



Intermediate Code Generation

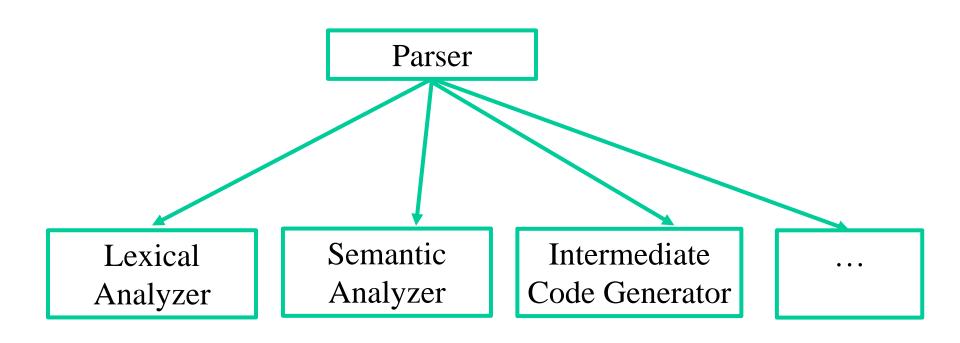
Lecture 8

Recommended Reading

P. Tremblay, J. Sorenson, Theory and Practice of Compiler Writing, McGraw Hill, Chapter 11.

(Not required if you attend this lecture and the next Exercise session)

Intermediate Code Generation



Intermediate Representation

- Translating source program into an "intermediate language."
 - -Simple
 - -CPU Independent,
 - —...yet, close in spirit to machine language.
- Three Address Code (quadruples)
- Two Address Code, etc.

Three Address Codes

- Statements of general form x:=y op z here represented as (op, y, z, x)
- No built-up arithmetic expressions are allowed.
- As a result, x:=y + z * w should be represented as:
 t₁:=z * w
 t₂:=y + t₁
 x:=t₂

 Three-address code is useful: related to machinelanguage/ simple/ optimizable.

Types of Three-Address Statements

if x relop y goto L (relop, x, y, L), or (jpf, A1, A2,), (jpt, A1, A2,), etc.

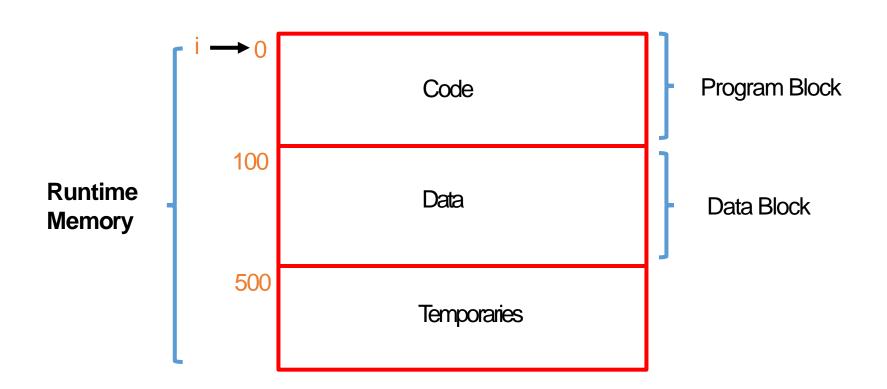
Different Addressing Modes

```
(+, 100, 101, 102)
  location 102 \leftarrow content of 100 + content of 101
(+, #100, 101, 102)
  location 102 \leftarrow constant 100 + content of 101
(+, @100, 101, 102)
  location 102 \leftarrow content of content of 100 + content of 101
```

Definitions

- Action Symbols (eg., #pid, #add, #mult, etc.): special symbols added to the grammar to signal the need for code generation
- Semantic Action (or, Semantic Routine): Each action symbol is associated with a sub-routine to perform
- Semantic Stack (here referred to by "ss"): a stack dedicated to the both semantic analyzer and intermediate code generator to store and use the required information
- Program Block (here referred to by "PB"): part of run time memory to be filled by the generated code

Run Time Memory Organization



Top-Down vs. Bottom-Up Generation

- Intermediate Code Generation can be performed in a top-down or bottom-up fashion (depending on the parsing direction)
- · We first explain the Top-Down Approach
- Then we explain the minor differences that exist between the two approaches

Top-Down Intermediate Code Generation

PRODUCTION Rules with action symbols:

```
1. S \rightarrow \#pid id := E \#assign
2. E \rightarrow T E'
3. E' \rightarrow \epsilon
4. E' \rightarrow + T #add E'
5. T \rightarrow F T'
6. T' \rightarrow \epsilon
7. T' \rightarrow * F \# \text{mult } T'
8. F \rightarrow (E)
9. F \rightarrow \#pid id
e.g., input: a := b + c * d
```

Code Generator Program

End codegen

```
Proc codegen(Action)

    Function <u>gettemp</u> returns a
new temporary variable that

      case (Action) of
                                                              we can use.
               #pid: begin
                        p \leftarrow findaddr(input);

    Function findaddr(input)
looks up the current input's
address from Symbol Table.

                        push(p)
               end
               #add | #mult : begin
                       t \leftarrow gettemp
                       PB[i] \leftarrow (+ \mid *, ss(top), ss(top-1), t);
                        i \leftarrow i + 1; pop(2); push(t)
               end
               #assign: begin
                        PB[i] \leftarrow (:=, ss(top), ss(top-1),);
                        i \leftarrow i + 1; pop(2)
               end
    end
```

Example

```
S \rightarrow \#pid id := E \#assign

E \rightarrow T E'

E' \rightarrow \epsilon \mid + T \#add E'

T \rightarrow F T'

T' \rightarrow \epsilon \mid * F \#mult T'

F \rightarrow (E)

F \rightarrow \#pid id
```

Parse Table

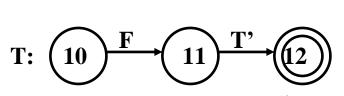
Non- terminal	INPUT SYMBOL						
terminal	id	+	*	()	\$:=
E	TE'			TE'			
E'		+ T E'			€	€	
T	FT'			FT'			
Т'		€	* F T'		€	€	
F	id			(E)			
S	id := E						

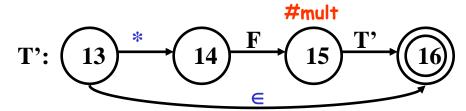
Example (Cont.)

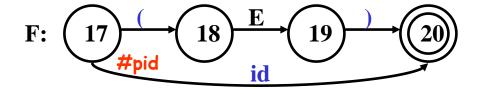
Parse Stack	Input	Operations
S \$	id1 := id2 + id3 * id4 \$	Pop
#pid id := E #assign \$	id1 := id2 + id3 * id4 \$	Call codegen(#pid), pop
id := E #assign \$	id1 := id2 + id3 * id4 \$	2 Matching ,2 pop
T E'#assign \$	id2 + id3 * id4 \$	Pop
F T'E'#assign \$	id2 + id3 * id4 \$	pop
#pid id T' E' #assign \$	id2 + id3 * id4 \$	Call codegen(#pid), pop
id T' E' #assign \$	id2 + id3 * id4 \$	Matching, pop
E'#assign \$	+ id3 * id4 \$	pop
+ T #add E' #assign \$	+ id3 * id4 \$	Matching, pop
F T' #add E' #assign \$	id3 * id4 \$	Pop
#pid id T' #add E' #assign \$	id3 * id4 \$	Call codegen(#pid), pop
id T' #add E' #assign \$	id3 * id4 \$	Matching, 2 pop
* F #mult T' #add E' #assign \$	* id4 \$	Matching, pop
#pid id #mult T' #add E' #assign \$	id4 \$	Call codegen(#pid), pop
id #mult T' #add E' #assign \$	id4 \$	Matching, pop
#mult T' #add E' #assign \$	\$	Call codegen(#mult), pop
T' #add E' #assign \$	\$	pop
#add E' #assign \$	\$	Call codegen(#add), pop
E'#assign \$	\$	pop
#assign \$	\$	Call codegen(#assign), pop
\$	\$	Finish!!

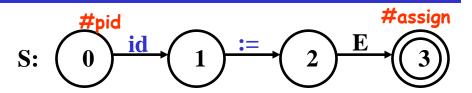
Transition Diagrams

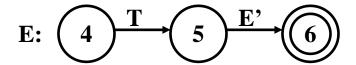
$$S \rightarrow \#pid id := E \#assign$$
 $E \rightarrow T E'$
 $E' \rightarrow \varepsilon \mid + T \#add E'$
 $T \rightarrow F T'$
 $T' \rightarrow \varepsilon \mid * F \#mult T'$
 $F \rightarrow (E)$
 $F \rightarrow \#pid id$

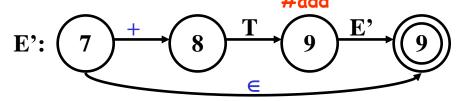






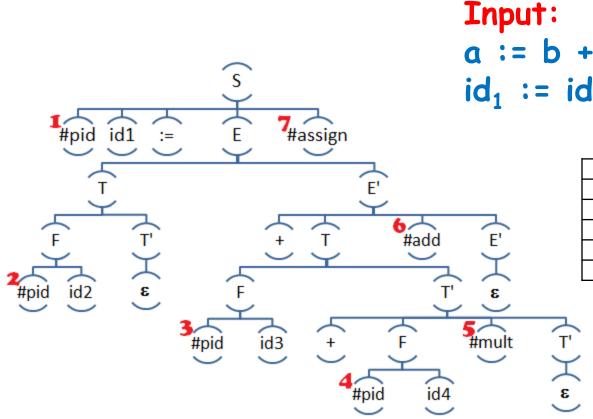


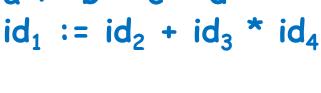




Example (Cont.)

 The order of Semantic Routines can be shown by the parse tree of the input sentence:

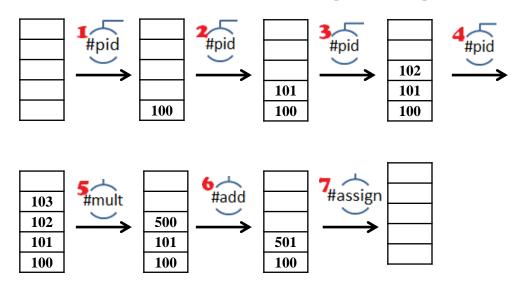




no	lexeme	address	
1	а	100	
2	b	101	
3	С	102	
4 d 103			
Symbol Table			

Example (Cont.)

• Semantic Stack (SS) status during code generation:

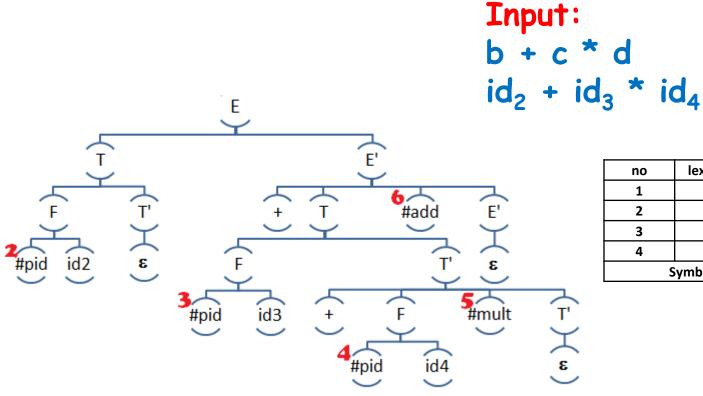


• Program Block (PB):

i	PB[i]	Semantic Action called
0	(*,103,102,500)	#mult
1	(+,500,101,501)	#add
2	(=,501,100,)	#assign

Example 2

· Suppose we only had the right hand side expression



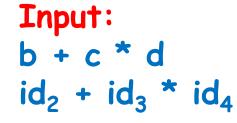
no	lexeme	address
1	а	100
2	b	101
3	С	102
4	d	103

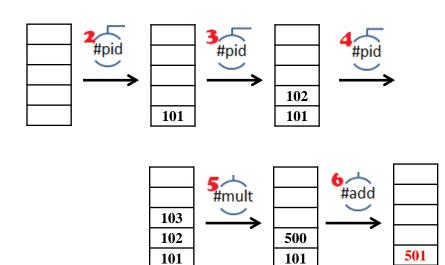
Symbol Table

Example 2 (Cont.)

• A temporary memory address will remain in SS:

i	PB[i]	Semantic Actions
0	(*,103, 102, 500)	#mult
	(+, 500, 101, <mark>501</mark>)	#add
2		





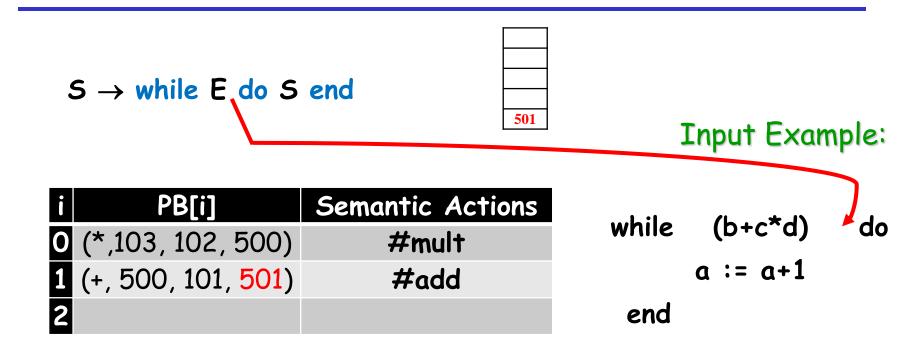
no	lexeme	address	
1	а	100	
2	b	101	
3	С	102	
4 d 103			
Symbol Table			

$$S \rightarrow \text{while E do S end}$$

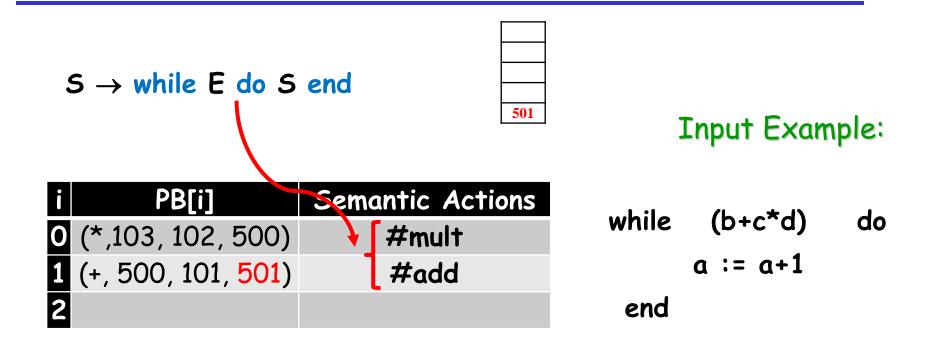
Input Example:

while
$$(b+c*d)$$
 do $a := a+1$ end

 We need to find the appropriate place for inserting Semantic Actions symbols.



 After parsing E, the code for while's condition has been generated; and the allocated temporary memory address will be on top of the Semantic Stack.



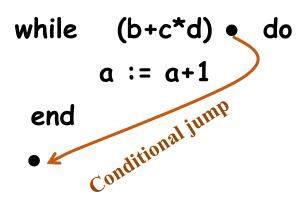
This is done by the semantic routines of non-terminal E

$S \rightarrow \text{while E do S end}$

 We need a conditional jump, based on the result of while's condition, to outside of the loop.

 But, we haven't yet compiled the loop body and thus don't know where the end of loop is!

Input Example:



#save
$$S \rightarrow \text{while E do S end}$$

Solution: BACKPATCHING!

#save: begin

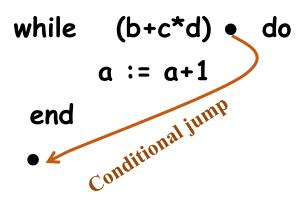
push(i)

i ← i + 1

end

 Reserve a place in the program block for the jump, and generate the code at the end of the loop

Input Example:



#label $S \rightarrow \text{while E do #save S end}$

- We also need an unconditional jump back to while condition.
- Target of the unconditional jump must be saved in SS
- This is done by a semantic action; let's call it #label

```
#label: begin

push(i)

end
```

Input Example: Unconditional jump while • (b+c*d) • do a := a+1• end Conditional jump

#while $S \rightarrow \text{while #label E do #save S end}$

At the end of while, the destination of conditional jump is known. So, the place saved by #save can be filled by #while.

An unconditional jump to the start of expression (saved by #label) is generated, too.

```
#while: begin

PB[ss(top)] \leftarrow (jpf, ss(top-1), i+1, );

PB[i] \leftarrow (jp, ss(top-2), , );

i \leftarrow i + 1; Pop(3)

end
```

Input Example: Unconditional jump while • (b+c*d) • do a := a+1• end Conditional jump Conditional jump

S -> while #label E do #save S #while end

i	PB[i]	Semantic Actions
<mark>→</mark> 0		
1		
2		
3		
4		
1 2 3 4 5		
6		

Input Example: Unconditional jump while • (b+c*d) • do a := a+1• end Conditional jump



S -> while #label E do #save S #while end

PB[i]	Semantic Actions
	PB[i]

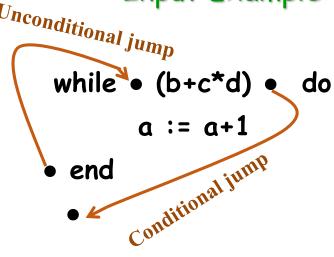
Input Example: Unconditional jump while • (b+c*d) • do a := a+1• end Conditional jump Conditional jump

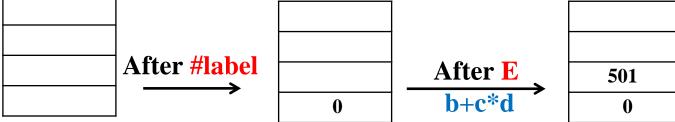


S -> while #label E do #save S #while end

i PB[i]	Semantic Actions	II.
O (*,103, 102, 500)	by E	<i>Unconditio</i>
1 (+, 500, 101, 501)	by E	whi
<mark>→</mark> 2		\ V
3		
4		• er
5		• 4
6		
_		

Input Example:





S -> while #label E do #save S #while end

i	PB[i]	Semantic Actions
0	(*,103, 102, 500)	by E
1	(+, 500, 101, 501)	by E
2		
→ 3		
4		
5		
6		

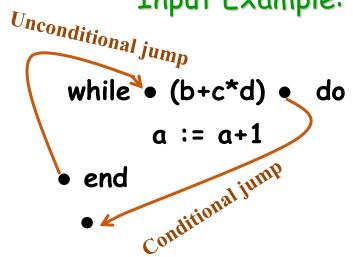
Input Example: Unconditional jump while • (b+c*d) • do a := a+1• end Conditional jump Conditional jump

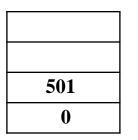


S -> while #label E do #save S #while end

i	PB[i]	Semantic Actions
0	(*,103, 102, 500)	by E
1	(+, 500, 101, 501)	by E
2		
3	(+, 100, #1, 503)	by S
4	(=, 503, 100,)	by S
5		
6		

Input Example:





After

2		
501		
0		

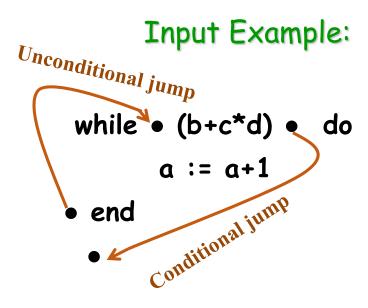
After S

2		
501		
0		

31

S -> while #label E do #save S #while end

i	PB[i]	Semantic Actions
0	(*,103, 102, 500)	by E
1	(+, 500, 101, 501)	by E
2	(jpf, 501, 6,)	#while
3	(+, 100, #1, 503)	by S
4	(=, 503, 100,)	by S
5	(jp, 0, ,)	#while
6		





```
S \rightarrow \text{if E then } S S'

S' \rightarrow \text{else } S

S' \rightarrow \epsilon
```

Input Example:

If
$$(a+b)$$
 then $a := a+1$ else $b := b+1$

$$S \rightarrow \text{if E \#save then } S S'$$

 $S' \rightarrow \text{else } S$
 $S' \rightarrow \epsilon$

Conditional jump: A place for jump should be saved by #save and to be later filled (by back patching).

Input Example:

If
$$(a+b)$$
 then $a := a+1$
else $b := b+1$
conditional jump

```
#save: begin

push(i)

i ← i + 1

end
```

```
S \rightarrow \text{if E \#save then } S S'

S' \rightarrow \text{else \#jpf\_save } S

S' \rightarrow \epsilon
```

When compiler reaches to else, the conditional jump can be generated by #jpf_save.

unconditional jump: A place for jump should be saved by #jpf_save and to be later filled (by back patching).

Input Example:

```
If (a+b) then a := a+1 else b := b+1 conditional jump Unconditional jump
```

```
#jpf_save: begin

PB[ss(top)] \leftarrow (jpf,ss(top-1), i+1, )

Pop(2), push(i), i \leftarrow i + 1;
```

end

```
S \rightarrow \text{if E \#save then } S S'

S' \rightarrow \text{else \#jpf\_save } S \# \text{jp}

S' \rightarrow \epsilon
```

When compiler is at the end of else statement, the unconditional jump can be generated by #jp.

Input Example:

```
If (a+b) then a := a+1 else b := b+1 conditional jump Unconditional jump
```

```
#jp: begin

PB[ss(top)] \leftarrow (jp, i, , )

Pop(1)

end
```

Conditional Statements (if-then-else)

```
S \rightarrow if E \#save then S S'
       S' \rightarrow else \# jpf save S \# jp
                                                                      Input Example:
       S' \rightarrow \# jpf
                                                      If (a+b) \bullet then a := a+1 \bullet
If there isn't an else statement
(S' \rightarrow \varepsilon, \text{ is used}),
                                                                  conditional jump
only a conditional jump is generated by
#jpf.
                                                     Compare with
                                                      #jpf_save
  #jpf: begin
            PB[ss(top)] \leftarrow (jpf, ss(top-1), i, )
            Pop(2)
                                                                                   37
```

end

Conditional Statements (if-then-else)

```
S \rightarrow \text{if E #save then } S S'

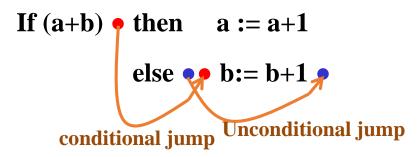
S' \rightarrow \text{else #jpf_save } S \# \text{jp}

S' \rightarrow \# \text{jpf}
```

Program Block:

i	PB[i]	Semantic Actions
0	(+, a, b, †1)	#add
1	(jpf, t1, ?=5,)	#save
2	(+, a, #1, †2)	#add
3	(:=, †2, a,)	#assign
4	(jp,?=7, ,)	#jpf_save
5	(+, b, #1, †3)	#add
6	(:=, t3, b,)	#assign
7		#jp

Input Example:

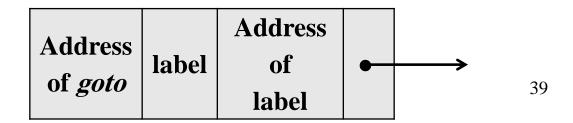


Goto Statements

```
S \rightarrow goto id;
S \rightarrow id: S
```

Difficulty: Unknown number forward goto statements. Semantic Stack is not enough!

- Implemented by a linked list. Each node of linked list has:
- - Address of goto (in PB)
 - Label name
 - Label address (in PB)
 - Pointer to next node

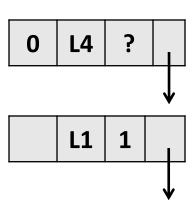


```
goto L4
   L1: Statement1
       goto L1;
       goto L2;
       goto L3;
  L3: Statement 2
       goto L1;
       goto L3;
       goto L2;
   L2: Statement 3
10 L4: Statement 4
```



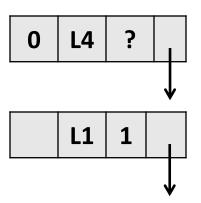
i	PB[i]
0	(jp,?,,)
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

```
goto L4
   L1: Statement1
       goto L1;
       goto L2;
       goto L3;
  L3: Statement 2
       goto L1;
       goto L3;
       goto L2;
   L2: Statement 3
10 L4: Statement 4
```



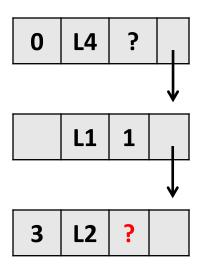
i	PB[i]
0	(jp,?,,)
1	statement 1
2	
3	
4	
5	
6	
7	
8	
9	
10	

```
goto L4
   L1: Statement1
       goto L1;
       goto L2;
       goto L3;
  L3: Statement 2
       goto L1;
       goto L3;
       goto L2;
   L2: Statement 3
10 L4: Statement 4
```



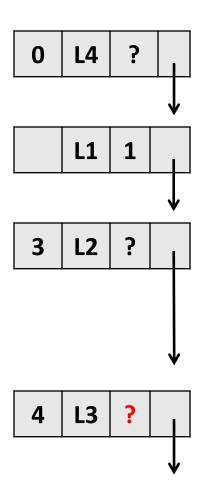
i	PB[i]
0	(jp,?,,)
1	statement 1
2	(jp , 1, ,)
3	
4	
5	
6	
7	
8	
9	
10	

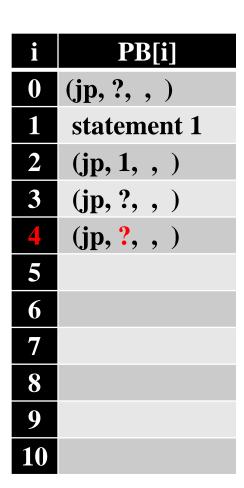
```
goto L4
   L1: Statement1
       goto L1;
       goto L2;
       goto L3;
  L3: Statement 2
       goto L1;
       goto L3;
       goto L2;
   L2: Statement 3
10 L4: Statement 4
```



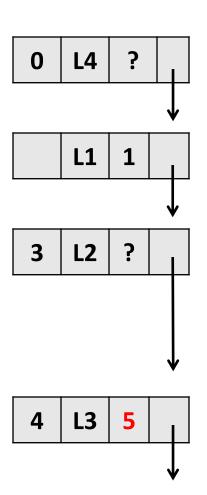
i	PB[i]
0	(jp,?,,)
1	statement 1
2 3	(jp, 1, ,)
3	(jp , ?, ,)
4	
5	
6	
7	
8	
9	
10	

```
goto L4
   L1: Statement1
       goto L1;
       goto L2;
       goto L3;
  L3: Statement 2
       goto L1;
       goto L3;
       goto L2;
   L2: Statement 3
10 L4: Statement 4
```



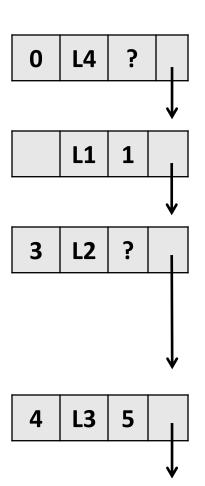


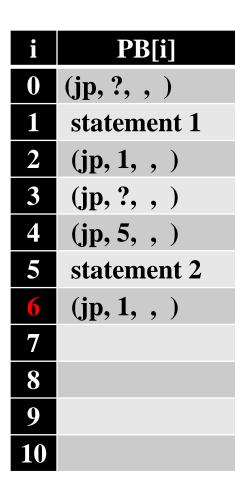
```
goto L4
   L1: Statement1
       goto L1;
       goto L2;
       goto L3;
  L3: Statement 2
       goto L1;
       goto L3;
       goto L2;
   L2: Statement 3
10 L4: Statement 4
```



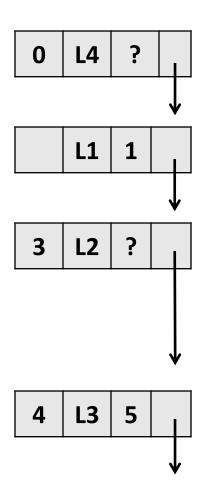
i	PB[i]
0	(jp,?,,)
1	statement 1
2	(jp, 1, ,)
3	(jp,?,,)
4	$(\mathbf{jp},5,\;,\;)$
5	statement 2
6	
7	
8	
9	
10	

```
goto L4
   L1: Statement1
       goto L1;
       goto L2;
       goto L3;
  L3: Statement 2
       goto L1;
       goto L3;
       goto L2;
   L2: Statement 3
10 L4: Statement 4
```



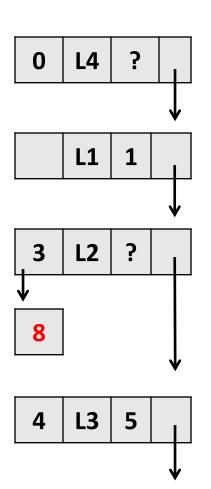


```
goto L4
   L1: Statement1
       goto L1;
       goto L2;
       goto L3;
  L3: Statement 2
       goto L1;
       goto L3;
       goto L2;
   L2: Statement 3
10 L4: Statement 4
```



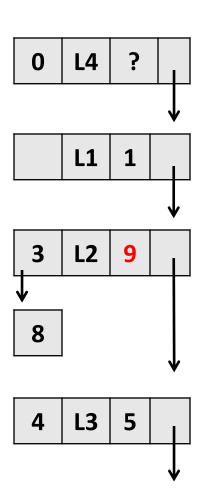
i	PB[i]
0	(jp,?,,)
1	statement 1
2	(jp , 1, ,)
3	(jp , ?, ,)
4	(jp, 5, ,)
5	statement 2
6	$(\mathbf{jp}, 1, ,)$
7	$(\mathbf{jp}, 5, ,)$
8	
9	
10	

```
goto L4
   L1: Statement1
       goto L1;
       goto L2;
       goto L3;
  L3: Statement 2
       goto L1;
       goto L3;
       goto L2;
   L2: Statement 3
10 L4: Statement 4
```



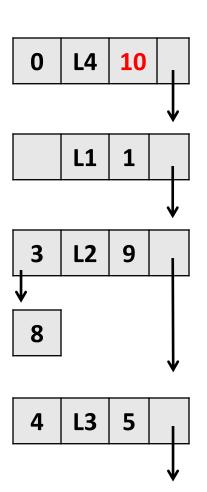
i	PB[i]
0	(jp,?,,)
1	statement 1
2	$(\mathbf{jp}, 1, ,)$
3	(jp , ?, ,)
4	(jp, 5, ,)
5	statement 2
6	(jp, 1, ,)
7	$(\mathbf{jp}, 5, ,)$
8	(jp , ?, ,)
9	
10	

```
goto L4
   L1: Statement1
       goto L1;
       goto L2;
       goto L3;
  L3: Statement 2
       goto L1;
       goto L3;
       goto L2;
   L2: Statement 3
10 L4: Statement 4
```



i	PB[i]
0	(jp,?,,)
1	statement 1
2	$(\mathbf{jp}, 1, ,)$
3	(jp , 9 , ,)
4	(jp, 5, ,)
5	statement 2
6	(jp, 1, ,)
7	(jp, 5, ,)
8	(jp , 9 , ,)
9	statement 3
10	

```
goto L4
   L1: Statement1
       goto L1;
       goto L2;
       goto L3;
  L3: Statement 2
       goto L1;
       goto L3;
       goto L2;
   L2: Statement 3
10 L4: Statement 4
```



i	PB[i]
0	(jp, 10, ,)
1	statement 1
2	$(\mathbf{jp}, 1, ,)$
3	(jp, 9, ,)
4	(jp, 5, ,)
5	statement 2
6	$(\mathbf{jp}, 1, ,)$
7	(jp, 5, ,)
8	(jp, 9, ,)
9	statement 3
10	statement 4

Bottom-Up Code Generation

```
1. S \rightarrow X \text{ id} := E

2. E \rightarrow T E'

3. E' \rightarrow \varepsilon

4. E' \rightarrow + T Y E'

5. T \rightarrow F T'

6. T' \rightarrow \varepsilon

7. T' \rightarrow * F Z T'

8. F \rightarrow (E)

9. F \rightarrow X \text{ id}

10. X \rightarrow \varepsilon

11. Y \rightarrow \varepsilon

new rules
```

12. $Z \rightarrow \varepsilon$

- Intermediate code generator is called by the parser, after each reduction
- Thus, only action symbols that are at the end of rules, are at a suitable positions
- Others action symbols must be replaced by new non-terminals producing only empty strings

Bottom-Up Code Generation

```
S \rightarrow X \text{ id} := E
2. E \rightarrow T E'
3. E' \rightarrow \varepsilon
4. E' \rightarrow + TYE'
5. T \rightarrow F T'
6. T' \rightarrow \epsilon
7. T' \rightarrow * F Z T'
8. F \rightarrow (E)
9. F \rightarrow X id
10. X \to \varepsilon
11. Y \to \varepsilon
12. Z \to \varepsilon new rules
```

Instead of action symbols, rule numbers are passed by the parser to the code generator

```
Proc codegen(Action)
    case (Action) of
            10 : begin
                        p \leftarrow findaddr(input);
                        push(p)
            end
            11 | 12 : begin
                        t \leftarrow gettemp
                        PB[i] \leftarrow (+ | *, ss(top), ss(top-1), t);
                        i \leftarrow i + 1; pop(2); push(t)
            end
          1: begin
            PB[i] \leftarrow (:=, ss(top), ss(top-1),);
            i \leftarrow i + 1; pop(2)
         end
      end
End codegen
```

Question?

$S \rightarrow repeat S until E end$

The above grammar defines repeat-until loops, where the loop body is executed at least once; we exit loop when its condition is true.

 Add the required action symbols and write the required semantic routines for such loops. Generate three address codes of the given example.

Input Example:

repeat

a := a-1

b := b+1

until (a-b) end

Question?

- The following grammar defines syntax of for loops.
- Add the required action symbols and write the required semantic routines for such loops. Generate three address codes of the given example.

$$S \rightarrow \text{for id} := E_1 \text{ to } E_2 \text{ A do S end}$$

 $A \rightarrow \text{by } E_3$
 $A \rightarrow \epsilon$

b+c : loop variable (j) initial value
a*b : loop variable (j) limit (constant)
c*d : loop variable (j) step (constant)

Input Example:

```
for j := b+c to a*b by c*a do
d := d+j
end
```

Question?

- The following grammar defines syntax of case statements, where at most one of case statements is to be executed.
- Can we generate intermediate code for these statements by just using a sematic stack to store the addresses that are required for back-patching?

Example:

```
case (c * d) of

a: a := a + 1;

b: b := b + 2;

c: c := c + 3;

otherwise: e := c*d

end
```

 $S \rightarrow case E of L end$ $L \rightarrow id: S B$ $B \rightarrow \epsilon \mid otherwise S \mid ; is : S B$