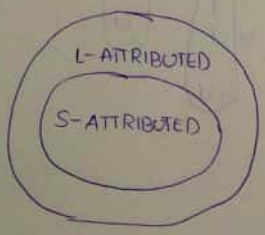
(A) S-ATTRIBUTED (B) L-ATTRIBUTED

(C) BOTH (D) NONE ✓

② A → BC {B.S = A.S} → Inherited Attribute ⇒ NOT S-ATTRIBUTED

a) S-ATTRIBUTED (B) L-ATTRIBUTED ✓

(C) BOTH (D) NONE



SDT TO A

(11) D → TL

T → int

/cha

L → L, I

/id

S-ATTRIB

(12) D → D,

/T

T → in

/c

ED SDT

DT

nd Synthesized
d attribute is
either from parent

S= ✓ X.S, Y.S = ✗ Z.S]

laced anywhere

ted by traversing
ft to Right

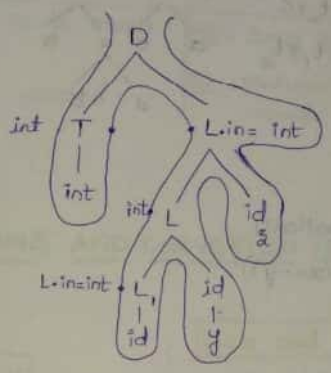
SDT TO ADD TYPE INFORMATION INTO SYMBOL TABLE

(11) $D \rightarrow TL$ {L.in = T.type} ⇒ Inherited Attribute
$T \rightarrow$ int {T.type = int;} ⇒ Synthesized Attribute
/char {T.type = char;}
$L \rightarrow L_1$, id {L_1.in = L.in, add type (id.name, L_1.in);}
/id {add type (id.name, L.in)}

L-ATTRIBUTED

int x, y, z;

x	int
y	int
z	int

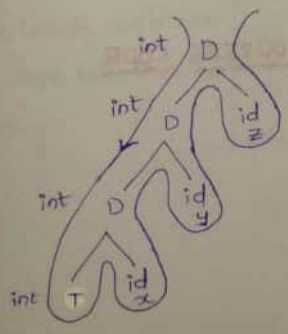


⇒ Evaluate the synthesized attribute when you last visit it

⇒ Evaluate the Inherited attribute when you first visit it

S-ATTRIBUTED SDT FOR THE SAME QUESTION

(12) $D \rightarrow D_1$, id {add-type (id.name, D_1.type)}
/T id {add-type (id.name, T.type), D.type = T.type}
$T \rightarrow$ int {T.type = int;}
/char {T.type = char;}

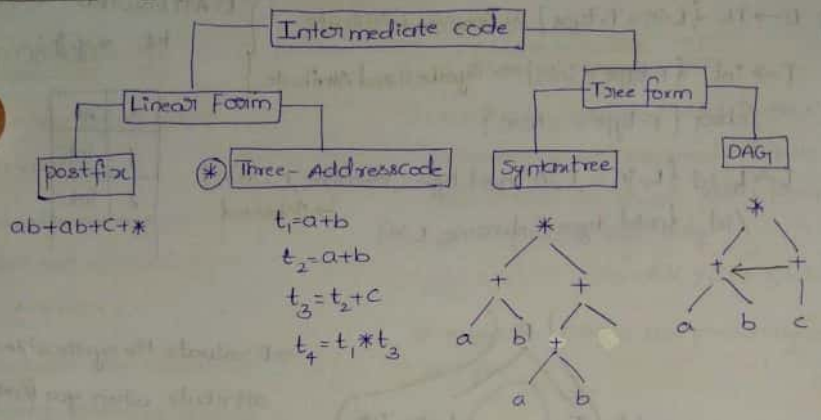


x	int
y	int
z	int

INTERMEDIATE CODE GENERATION

INTRODUCTION TO INTERMEDIATE CODE

Ex: (a+b) * (a+b+c)



TYPES OF 3-ADDRESS CODE

1) x = y op z (x=a+b (Binary operation))

2) x = op z (Unary operation (x=-y))

3) x=y (Assignment)

4) if x (rel op) y goto L (If 'x' (relational operator) y goto L)

5) goto L ⇒ (unconditional)

6) A[i] = x (Array Indexing)
   y = A[i]

7) x = *p ⇒ pointer
   y = &y ⇒ Address of variable assigned to another variable

VARIOUS REPRESENTATIONS OF 3-ADDRESS CODE

⇒ (a+b) * (c+d) + (a+b+c)

1) $t_1 = a+b$
2) $t_2 = -t_1$
3) $t_3 = c+d$
4) $t_4 = t_2 * t_3$
5) $t_5 = a+b$
6) $t_6 = t_5 + c$
7) $t_7 = t_4 + t_6$

(a+b+c)

QUADRAPLE

	Opr	Op1	Op2	Result
0)	+	a	b	t_1
1)	−	t_2	NULL	t_2
2)	+	c	d	t_3
3)	*	t_2	t_3	t_4
4)	+	a	b	t_5
5)	+	t_5	c	t_6
6)	+	t_4	t_6	t_7

Adv: Statements can be moved Around

Dis: More Space wasted

TRIPLE

	Opr	op1	op2
1)	+	a	b
2)	−	(1)	
3)	+	c	d
4)	*	(2)	(3)
5)	+	a	b
6)	+	(5)	c
7)	+	(4)	(6)

Adv: Space is not wasted

Dis: Statements cannot be moved

INDIRECT TRIPLE

i)	(1)
ii)	(2)
iii)	(3)
iv)	(4)
v)	(5)
vi)	(6)
vii)	(7)

Adv: State ments can be moved

Dis: Two Access of Memory.

BACK PATCHING AND CONVERTION TO 3-ADDRESS CODE

Back Patching

If (a<b) then t=1
else t=0

(i): if a<b goto (i+3)
(i+1): t=0
(i+2): goto (i+4)
(i+3): t=1
(i+4): Return.

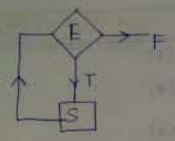
Leaving the Labels as blanks and filling them later is called Back patching

t_1 t_2 t_3

a<b and c<d or e<f

100) If (a<b) goto 103
101) t_1=0
102) goto 104
103) t_1=1
104) if (c<d) goto 107
105) t_2=0
106) goto 108
107) t_2=1
108) if (e<f) goto 111
109) t_3=0
110) goto 112
111) t_3=1
112) t_4=[t_1 and t_2]
113) t_5=[t_4 or t_3]

① While E do S



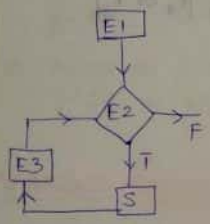
L: If (E==0) goto L1
S
Goto L
L1:

(or) if (E) goto L1
goto last
L1: S
goto L

② while (a<b) do
x = y+z

L: if a<b goto L1
goto last
L1: t = y+z
x = t
goto L
last:

③ for (E1; E2; E3)



for (i=0; i<10; i++)
a = b+c;

i = 0
L: if (i<10) goto L1
goto last
L1: t_1 = b+c
a = t_1
t = i+1
i = t
goto L
last:

④ Switch
{

}

TWO

x = A
t_1 = y
t_2 = t_1
t_3 = t
t_4 = B
x =

Row M

colum

```
switch (i+j)
{
  case (1): a=b+c;
            break;
  case (2): p=Q+R;
            break;
  default : x=y+z;
            break;
}
```

```
t = i+j
goto test
L1: t1 = b+c
    a = t1
    goto last
L2: t2 = Q+R
    p = t2
    goto last
L3: t3 = y+z
    x = t3
    goto last
test: if (t==1) goto L1
      if (t==2) goto L2
      goto L3
last :
```


TWO DIMENSIONAL ARRAY TO 3-ADDRESS CODE

$x = A[y\ z]$

$t_1 = y * 20$
$t_2 = t_1 + z$
$t_3 = t_2 * 4$
} $(y*20+z)\times 4$

t_4 = Base Address of A

$x = t_4[t_3]$

↳ Base offset Addressing
(Base Address + offset)

A : 10X20 –

A [4] [4]

00	01	02	03
10	11	12	13
20	21	22	23
30	31	32	33

A [2 3] → cross 2 rows, cross 3 columns
= 2*4+3 → No of elements in row
= 11

Row Major order: 00 01 02 03 10 11 12 13 20 21 22 23 30 31 32 33

column major order: 00 10 20 30 01 11 21 31 02 12 22 32 03 13 23 33

RUN ENVIRONMENT

⇒ Run time Environment means when you run the program what is the support that you need from the operating system.

STORAGE ALLOCATION STRATEGIES

1) Static

1) Allocation is done at compile time
2) Bindings do not change at Runtime
3) One Activation Record per procedure

Disadvantages

1) Recursion is not supported.
2) Size of data objects must be known at compile time
3) Data Structures cannot be created Dynamically

2) Stack

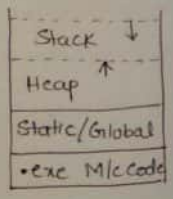
whenever a new activation begins, Activation record is pushed onto the Stack and whenever Activation ends, Activation record is popped off
⇒ Local variables are bound to fresh storage

Disadvantages

1> Local variables cannot be retained once activation end

3) Heaps

⇒ Allocation and deallocation can be in any order
⇒ Disadv : Heap management is overhead



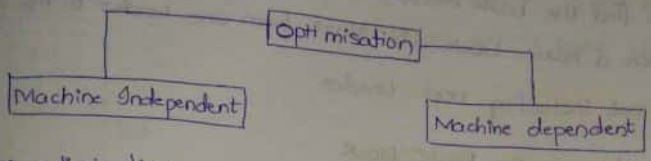
SUMMARY

Activations can have

1) permanent lifetime in case of static allocation
2) Nested lifetime in case of stack Allocation
3) Arbitrary lifetime incase of Heap Allocation.

CODE OPTIMISATION

INTRODUCTION TO CODE OPTIMISATION



1) Loop optimisations
   (a) code motion (or) Frequency reduction
   (b) Loop unrolling
   (c) Loop Jamming
2) folding
   - constant propagation
3) Redundancy Elimination
4) Strength Reduction

1) Register Allocation
2) Use of Addressing modes
3) peephole optimisation
   (a) Redundant = load / store
   (b) Strength Reduction
   (c) flow of control options
   (d) Use of M/c idioms

LOOP OPTIMISATION AND BASIC BLOCKS

→ To apply optimisations, we must first detect loops

→ For detecting loops we can use control flow Analysis (CFA) using program Flow Graph (PFG)

→ To find PFG, we need to find Basic blocks

A basic block is a sequence of 3-Address statements where control enters at the begining and leaves at the end without any jumps or halts

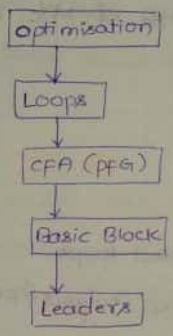
ALGORITHM TO FIND LEADERS

Finding the BASIC BLOCKS

→ In order to find the basic blocks, we need to find the leaders in the program then a basic block will start from one leader to the next Leader but not including next Leader.

Identifying Leaders in the Basic Block

1) I Statement is a leader
2) Statement that is target of conditional or unconditional Statement is a Leader → if() goto 200 (or) goto 300 → leaders
   ↳ Leaders
3) Statement that follows immediately a conditional or unconditional Statement is a Leader.



Example

```
fact(x)
{
    int f=1
    for(i=2;i<=x;i++)
    f=f*i;
    return f;
}
```

Now, the

*1) f = 1
2) i = 2
*3) if (i>x
*4) t_1 = f
5) t_2 = i
7) i = t_2;
8) goto (3)
*9) Goto c

TYPES

Frequenc
Moving
frequenc
frequenc
code m
Ex: whi
{
A
i
}
t =
whi
A

Now, the 3-Address code for the above problem will be

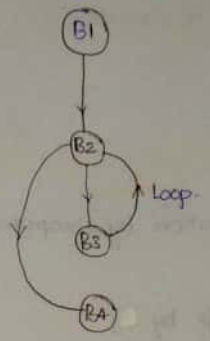
(1) f = 1 — B1
(2) i = 2

(3) If (i>x) goto 9 — B2

(4) t_1 = f * i ;
(5) t_2 = i+1 ;
(6) i = t_2 ;
(7) goto (3) — B3

(8) Goto calling program — B4

⇒ Incase you have 4 leaders we will get 4 Basic Blocks ⇒ If you have n' basic Leaders will get n' basic blocks



TYPES OF LOOP OPTIMIZATION

Frequency Reduction

Moving the code from high frequency region to low frequency Region is called code motion.

Ex: while (i<5000)
{
A = Sin(x)/cos(x) * i;
i++;
}

⇓

t = Sinx/cos x
while (i<5000)
A = t*i;

Loop unrolling

while (i<10)
{
x[i] = 0
i++
}

⇓

while (i<10)
{
x[i] = 0;
i++;
x[i] = 0;
i++;
}

Loop jamming

combines the bodies of two loops

for(i=0; i<10; i++)
for(j=0; j<10; j++)
x[i,j] = 0;
for(i=0; i<10; i++)
x[i,i] = 0

⇓

for(i=0, i<10; i++)
{
for(j=0; j<10; j++)
{
x[i,j] = 0;
} x[i,j] = 0;
}

MACHINE INDEPENDENT OPTIMISATION

Folding:

Replacing an expression that can be computed at compile time by its value

Ex: 2+3+C+B = 5+C+B

Redundancy Elimination (DAG)

A = B+C

D = 2+B+3+C

D = 2+3+A

Strength Reduction

Replacing a costly operation by cheaper one

Ex: B = A * 2

B = A << 1 [Left shift by 1]

Algebraic Simplification

A = A+0 }
x = x*1 } eliminate such statements

MACHINE DEPENDENT OPTIMISATION

1) Register Allocation
   - Local Allocation
   - Global Allocation

2) Use of Addressing modes

3) peephole optimisation

   a) Redundant Load and store elimination

x = y+z

```
Mov y,R0
Add z,R0
Mov R0,x
```

```
a = b+c | Mov b,R0
d = a+e | Add c,R0
        | Mov R0,a   X
        | Mov a,R0
        | Add e,R0
        | Mov R0,d
```

(b) Flow control optimisation

(41)

Avoid jumps on jumps

L1: Jump L2 → L4

L2: Jump L3

L3: Jump L4

Eliminate dead code

```
#define x 0
if (x)
{
    { dead code ×
}
```

d) Use of M/c idioms

i = i+1

```
MOV  R0, i
add  R0, 1
mov  i, R0
```

} increment 'i' [inc i]

⇒ Handle of the string is a substring that matches with RHS of production

⇒ RR conflicts occur in LALR(1) parser when merging of the states

⇒ SR conflicts doesnot occur in LALR(1) parser

⇒ If the attribute can be evaluated in Depth-first-order then the attribute is L-attributed.

Thank You

ADARSH CHETAN