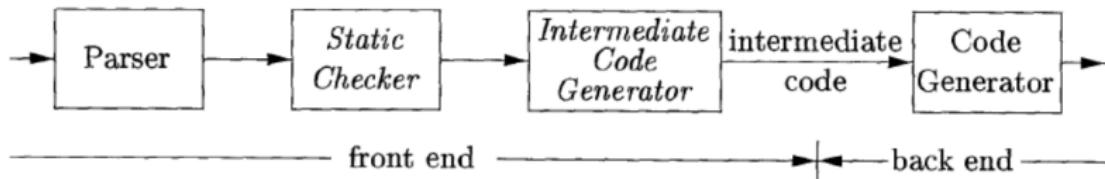


Intermediate-Code Generation

Intermediate-Code Generation



Three-Address Code

- In three-address code, there is at **most one operator** on the **right side of an instruction**
- Thus a **source-language expression like $x+y*z$** might be translated into the sequence of three-address instructions

$$\begin{aligned}t_1 &= y * z \\t_2 &= x + t_1\end{aligned}$$

Common three-address instructions

1. Assignment instructions of the form $x = y \ op \ z$, where op is a binary arithmetic or logical operation, and x , y , and z are addresses.
2. Assignments of the form $x = op \ y$, where op is a unary operation. Essential unary operations include unary minus, logical negation, shift operators, and conversion operators that, for example, convert an integer to a floating-point number.
3. *Copy instructions* of the form $x = y$, where x is assigned the value of y .
4. An unconditional jump `goto L`. The three-address instruction with label L is the next to be executed.
5. Conditional jumps of the form `if x goto L` and `ifFalse x goto L`. These instructions execute the instruction with label L next if x is true and false, respectively. Otherwise, the following three-address instruction in sequence is executed next, as usual.

6. Conditional jumps such as `if x relop y goto L`, which apply a relational operator (`<`, `==`, `>=`, etc.) to x and y , and execute the instruction with label L next if x stands in relation $relop$ to y . If not, the three-address instruction following `if x relop y goto L` is executed next, in sequence.
7. Procedure calls and returns are implemented using the following instructions: `param x` for parameters; `call p, n` and `y = call p, n` for procedure and function calls, respectively; and `return y`, where y , representing a returned value, is optional. Their typical use is as the sequence of three-address instructions

```
param x1
param x2
...
param xn
call p, n
```

generated as part of a call of the procedure $p(x_1, x_2, \dots, x_n)$. The integer n , indicating the number of actual parameters in “`call p, n`,”

8. Indexed copy instructions of the form $x = y[i]$ and $x[i] = y$. The instruction $x = y[i]$ sets x to the value in the location i memory units beyond location y . The instruction $x[i] = y$ sets the contents of the location i units beyond x to the value of y .

Common three-address instructions

```
do i = i+1; while (a[i] < v);
```

```
L:   t1 = i + 1
      i = t1
      t2 = i * 8
      t3 = a [ t2 ]
      if t3 < v goto L
```

Data structures for TAC

Quadruples

A *quadrule* (or just “*quad*”) has four fields, which we call *op*, *arg*₁, *arg*₂, and *result*. The *op* field contains an internal code for the operator. For instance, the three-address instruction $x = y + z$ is represented by placing `+` in *op*, *y* in *arg*₁, *z* in *arg*₂, and *x* in *result*. The following are some exceptions to this rule:

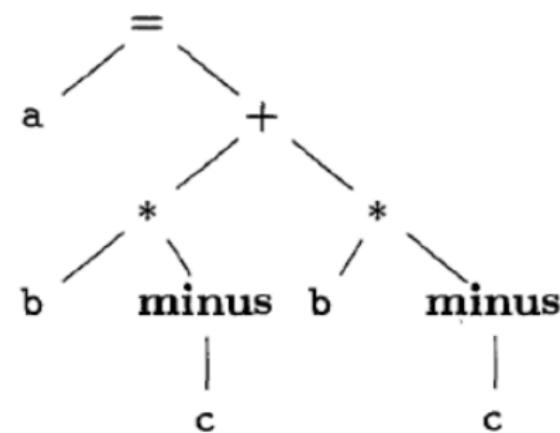
1. Instructions with unary operators like $x = \text{minus } y$ or $x = y$ do not use *arg*₂. Note that for a copy statement like $x = y$, *op* is `=`, while for most other operations, the assignment operator is implied.
2. Operators like `param` use neither *arg*₂ nor *result*.
3. Conditional and unconditional jumps put the target label in *result*.

Data structures for TAC

Quadruples

Three-address code for the assignment $a = b * - c + b * - c ;$

```
t1 = minus c  
t2 = b * t1  
t3 = minus c  
t4 = b * t3  
t5 = t2 + t4  
a = t5
```



(a) Three-address code

(a) Syntax tree

Data structures for TAC

Quadruples

Three-address code for the assignment $a = b * - c + b * - c ;$

```
t1 = minus c
t2 = b * t1
t3 = minus c
t4 = b * t3
t5 = t2 + t4
a = t5
```

	<i>op</i>	<i>arg₁</i>	<i>arg₂</i>	<i>result</i>
0	minus	c		t ₁
1	*	b	t ₁	t ₂
2	minus	c		t ₃
3	*	b	t ₃	t ₄
4	+	t ₂	t ₄	t ₅
5	=	t ₅		a
				...

(a) Three-address code

(b) Quadruples

Data structures for TAC

Triples

- A triple has **only three fields**, which we call **op**, **arg1**, and **arg2**
- Note that the **result field in Quad** is used primarily for temporary names.
- Using **triples**, we refer to the **result of an operation** x op y **by its position**, rather than by an explicit temporary name.
- Thus, instead of the **temporary t**, a triple representation would refer to **position (0)**.
- **Parenthesized numbers** represent pointers into the **triple structure** itself.

Data structures for TAC

Triples

Three-address code for the assignment $a = b * - c + b * - c ;$

```
t1 = minus c
t2 = b * t1
t3 = minus c
t4 = b * t3
t5 = t2 + t4
a = t5
```

	<i>op</i>	<i>arg₁</i>	<i>arg₂</i>
0	minus	c	
1	*	b	(0)
2	minus	c	
3	*	b	(2)
4	+	(1)	(3)
5	=	a	(4)
			...

(a) Three-address code

(b) Triples

Benefit of Quad over Triples

- A benefit of **quadruples** over triples can be seen in an **optimizing compiler**, where **instructions are often moved around**.
 - With quadruples, if we move an instruction that **computes a temporary t**, then the instructions that **use t** require **no change**.
- With **triples**, the **result** of an operation is referred to by **its position**
- So **moving an instruction** may require us to **change all references** to that result.

Indirect triples

- Consist of a **listing of pointers** to triples,
 - Rather than a listing of triples themselves.
- For example, use an **array instruction** to list **pointers to triples** in the desired order.

With indirect triples, an **optimizing compiler** can move an instruction by **reordering the instruction list**, without affecting the triples themselves.

<i>instruction</i>
35 (0)
36 (1)
37 (2)
38 (3)
39 (4)
40 (5)
...

	<i>op</i>	<i>arg₁</i>	<i>arg₂</i>
0	minus	c	
1	*	b	(0)
2	minus	c	
3	*	b	(2)
4	+	(1)	(3)
5	=	a	(4)
		...	

Static Single-Assignment Form

Two distinctive aspects distinguish SSA from three-address code.

(a) The first is that **all assignments in SSA are to variables with distinct names**; hence the term static single-assignment.

$$p = a + b$$

$$q = p - c$$

$$p = q * d$$

$$p = e - p$$

$$q = p + q$$

$$p_1 = a + b$$

$$q_1 = p_1 - c$$

$$p_2 = q_1 * d$$

$$p_3 = e - p_2$$

$$q_2 = p_3 + q_1$$

(a) Three-address code.

(b) Static single-assignment form.

Static Single-Assignment Form

Two distinctive aspects distinguish SSA from three-address code.

(a) The first is that **all assignments in SSA are to variables with distinct names**; hence the term static single-assignment.

(b)

The same variable may be defined in two different control-flow paths in a program. For example, the source program

```
if ( flag ) x = -1; else x = 1;  
y = x * a;
```

has two control-flow paths in which the variable x gets defined. If we use different names for x in the true part and the false part of the conditional statement, then which name should we use in the assignment $y = x * a$? Here is where the second distinctive aspect of SSA comes into play. SSA uses a notational convention called the ϕ -function to combine the two definitions of x :

```
if ( flag ) x1 = -1; else x2 = 1;  
x3 =  $\phi(x_1, x_2);$ 
```

Static Single-Assignment Form

Here, $\phi(x_1, x_2)$ has the value x_1 if the control flow passes through the true part of the conditional and the value x_2 if the control flow passes through the false part. That is to say, the ϕ -function returns the value of its argument that corresponds to the control-flow path that was taken to get to the assignment-statement containing the ϕ -function.

Major translation classes of Three address code generation

- (a) Declaration statements (+ handling data type and storage)
- (b) Expressions and statements
- (a) Control flow statements

Declaration statement

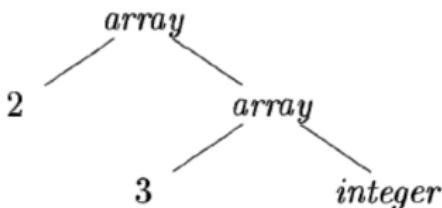
Representing data types: **Type Expressions**

Types have structure, which we shall represent using type expressions.

- A **type expression** is either a **basic type** (*boolean, char, integer, float, and void*)
or
- is formed by applying an operator called a **type constructor** to a type expression.
- A **type expression** can be formed by applying the **array type constructor** to a **number and a type expression**.

Declaration statement

- The array type int [2] [3] can be read as "array of 2 arrays of 3 integers each"
- Can be represented as a **type expression** array(2, array(3, integer)).
- This type is represented by the tree.



- The **operator array** takes **two parameters**, a number and a type.
 - Here the **type expression** can be formed by applying the **array type constructor** to a number and a type expression.

Declaration statement

Example SDD

PRODUCTION	SEMANTIC RULES
$T \rightarrow B\ C$	$T.t = C.t$ $C.b = B.t$
$B \rightarrow \text{int}$	$B.t = \text{integer}$
$B \rightarrow \text{float}$	$B.t = \text{float}$
$C \rightarrow [\text{ num}] C_1$	$C.t = \text{array}(\text{num}.val, C_1.t)$ $C_1.b = C.b$
$C \rightarrow \epsilon$	$C.t = C.b$



Type Expressions

- Nonterminal T generates either a **basic type** or an **array type**.
- Nonterminal B generates one of the basic types int and float.
- T generates a basic type when C derives ϵ .
- Otherwise, C generates array components consisting of a sequence of integers, each integer surrounded by brackets.

Declaration statement

Example SDD

PRODUCTION	SEMANTIC RULES
$T \rightarrow B\ C$	$T.t = C.t$ $C.b = B.t$
$B \rightarrow \text{int}$	$B.t = \text{integer}$
$B \rightarrow \text{float}$	$B.t = \text{float}$
$C \rightarrow [\text{num}] C_1$	$C.t = \text{array}(\text{num}.val, C_1.t)$ $C_1.b = C.b$
$C \rightarrow \epsilon$	$C.t = C.b$



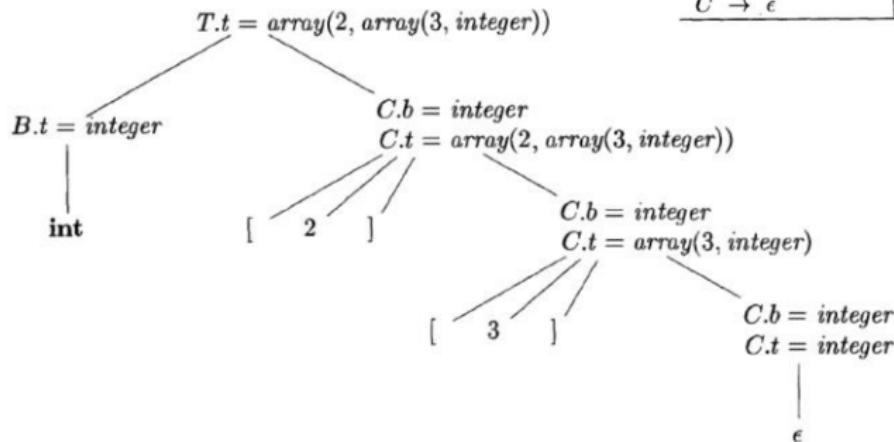
Type Expressions

- The nonterminals B and T have a synthesized attribute t representing a type.
- The nonterminal C has two attributes: an inherited attribute b and a synthesized attribute t .

Declaration statement

Example SDD

input string int [2][3]

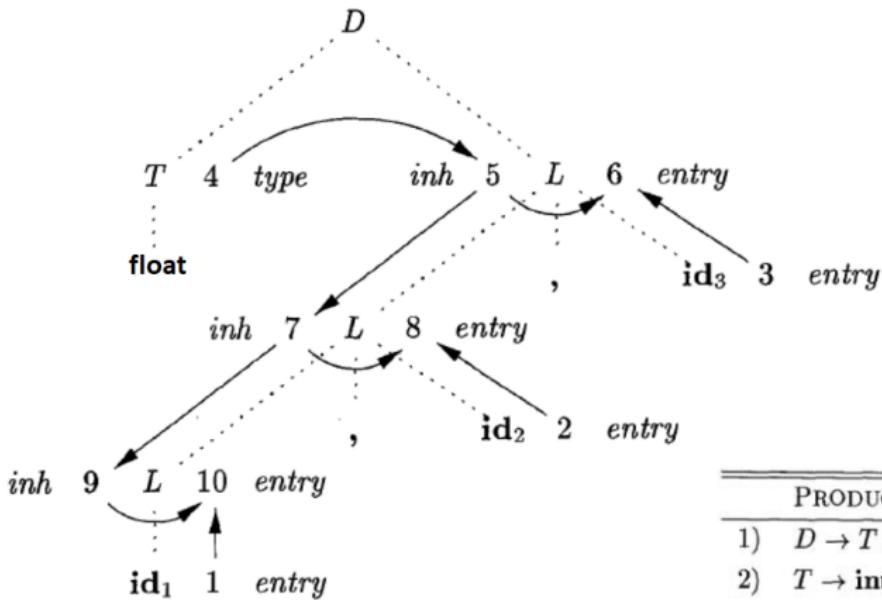


- The nonterminal C has two attributes: an inherited attribute b and a synthesized attribute t .
- The inherited b attributes pass a basic type down the tree, and the synthesized t attributes accumulate the result.

PRODUCTION	SEMANTIC RULES
$T \rightarrow B C$	$T.t = C.t$ $C.b = B.t$
$B \rightarrow \text{int}$	$B.t = \text{integer}$
$B \rightarrow \text{float}$	$B.t = \text{float}$
$C \rightarrow [\text{num}] C_1$	$C.t = \text{array}(\text{num}.val, C_1.t)$ $C_1.b = C.b$
$C \rightarrow \epsilon$	$C.t = C.b$

Declaration statement: Example SDD

float id₁ , id₂ , id₃



PRODUCTION	SEMANTIC RULES
1) $D \rightarrow T L$	$L.\text{inh} = T.\text{type}$
2) $T \rightarrow \text{int}$	$T.\text{type} = \text{integer}$
3) $T \rightarrow \text{float}$	$T.\text{type} = \text{float}$
4) $L \rightarrow \text{id}_1 \text{, id}_2$	$L_1.\text{inh} = L.\text{inh}$ $\text{addType}(\text{id}_1.\text{entry}, L.\text{inh})$
5) $L \rightarrow \text{id}_3$	$\text{addType}(\text{id}_3.\text{entry}, L.\text{inh})$

Symbol table

ST(global)

This is the Symbol Table for global symbols

Name	Type	Initial Value	Size	Offset	Nested Table
d	float	2.3	8	0	null
i	int	null	4	8	null
w	array(10, int)	null	40	12	null
a	int	4	4	52	null
p	ptr(int)	null	4	56	null
b	int	null	4	60	null
func	function	null	0	64	ptr-to-ST(func)
c	char	null	1	64	null

Find the storage for each variable

More on Declaration statement

Data type + Storage layout

$D \rightarrow T \text{ id } ; D \mid \epsilon$

$T \rightarrow B C$

$B \rightarrow \text{int} \mid \text{float}$

$C \rightarrow \epsilon \mid [\text{num}] C$

- Simplified grammar that declares just **one name at a time**;
- We already explored the declarations with lists of names

Storage layout:

- **Relative address** of all the variables
- From the **type of a name**, we can determine the **amount of storage** that will be needed for the name at run time.
- At **compile time**, we can use these amounts to **assign each name a relative address**.
- The **type** and **relative address** are saved in the **symbol-table entry** for the name.

More on Declaration statement

Data type + Storage layout

- The **width** of a **type** is the number of **storage units** needed for objects of that type (**offset**).
- A **basic type**, such as a character, integer, or float, requires an integral number of bytes.
- Arrays allocated in one **contiguous block of bytes**

$T \rightarrow B$ $\{ t = B.type; w = B.width; \}$

$B \rightarrow \text{int}$ $\{ B.type = \text{integer}; B.width = 4; \}$

$B \rightarrow \text{float}$ $\{ B.type = \text{float}; B.width = 8; \}$

$C \rightarrow \epsilon$ $\{ C.type = t; C.width = w; \}$

$C \rightarrow [\text{num}] C_1$ $\{ \text{array}(\text{num.value}, C_1.type); C.width = \text{num.value} \times C_1.width; \}$

Computes **data types** and their **widths** for basic and array types

Type Expressions

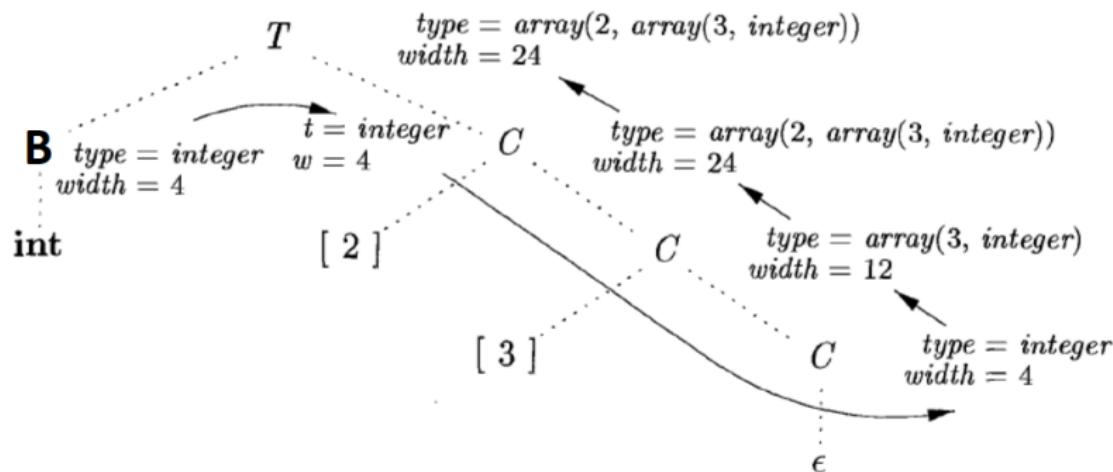
The **width of an array** is obtained by **multiplying** the **width of an element** by the **number of elements** in the array.

More on Declaration statement

Data type + Storage layout

int [2] [3]

$T \rightarrow B$	$\{ t = B.type; w = B.width; \}$
$B \rightarrow \text{int}$	$\{ B.type = \text{integer}; B.width = 4; \}$
$B \rightarrow \text{float}$	$\{ B.type = \text{float}; B.width = 8; \}$
$C \rightarrow \epsilon$	$\{ C.type = t; C.width = w; \}$
$C \rightarrow [\text{num}] C_1$	$\{ \text{array}(\text{num.value}, C_1.type); C_1.width = \text{num.value} \times C_1.width; \}$



Relative address

Name	Data type
d	float
i	int
w	array(10, int)

Relative address

0
8
12
..

Sequences of Declarations

```
int x;  
float y;
```

- Relative address: **offset**
- Keeps track of the **next available relative address**

$P \rightarrow \{ offset = 0; \} D$

$D \rightarrow T \text{id} ; \quad \{ top.put(\text{id}.lexeme, T.type, offset);$
 $\qquad \qquad \qquad offset = offset + T.width; \}$

D_1
 $D \rightarrow \epsilon$

- The translation scheme deals with a **sequence of declarations** of the form $T \text{id}$, where T generates a data type
- Before the first declaration is considered, **offset is set to 0**.
- As **each new name x** is seen, x is entered into the **symbol table** with its **relative address = current value of offset**,
 - which is **then incremented** by the width of the type of x .

Sequences of Declarations

- Relative address: offset
- Keeps track of the next available relative address

$$P \rightarrow \{ \text{offset} = 0; \} \ D$$
$$D \rightarrow T \text{id} ; \quad \{ \text{top.put(id.lexeme, T.type, offset);}$$
$$\quad \quad \quad \text{offset} = \text{offset} + T.\text{width}; \}$$
$$D_1$$
$$D \rightarrow \epsilon$$

The semantic action within the production $D \rightarrow T \text{id} ; D_1$ creates a symbol-table entry by executing $\text{top.put(id.lexeme, T.type, offset)}$. Here top denotes the current symbol table. The method top.put creates a symbol-table entry for id.lexeme , with type $T.\text{type}$ and relative address offset in its data area.

Major translation classes of Three address code generation

(a) Declaration statements (+ handling data type and storage)

(b) Expressions and statements

(a) Control flow statements

Translation of Expressions statement $a = b + - c$

Three-address code for an assignment statement

PRODUCTION	SEMANTIC RULES
$S \rightarrow id = E ;$	$S.code = E.code \parallel$ $gen(top.get(id.lexeme) '=' E.addr)$
$E \rightarrow E_1 + E_2$	$E.addr = \text{new Temp}()$ $E.code = E_1.code \parallel E_2.code \parallel$ $gen(E.addr '=' E_1.addr '+' E_2.addr)$
$ - E_1$	$E.addr = \text{new Temp}()$ $E.code = E_1.code \parallel$ $gen(E.addr '=' 'minus' E_1.addr)$
$ (E_1)$	$E.addr = E_1.addr$ $E.code = E_1.code$
$ id$	$E.addr = top.get(id.lexeme)$ $E.code = ''$

Attribute **code** for S

attributes **addr** and **code** for an expression E.

Attributes S.code and E.code denote the three-address code for S and E, respectively.

Attribute E.addr denotes the address that will hold the value of E.

Translation of Expressions

Three-address code for an assignment statement

PRODUCTION	SEMANTIC RULES
$S \rightarrow \text{id} = E ;$	$S.\text{code} = E.\text{code} $ $\text{gen}(\text{top.get(id.lexeme)} ' =' E.\text{addr})$
$E \rightarrow E_1 + E_2$	$E.\text{addr} = \text{new Temp}()$ $E.\text{code} = E_1.\text{code} E_2.\text{code} $ $\text{gen}(E.\text{addr} ' =' E_1.\text{addr} '+' E_2.\text{addr})$
$ - E_1$	$E.\text{addr} = \text{new Temp}()$ $E.\text{code} = E_1.\text{code} $ $\text{gen}(E.\text{addr} ' =' '\text{minus}' E_1.\text{addr})$
$ (E_1)$	$E.\text{addr} = E_1.\text{addr}$ $E.\text{code} = E_1.\text{code}$
$ \text{id}$	$E.\text{addr} = \text{top.get(id.lexeme)}$ $E.\text{code} = ''$

When an expression is a single identifier, say `x`, then `x` itself holds the value of the expression.

The semantic rules for this production define `E.addr` to point to the symbol-table entry

Translation of Expressions

Three-address code for an assignment statement

PRODUCTION	SEMANTIC RULES
$S \rightarrow id = E ;$	$S.code = E.code $ $gen(top.get(id.lexeme) '=' E.addr)$
$E \rightarrow E_1 + E_2$	$E.addr = new Temp()$ $E.code = E_1.code E_2.code $ $gen(E.addr '=' E_1.addr '+' E_2.addr)$



The semantic rules for $E \rightarrow E_1 + E_2$, generate code to compute the value of E from the values of E_1 and E_2 . Values are computed into newly generated temporary names. If E_1 is computed into $E_1.addr$ and E_2 into $E_2.addr$, then $E_1 + E_2$ translates into $t = E_1.addr + E_2.addr$, where t is a new temporary name. $E.addr$ is set to t . A sequence of distinct temporary names t_1, t_2, \dots is created by successively executing **new Temp()**.

Translation of Expressions

Three-address code for an assignment statement

PRODUCTION	SEMANTIC RULES
$S \rightarrow \text{id} = E ;$	$S.\text{code} = E.\text{code} $ $\text{gen}(\text{top.get(id.lexeme)} ' =' E.\text{addr})$
$E \rightarrow E_1 + E_2$	$E.\text{addr} = \text{new Temp}()$ $E.\text{code} = E_1.\text{code} E_2.\text{code} $ $\text{gen}(E.\text{addr} ' =' E_1.\text{addr} '+' E_2.\text{addr})$

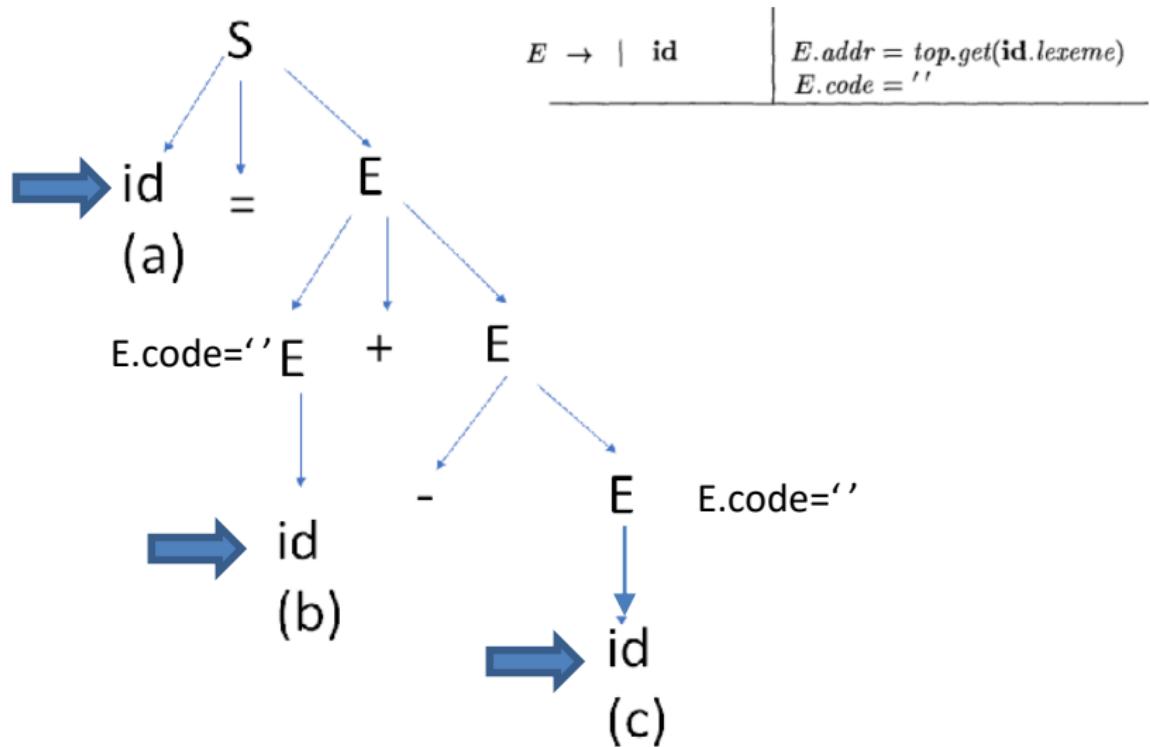
Finally, the production $S \rightarrow \text{id} = E ;$ generates instructions that assign the value of expression E to the identifier id . The semantic rule for this production uses function top.get to determine the address of the identifier represented by id , as in the rules for $E \rightarrow \text{id}$. $S.\text{code}$ consists of the instructions to compute the value of E into an address given by $E.\text{addr}$, followed by an assignment to the address $\text{top.get(id.lexeme)}$ for this instance of id .

Translation of Expressions

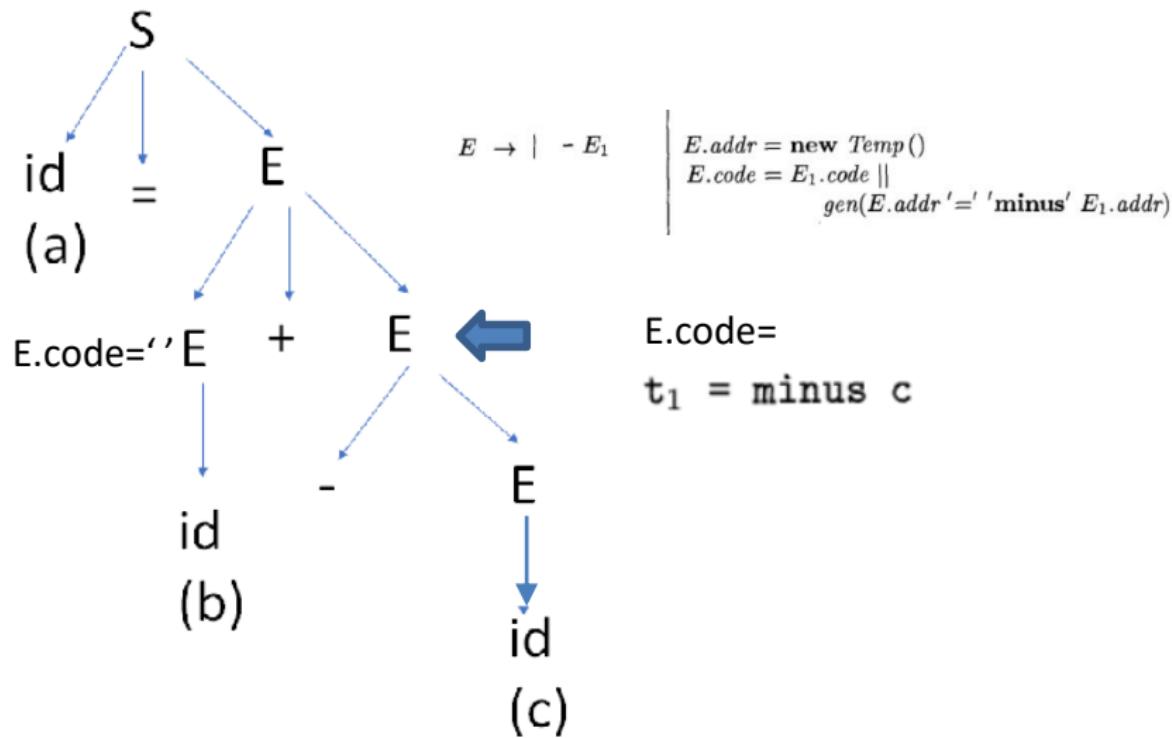
Three-address code for an assignment statement

PRODUCTION	SEMANTIC RULES
$S \rightarrow \text{id} = E ;$	$S.\text{code} = E.\text{code} $ $\text{gen}(\text{top.get(id.lexeme)} ' = ' E.\text{addr})$
$E \rightarrow E_1 + E_2$	$E.\text{addr} = \text{new Temp}()$ $E.\text{code} = E_1.\text{code} E_2.\text{code} $ $\text{gen}(E.\text{addr} ' = ' E_1.\text{addr} ' + ' E_2.\text{addr})$
$ - E_1$	$E.\text{addr} = \text{new Temp}()$ $E.\text{code} = E_1.\text{code} $ $\text{gen}(E.\text{addr} ' = ' \text{minus}' E_1.\text{addr})$
$ (E_1)$	$E.\text{addr} = E_1.\text{addr}$ $E.\text{code} = E_1.\text{code}$
$ \text{id}$	$E.\text{addr} = \text{top.get(id.lexeme)}$ $E.\text{code} = ''$

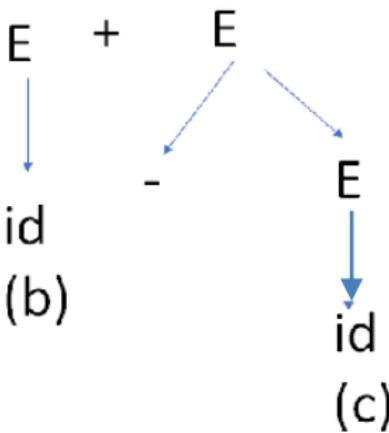
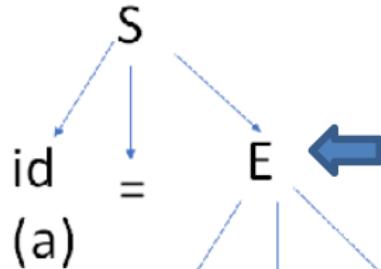
statement $a = b + - c$



statement **a = b + - c**



statement **a = b + - c**



$$E \rightarrow E_1 + E_2$$

$E.\text{addr} = \text{new Temp}()$

$E.\text{code} = E_1.\text{code} || E_2.\text{code} ||$

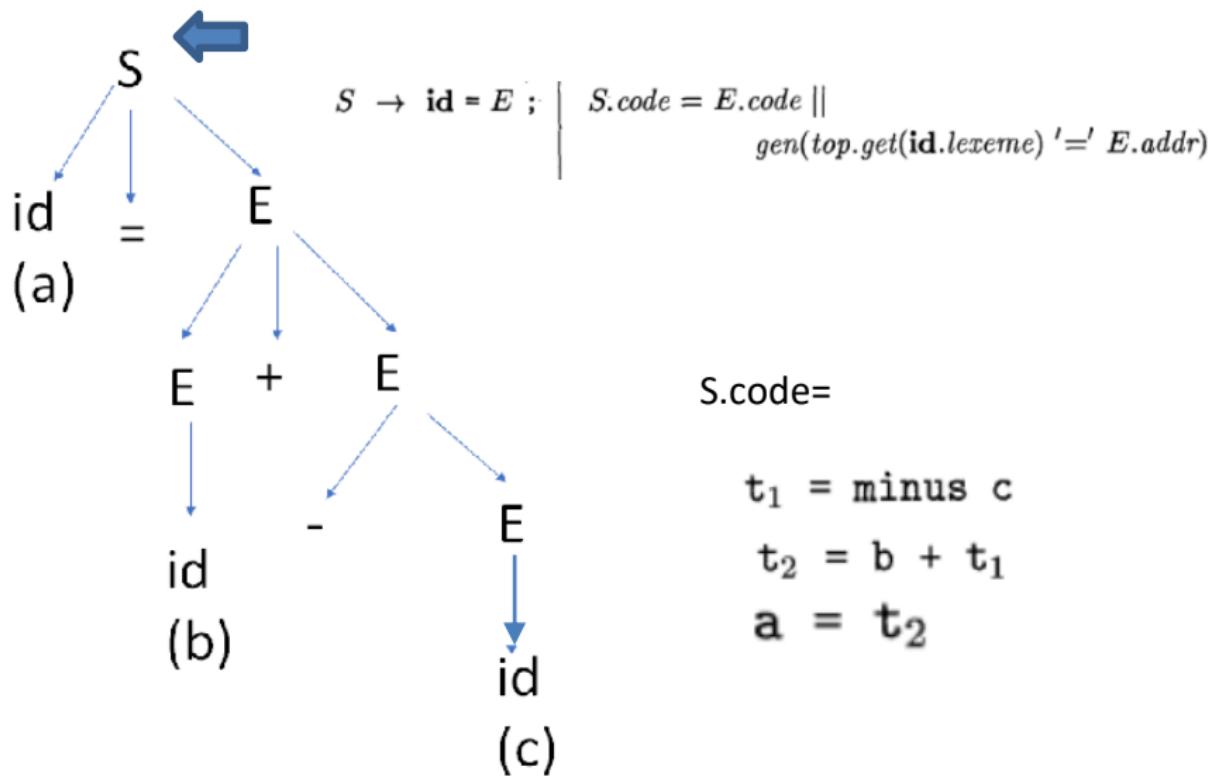
$\text{gen}(E.\text{addr}' = E_1.\text{addr}' + E_2.\text{addr})$

$E.\text{code} =$

$t_1 = \text{minus } c$

$t_2 = b + t_1$

statement $a = b + -c$



statement $a = b + - c$

Translation of Expressions

Three-address code for an assignment statement

PRODUCTION	SEMANTIC RULES
$S \rightarrow \text{id} = E ;$	$S.\text{code} = E.\text{code} $ $\text{gen}(\text{top.get(id.lexeme)} ' = ' E.\text{addr})$
$E \rightarrow E_1 + E_2$	$E.\text{addr} = \text{new Temp}()$ $E.\text{code} = E_1.\text{code} E_2.\text{code} $ $\text{gen}(E.\text{addr} ' = ' E_1.\text{addr} ' + ' E_2.\text{addr})$
$ - E_1$	$E.\text{addr} = \text{new Temp}()$ $E.\text{code} = E_1.\text{code} $ $\text{gen}(E.\text{addr} ' = ' \text{minus}' E_1.\text{addr})$
$ (E_1)$	$E.\text{addr} = E_1.\text{addr}$ $E.\text{code} = E_1.\text{code}$
$ \text{id}$	$E.\text{addr} = \text{top.get(id.lexeme)}$ $E.\text{code} = ''$

statement $a = b + - c$

$t_1 = \text{minus } c$
 $t_2 = b + t_1$
 $a = t_2$

Incremental Translation

- So far, **E.Code** attributes were long strings
 - Generated incrementally
- In incremental translation, generate only the **new three-address instructions**
- **Past sequence** may either be **retained in memory** for further processing, or it may be **output incrementally**.
- In the incremental approach, gen() not only constructs a three-address instruction,
 - it **appends** the instruction to the sequence of instructions generated so far.

Incremental Translation

$S \rightarrow \text{id} = E ; \{ \text{gen}(\text{top.get(id.lexeme)} '==' E.\text{addr}); \}$

$E \rightarrow E_1 + E_2 \quad \{ E.\text{addr} = \text{new Temp}();$
 $\quad \quad \quad \text{gen}(E.\text{addr} '==' E_1.\text{addr} '+' E_2.\text{addr}); \}$

| $- E_1 \quad \{ E.\text{addr} = \text{new Temp}();$
 $\quad \quad \quad \text{gen}(E.\text{addr} '==' '\text{minus}' E_1.\text{addr}); \}$

| $(E_1) \quad \{ E.\text{addr} = E_1.\text{addr}; \}$

| $\text{id} \quad \{ E.\text{addr} = \text{top.get(id.lexeme)}; \}$

- This **translation scheme generates the same code** as the previous syntax directed definition.
- With the incremental approach, the **E.code attribute is not used**,
 - Since there is a **single sequence of instructions** that is created by **successive calls to gen()**.

Translation of Array Expressions

```
int num[3][4] = {  
    {1, 2, 3, 4},  
    {5, 6, 7, 8},  
    {9, 10, 11, 12}  
};
```

row-wise memory allocation

	←→ row 0 →→				←→ row 1 →→				←→ row 2 →→			
value	1	2	3	4	5	6	7	8	9	10	11	12
address	1000	1002	1004	1006	1008	1010	1012	1014	1016	1018	1020	1022

↑
first element of the array num

Let w_1 be the length of a row (# of columns) and w_2 be the size of an element in a row.

The relative address of $A[i_1][i_2]$ can be calculated by the formula

$$\text{base} + i_1 \times w_1 + i_2 \times w_2$$

Translation of Array Expressions

$S \rightarrow id = E ; \quad \{ gen(top.get(id.lexeme) '==' E.addr); \}$

L.addr denotes a temporary variable containing the relative addresses (offset) of the array elements (array index)

| $L = E ; \quad \{ gen(L.addr.base '[' L.addr ']' '==' E.addr); \}$

$E \rightarrow E_1 + E_2 \quad \{ E.addr = new Temp();$
 $\quad gen(E.addr '==' E_1.addr +' E_2.addr); \}$

| $id \quad \{ E.addr = top.get(id.lexeme); \}$

$$base + i_1 \times w_1 + i_2 \times w_2$$

| $L \quad \{ E.addr = new Temp();$
 $\quad gen(E.addr '==' L.array.base '[' L.addr ']'); \}$

$L \rightarrow id [E] \quad \{ L.array = top.get(id.lexeme);$
 $\quad L.type = L.array.type.elem;$
 $\quad L.addr = new Temp();$
 $\quad gen(L.addr '==' E.addr '*' L.type.width); \}$

| $L_1 [E] \quad \{ L.array = L_1.array;$
 $\quad L.type = L_1.type.elem;$
 $\quad t = new Temp();$
 $\quad L.addr = new Temp();$
 $\quad gen(t '==' E.addr '*' L.type.width); \}$
 $\quad gen(L.addr '==' L_1.addr +' t); \}$

Translation of Array Expressions

$S \rightarrow id = E ; \quad \{ gen(top.get(id.lexeme) '==' E.addr); \}$

| $L = E ; \quad \{ gen(L.addr.base '[' L.addr ']' '==' E.addr); \}$

$E \rightarrow E_1 + E_2 \quad \{ E.addr = new Temp();$
 $\quad gen(E.addr '==' E_1.addr +' E_2.addr); \}$

| $id \quad \{ E.addr = top.get(id.lexeme); \}$

| $L \quad \{ E.addr = new Temp();$
 $\quad gen(E.addr '==' L.array.base '[' L.addr ']'); \}$

$L \rightarrow id [E] \quad \{ L.array = top.get(id.lexeme);$
 $\quad L.type = L.array.type.elem;$
 $\quad L.addr = new Temp();$
 $\quad gen(L.addr '==' E.addr '*' L.type.width); \}$

| $L_1 [E] \quad \{ L.array = L_1.array;$
 $\quad L.type = L_1.type.elem;$
 $\quad t = new Temp();$
 $\quad L.addr = new Temp();$
 $\quad gen(t '==' E.addr '*' L.type.width); \}$
 $\quad gen(L.addr '==' L_1.addr +' t); \}$

- **L.array** is a pointer to the symbol-table entry for the array name.
- **L.array.base** indicates base address of the array -- array name
- **L.type** is the type **t** of the subarray generated by **L**.
- For any **type t**, we assume that its width is given by **t.width**.
- For any array type **t**, **t.elem** gives the array element type.

Translation of Array Expressions

$S \rightarrow id = E ; \quad \{ gen(top.get(id.lexeme) '==' E.addr); \}$ Standard

| $L = E ; \quad \{ gen(L.addr.base '[' L.addr ']' '==' E.addr); \}$

$E \rightarrow E_1 + E_2 \quad \{ E.addr = new Temp(); \quad gen(E.addr '==' E_1.addr +' E_2.addr); \}$ Standard

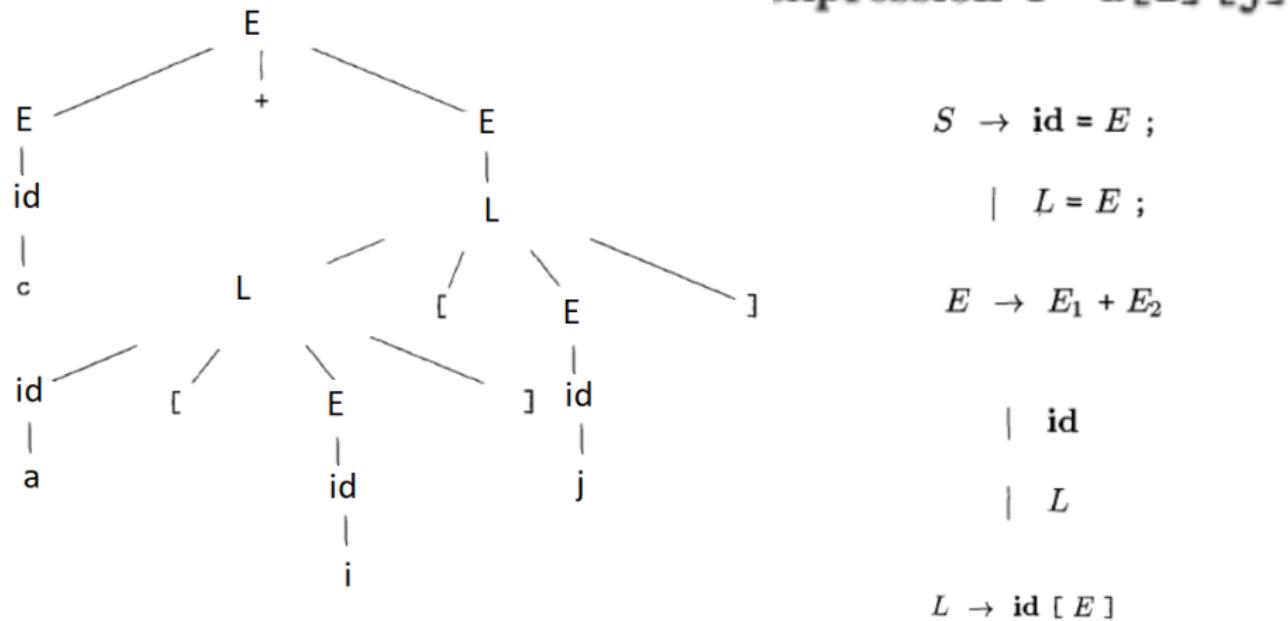
| $id \quad \{ E.addr = top.get(id.lexeme); \}$ Standard

| $L \quad \{ E.addr = new Temp(); \quad gen(E.addr '==' L.array.base '[' L.addr ']'); \}$

$L \rightarrow id [E] \quad \{ L.array = top.get(id.lexeme); \quad L.type = L.array.type.elem; \quad L.addr = new Temp(); \quad gen(L.addr '==' E.addr '*' L.type.width); \}$

| $L_1 [E] \quad \{ L.array = L_1.array; \quad L.type = L_1.type.elem; \quad t = new Temp(); \quad L.addr = new Temp(); \quad gen(t '==' E.addr '*' L.type.width); \quad gen(L.addr '==' L_1.addr +' t); \}$

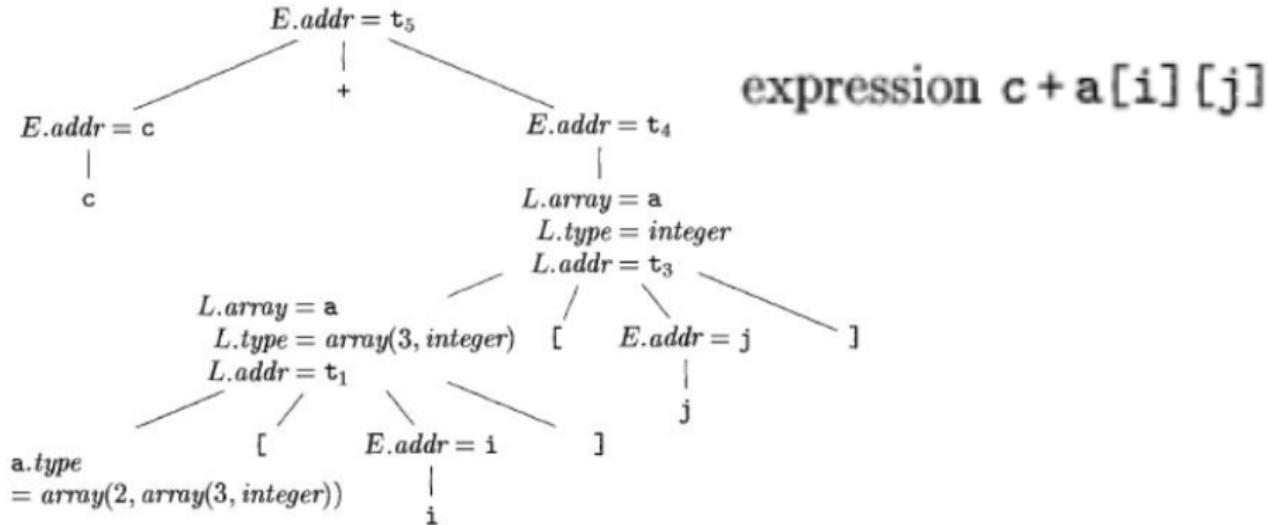
Translation of Array Expressions



Parse tree

| $L_1 [E]$

Translation of Array Expressions



```

 $L \rightarrow \text{id} [ E ] \quad \{ L.array = \text{top.get(id.lexeme)};$ 
 $\qquad L.type = L.array.type.elem;$ 
 $\qquad L.addr = \text{new Temp}();$ 
 $\qquad \text{gen}(L.addr \leftarrow E.addr *' L.type.width); \}$ 

```

$$E \rightarrow L \quad \{ E.\text{addr} = \text{new Temp}(); \\ \quad \quad \quad \text{gen}(E.\text{addr}' = L.\text{array.base}'[L.\text{addr}']); \}$$

| $L_1(E)$

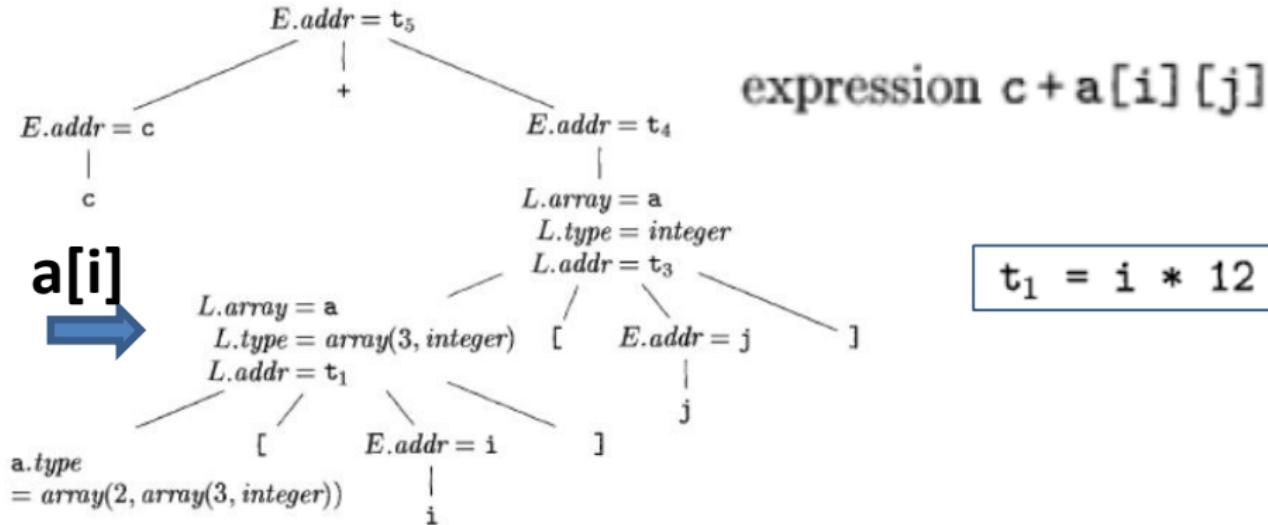
```

{ L.array = L1.array;
  L.type = L1.type.elem;
  t = new Temp();
  L.addr = new Temp();
  gen(t '=' E.addr '*' L.type.width); }
  gen(L.addr '=' L1.addr '+' t); }

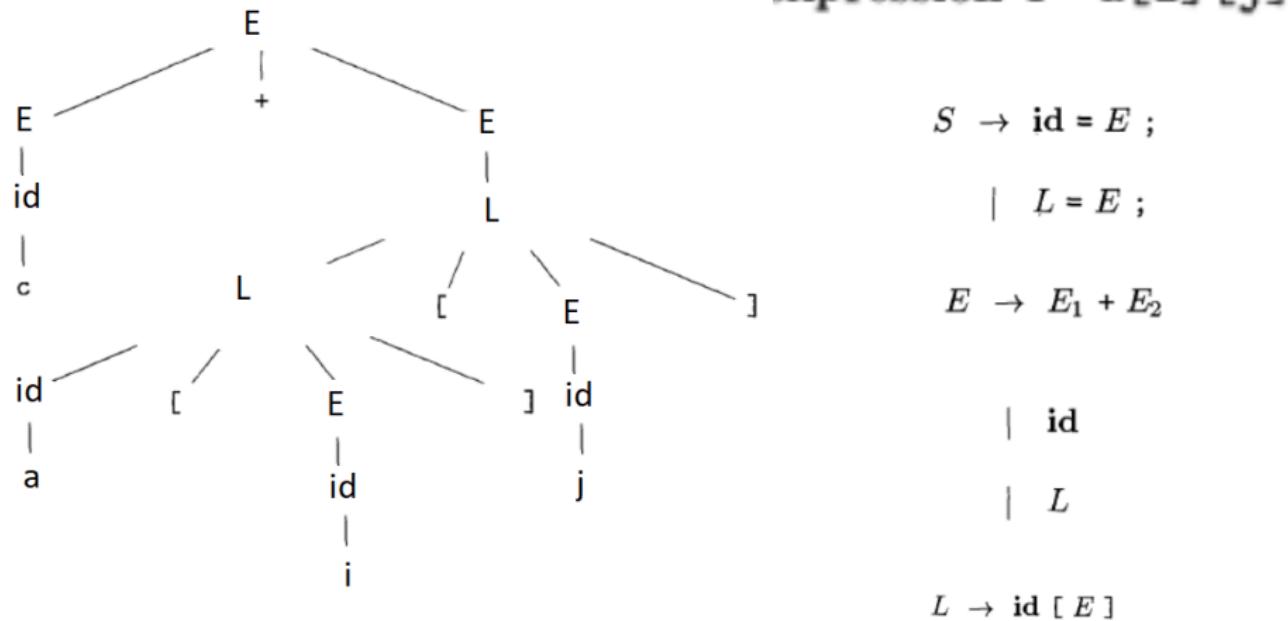
```

$E \rightarrow \text{id} \quad \{ E.\text{addr} = \text{top.get(id.lexeme)}; \}$

Translation of Array Expressions



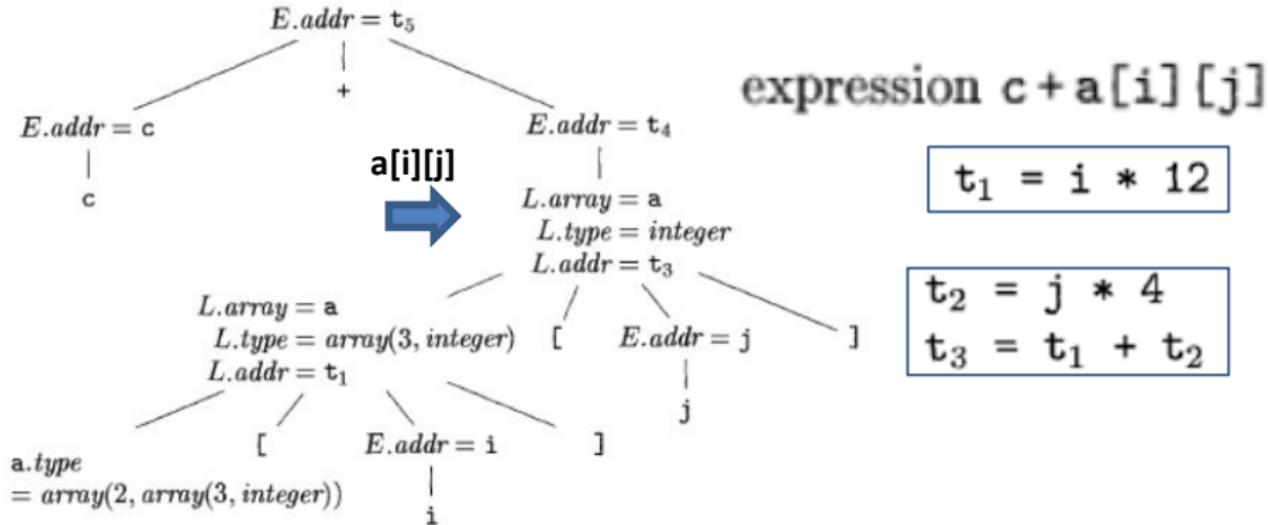
Translation of Array Expressions



Parse tree

$| \quad L_1 [E]$

Translation of Array Expressions



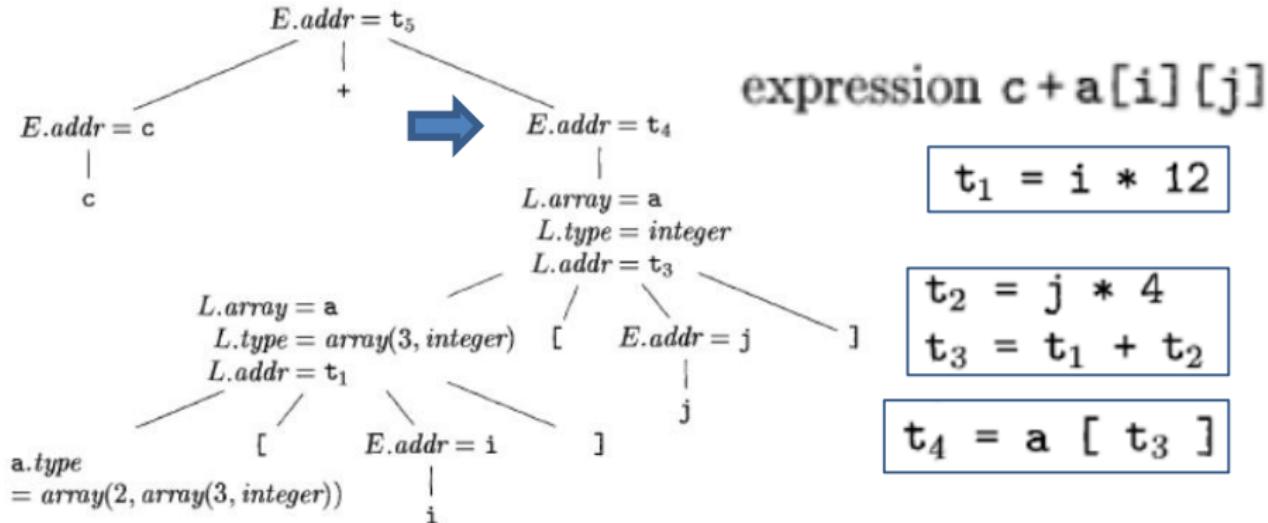
$L \rightarrow \text{id} [E] \quad \{ L.array = \text{top.get(id.lexeme)};$
 $L.type = L.array.type.elem;$
 $L.addr = \text{new Temp}();$
 $\text{gen}(L.addr '=' E.addr '*' L.type.width); \}$

$E \rightarrow \text{id}$

$\{ E.addr = \text{top.get(id.lexeme)}; \}$

$| \quad L_1 [E] \quad \{ L.array = L_1.array;$
 $L.type = L_1.type.elem;$
 $t = \text{new Temp}();$
 $L.addr = \text{new Temp}();$
 $\text{gen}(t '=' E.addr '*' L.type.width); \}$
 $\text{gen}(L.addr '=' L_1.addr '+ t); \}$

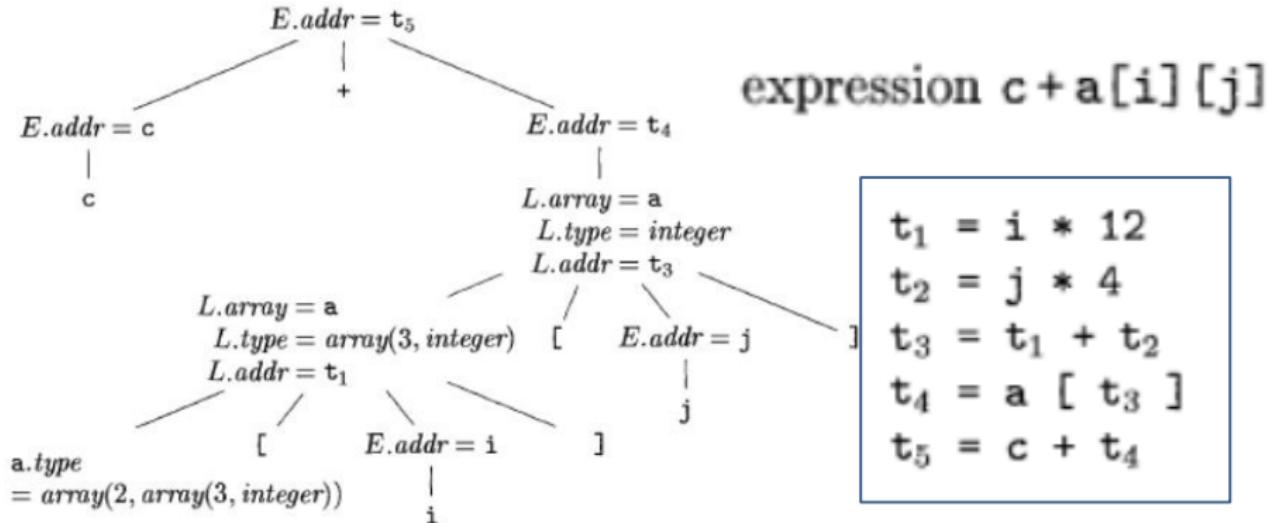
Translation of Array Expressions



$E \rightarrow L$

```
{ E.addr = new Temp();  
gen(E.addr '=' L.array.base '[' L.addr ']'); }
```

Translation of Array Expressions



```

 $E \rightarrow L \quad \{ E.addr = \text{new Temp}();$   

 $\quad \quad \quad \text{gen}(E.addr '!=' L.array.base '[' L.addr']); \}$   

 $E \rightarrow \text{id} \quad \{ E.addr = \text{top.get(id.lexeme)}; \}$   

 $L \rightarrow \text{id} [ E ] \quad \{ L.array = \text{top.get(id.lexeme)};$   

 $\quad \quad \quad L.type = L.array.type.elem;$   

 $\quad \quad \quad L.addr = \text{new Temp}();$   

 $\quad \quad \quad \text{gen}(L.addr '!=' E.addr '*' L.type.width); \}$   

 $L \rightarrow L_1 [ E ] \quad \{ L.array = L_1.array;$   

 $\quad \quad \quad L.type = L_1.type.elem;$   

 $\quad \quad \quad t = \text{new Temp}();$   

 $\quad \quad \quad L.addr = \text{new Temp}();$   

 $\quad \quad \quad \text{gen}(t '!=' E.addr '*' L.type.width); \}$   

 $\quad \quad \quad \text{gen}(L.addr '!=' L_1.addr '+' t); \}$ 

```

Home work

Generate TAC for

$c=a[i][j];$

$B[i][j]=b;$

Control Flow

Translation of statements such as **if-else-statements and while-statements**

Key step: Translation of **boolean expressions**

Boolean expressions are used as

- (i) Conditional expressions in statements that alter the **flow of control**
 - (ii) A boolean expression can evaluate true Or false as values. Such boolean expressions can be evaluated in analogy to **arithmetic expressions** using three-address instructions with logical operators
-
- The **intended use of boolean expressions** is determined by its **syntactic context**.
 - We concentrate on the use of boolean expressions to **alter the flow of control**.
 - For clarity, we introduce a **new nonterminal B**

Boolean Expressions

$B \rightarrow B \text{ || } B \mid B \text{ && } B \mid !B \mid (B) \mid E \text{ rel } E \mid \text{true} \mid \text{false}$

$<$, \leq , $=$, \neq , $>$, or \geq is represented by **rel**.

Given the expression $B_1 \text{ || } B_2$, if we determine that B_1 is true, then we can conclude that the entire expression is true without having to evaluate B_2 . Similarly, given $B_1 \&\& B_2$, if B_1 is false, then the entire expression is false.

The semantic definition of the programming language determines whether all parts of a boolean expression must be evaluated.

Short-Circuit Code

In *short-circuit* (or *jumping*) code, the boolean operators $\&\&$, || , and $!$ translate into jumps.

The operators themselves do not appear in the code; instead, the value of a boolean expression is represented by a position in the code sequence.

Control Flow

```
if ( x < 100 || x > 200 && x != y ) x = 0;
```

```
if x < 100 goto L2 ←
```

```
iffalse x > 200 goto L1
```

```
iffalse x != y goto L1
```

→ L₂: x = 0

→ L₁:

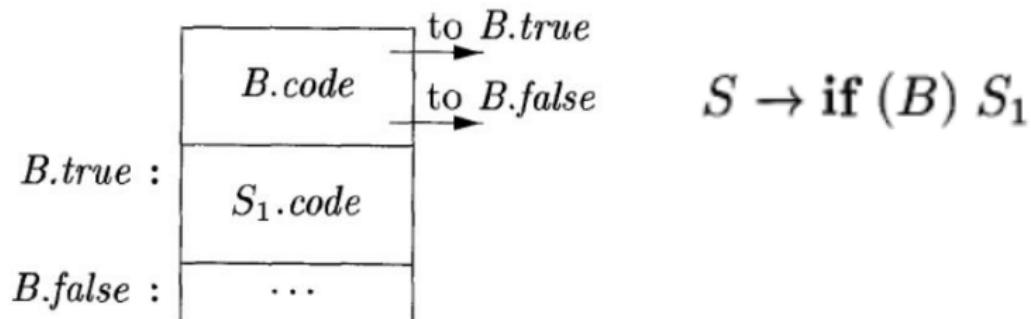
Translation of Flow-of-Control Statements into three-address code

$S \rightarrow \text{if } (B) S_1$

$S \rightarrow \text{if } (B) S_1 \text{ else } S_2$

$S \rightarrow \text{while } (B) S_1$

- Nonterminal B represents a boolean expression and nonterminal S represents a statement.
- Both B and S have a synthesized attribute **code**, which gives the translation into three-address instructions.
- Inherited attributes **B.true**, **B.false**, **S.next** generate **labels** for control flow
- **B.true the label** to which control flows if B is true,
- **B.false, the label** to which control flows if B is false.
- With a statement S , we associate an inherited attribute **S.next** denoting a **label** for the instruction **immediately after the code for S** .



(a) if

- Nonterminal **B** represents a boolean expression and nonterminal **S** represents a statement.
- Both **B** and **S** have a synthesized attribute **code**, which gives the translation into three-address instructions.
- Inherited attributes **B.true**, **B.false**, **S.next** generate labels for control flow
- **B.true** the label to which control flows if B is true,
- **B.false**, the label to which control flows if B is false.
- With a statement S, we associate an inherited attribute **S.next** denoting a label for the instruction immediately after the code for S.

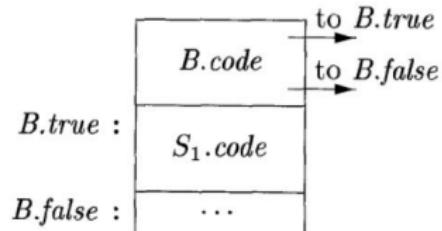
PRODUCTION	SEMANTIC RULES
$P \rightarrow S$	$S.next = newlabel()$ $P.code = S.code \parallel label(S.next)$
$S \rightarrow \text{assign}$	$S.code = \text{assign}.code$
$S \rightarrow \text{if} (B) S_1$	$B.true = newlabel()$ $B.false = S_1.next = S.next$ $S.code = B.code \parallel label(B.true) \parallel S_1.code$

- **newlabel()** creates a new **label** each time it is called,
- **label(L)** attaches label L to the next three-address instruction to be generated
- A **program consists** of a statement generated by $P \rightarrow S$.
- The semantic rules associated with this production initialize **S.next** to a new label.
- **P.code** consists of **S.code** followed by the new label **S.next**.

PRODUCTION	SEMANTIC RULES
$P \rightarrow S$	$S.next = newlabel()$ $P.code = S.code \parallel label(S.next)$
$S \rightarrow \text{assign}$	$S.code = \text{assign}.code$  standard
$S \rightarrow \text{if} (B) S_1$	$B.true = newlabel()$ $B.false = S_1.next = S.next$ $S.code = B.code \parallel label(B.true) \parallel S_1.code$

assign in the production $S \rightarrow \text{assign}$ is a placeholder for assignment statements.

- In translating $S \rightarrow \text{if} (B) S_1$, the semantic rules create a **new label $B.true$** and **attach it to the first three-address instruction** generated for the statement S_1 .
- Thus, **jumps to $B.true$** within the code for B will go to the **code S_1** .
- By **setting $B.false$ to $S.next$** , we ensure that control will **skip the code for S_1** if B evaluates to **false**.

$$S \rightarrow \text{if } (B) S_1$$
$$B.\text{true} = \text{newlabel}()$$
$$B.\text{false} = S_1.\text{next} = S.\text{next}$$
$$S.\text{code} = B.\text{code} \parallel \text{label}(B.\text{true}) \parallel S_1.\text{code}$$


(a) if

- In translating $S \rightarrow \text{if } (B) S_1$, the semantic rules create a **new label *B.true*** and **attach** it to the **first three-address instruction** generated for the statement S_1 .
- Thus, **jumps to *B.true*** within the code for B will go to the **code S_1** .
- By **setting *B.false* to $S.\text{next}$** , we ensure that control will **skip the code for S_1** if B evaluates to **false**.

Translation of Boolean Expressions

$B \rightarrow E_1 \text{ rel } E_2$

$B.\text{code} = E_1.\text{code} \parallel E_2.\text{code}$

$\parallel \text{gen('if' } E_1.\text{addr rel.op } E_2.\text{addr 'goto' } B.\text{true})$
 $\parallel \text{gen('goto' } B.\text{false})$

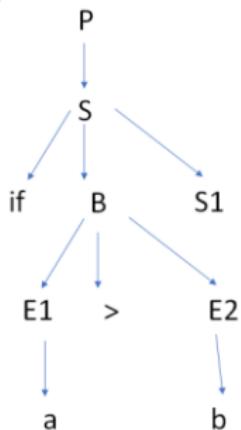
if a>b

x=0

$P \rightarrow S$

$S \rightarrow \text{assign}$

$S \rightarrow \text{if (} B \text{) } S_1$



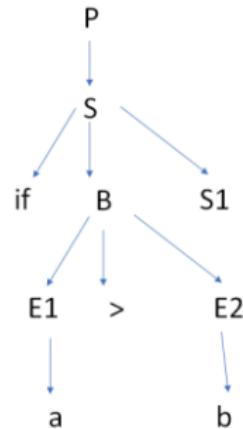
PRODUCTION	SEMANTIC RULES
$P \rightarrow S$	$S.next = newlabel()$ $P.code = S.code \parallel label(S.next)$
$S \rightarrow \text{assign}$	$S.code = \text{assign}.code$
$S \rightarrow \text{if} (B) S_1$	$B.true = newlabel()$ $B.false = S_1.next = S.next$ $S.code = B.code \parallel label(B.true) \parallel S_1.code$

L1=S.next

P.Code:

S.code

L1:

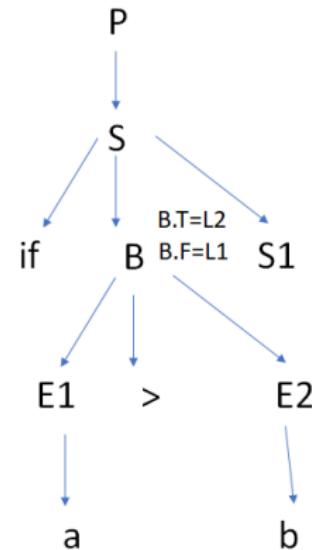


if a>b
x=0

PRODUCTION	SEMANTIC RULES
$P \rightarrow S$	$S.next = newlabel()$ $P.code = S.code \parallel label(S.next)$
$S \rightarrow \text{assign}$	$S.code = \text{assign}.code$
$S \rightarrow \text{if} (B) S_1$	$B.true = newlabel()$ $B.false = S_1.next = S.next$ $S.code = B.code \parallel label(B.true) \parallel S_1.code$

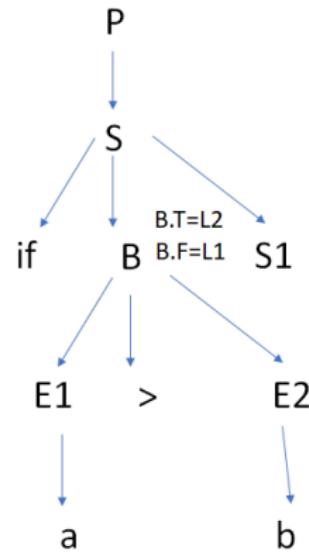
$L1 = S.next$
 $B.true = L2$
 $B.false = L1$

P.Code:
 B.code
 L2
 L1:



if a>b
x=0

PRODUCTION	SEMANTIC RULES
$P \rightarrow S$	$S.next = newlabel()$ $P.code = S.code \parallel label(S.next)$
$S \rightarrow \text{assign}$	$S.code = \text{assign}.code$
$S \rightarrow \text{if} (B) S_1$	$B.true = newlabel()$ $B.false = S_1.next = S.next$ $S.code = B.code \parallel label(B.true) \parallel S_1.code$
$L1 = S.next$ $B.true = L2$ $B.false = L1$	P.Code: if a>b goto L2 goto L1 L2: x=0 L1:.....
$B \rightarrow E_1 \text{ rel } E_2$	$B.code = E_1.code \parallel E_2.code$ $\parallel \text{gen('if' } E_1.\text{addr rel.op } E_2.\text{addr 'goto' } B.true)$ $\parallel \text{gen('goto' } B.false)$

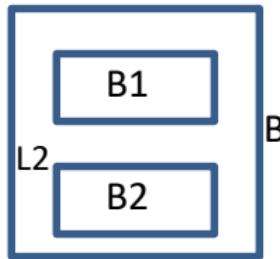


$B \rightarrow E_1 \text{ rel } E_2$

$B.code = E_1.code \parallel E_2.code$
 $\parallel \text{gen('if' } E_1.\text{addr rel.op } E_2.\text{addr 'goto' } B.true)$
 $\parallel \text{gen('goto' } B.false)$

Translation of Boolean Expressions

PRODUCTION	SEMANTIC RULES
$B \rightarrow B_1 \parallel B_2$	$B_1.\text{true} = B.\text{true}$ $B_1.\text{false} = \text{newlabel}()$ $B_2.\text{true} = B.\text{true}$ $B_2.\text{false} = B.\text{false}$ $B.\text{code} = B_1.\text{code} \parallel \text{label}(B_1.\text{false}) \parallel B_2.\text{code}$

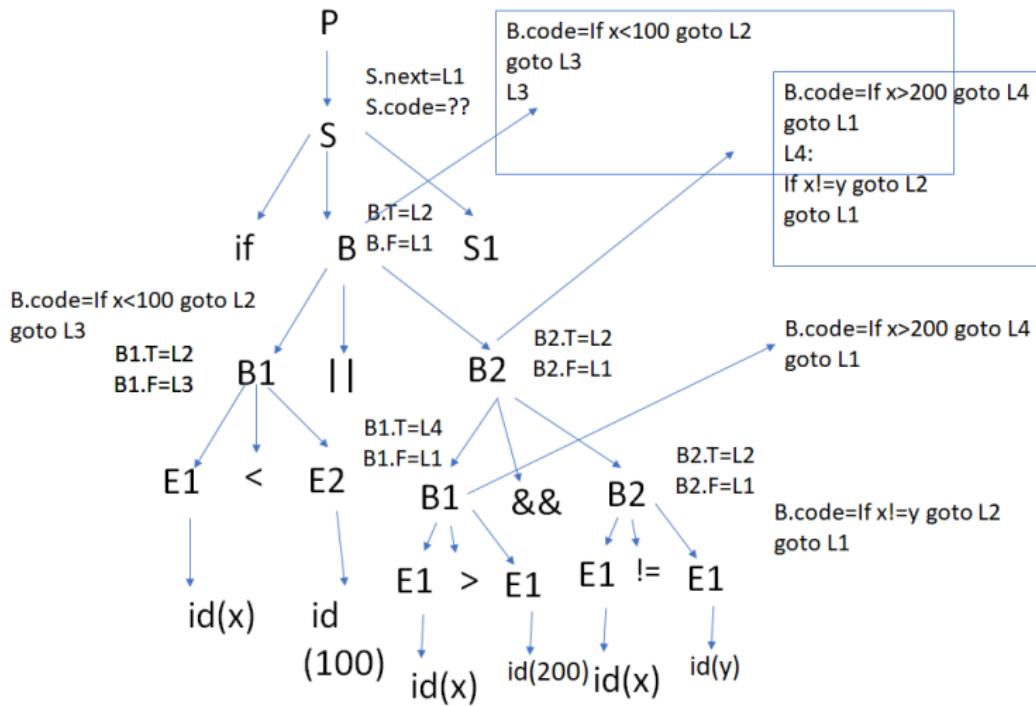


The true and false exits of B_2 are the same as the true and false exits of B , respectively.

Translation of Boolean Expressions

PRODUCTION	SEMANTIC RULES
$B \rightarrow B_1 \mid\mid B_2$	$B_1.\text{true} = B.\text{true}$ $B_1.\text{false} = \text{newlabel}()$ $B_2.\text{true} = B.\text{true}$ $B_2.\text{false} = B.\text{false}$ $B.\text{code} = B_1.\text{code} \mid\mid \text{label}(B_1.\text{false}) \mid\mid B_2.\text{code}$
$B \rightarrow B_1 \&\& B_2$	$B_1.\text{true} = \text{newlabel}()$ $B_1.\text{false} = B.\text{false}$ $B_2.\text{true} = B.\text{true}$ $B_2.\text{false} = B.\text{false}$ $B.\text{code} = B_1.\text{code} \mid\mid \text{label}(B_1.\text{true}) \mid\mid B_2.\text{code}$
$B \rightarrow ! B_1$	$B_1.\text{true} = B.\text{false}$ $B_1.\text{false} = B.\text{true}$ $B.\text{code} = B_1.\text{code}$
$B \rightarrow E_1 \text{ rel } E_2$	$B.\text{code} = E_1.\text{code} \mid\mid E_2.\text{code}$ $\mid\mid \text{gen('if' } E_1.\text{addr rel.op } E_2.\text{addr 'goto' } B.\text{true})$ $\mid\mid \text{gen('goto' } B.\text{false})$
$B \rightarrow \text{true}$	$B.\text{code} = \text{gen('goto' } B.\text{true})$
$B \rightarrow \text{false}$	$B.\text{code} = \text{gen('goto' } B.\text{false})$

```
if( x < 100 || x > 200 && x != y ) x = 0;
```

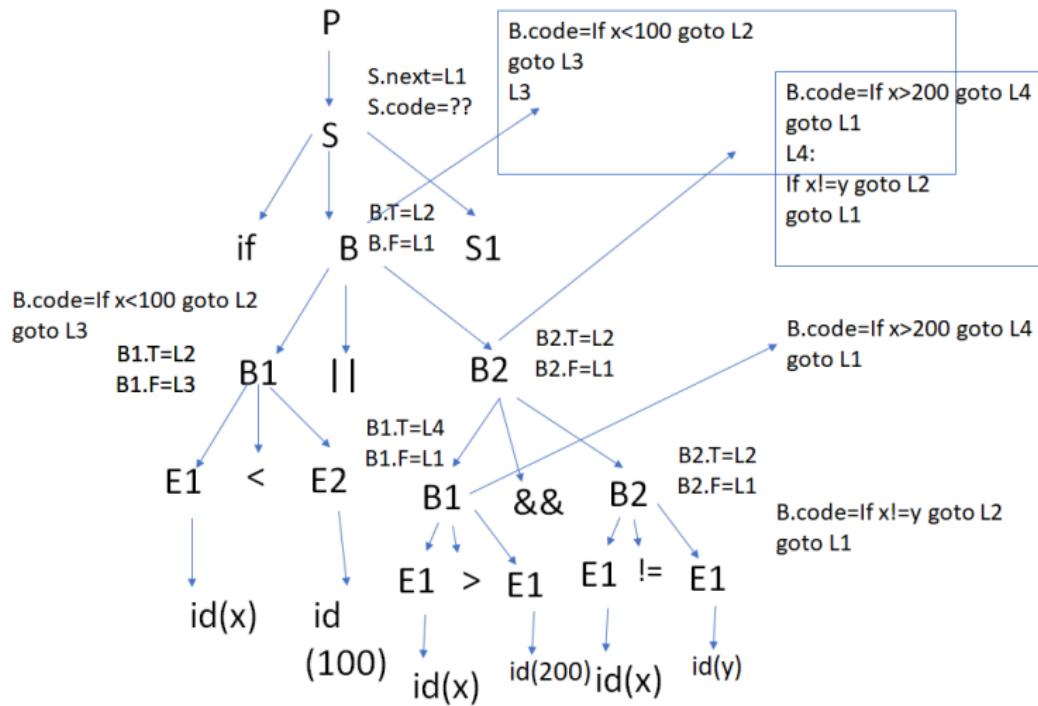


```

        if x < 100 goto L2
        goto L3
L3:   if x > 200 goto L4
        goto L1
L4:   if x != y goto L2
        goto L1
L2:   x = 0
L1:

```

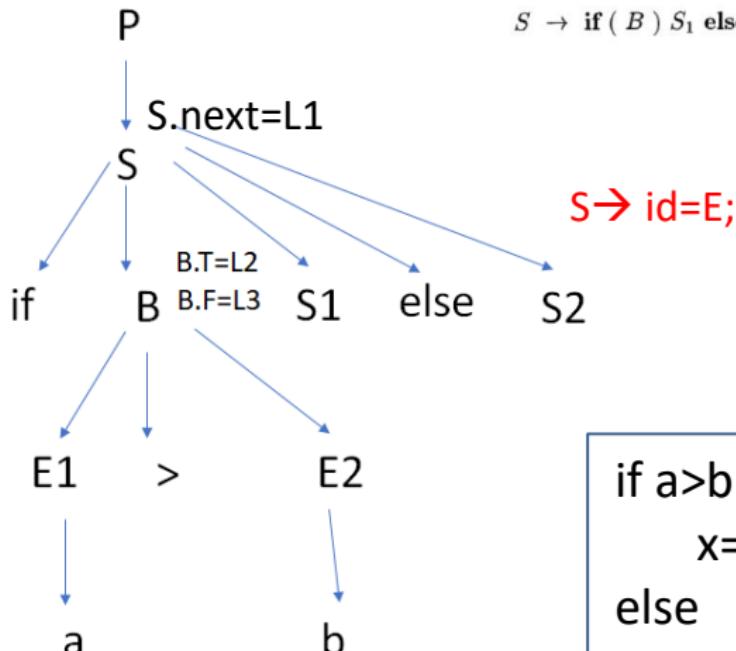
```
if( x < 100 || x > 200 && x != y ) x = 0;
```



Translation of if-else statement

	PRODUCTION	SEMANTIC RULES								
	$P \rightarrow S$	$S.next = newlabel()$ $P.code = S.code \parallel label(S.next)$								
	$S \rightarrow \text{if} (B) S_1 \text{ else } S_2$	$B.true = newlabel()$ $B.false = newlabel()$ $S_1.next = S_2.next = S.next$ $S.code = B.code$ $\parallel label(B.true) \parallel S_1.code$ $\parallel \text{gen('goto' } S.next)$ $\parallel label(B.false) \parallel S_2.code$								
$B.true :$	<table border="1"><tr><td>$B.code$</td><td>to $B.true$</td></tr><tr><td></td><td>to $B.false$</td></tr><tr><td>$S_1.code$</td><td></td></tr><tr><td>$\text{goto } S.next$</td><td></td></tr></table>	$B.code$	to $B.true$		to $B.false$	$S_1.code$		$\text{goto } S.next$		
$B.code$	to $B.true$									
	to $B.false$									
$S_1.code$										
$\text{goto } S.next$										
$B.false :$	<table border="1"><tr><td>$S_2.code$</td><td></td></tr></table>	$S_2.code$								
$S_2.code$										
$S.next :$	<table border="1"><tr><td>...</td><td></td></tr></table>	...								
...										
(b) if-else		<ul style="list-style-type: none">In translating the if-else-statement, if B is true, the code for the boolean expression B has jumps to the first instruction of the code for S1,if B is false, control jumps to the first instruction of the code for S2.Further, control flows from both S1 and S2 to the three-address instruction immediately following the code for S — its label is given by the inherited attribute S.next.An explicit goto S.next appears after the code for S1 to skip over the code for S2. No goto is needed after S2, since S2.next is the same as S.next.								

PRODUCTION	SEMANTIC RULES
$P \rightarrow S$	$S.next = newlabel()$ $P.code = S.code \parallel label(S.next)$
$S \rightarrow \text{if} (B) S_1 \text{ else } S_2$	$B.true = newlabel()$ $B.false = newlabel()$ $S_1.next = S_2.next = S.next$ $S.code = B.code$ $\parallel label(B.true) \parallel S_1.code$ $\parallel \text{gen('goto' } S.next)$ $\parallel label(B.false) \parallel S_2.code$
$S \rightarrow id = E;$	



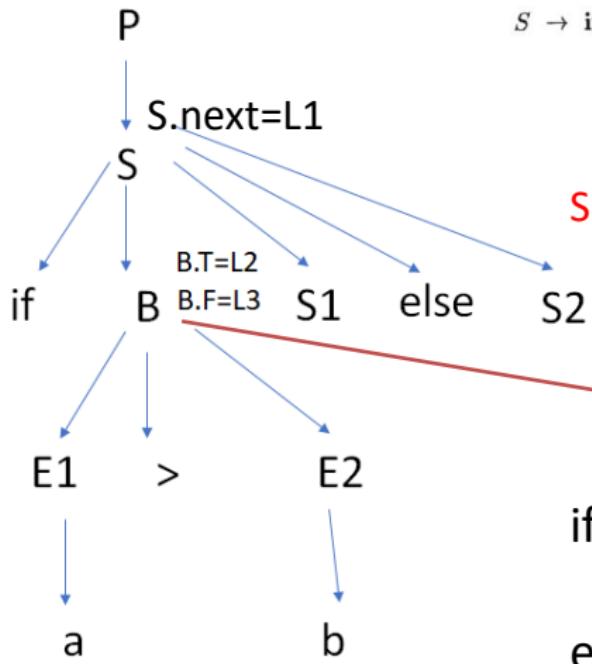
```

if a>b
  x=0;
else
  y=0;
  
```

L1:

$B \rightarrow E_1 \text{ rel } E_2$	$B.code = E_1.code \parallel E_2.code$ $\parallel \text{gen('if' } E_1.addr \text{ rel.op } E_2.addr \text{ 'goto' } B.true)$ $\parallel \text{gen('goto' } B.false)$
--------------------------------------	---

PRODUCTION	SEMANTIC RULES
$P \rightarrow S$	$S.next = newlabel()$ $P.code = S.code \parallel label(S.next)$
$S \rightarrow \text{if } (B) S_1 \text{ else } S_2$	$B.true = newlabel()$ $B.false = newlabel()$ $S_1.next = S_2.next = S.next$ $S.code = B.code$ $\quad \parallel label(B.true) \parallel S_1.code$ $\quad \parallel gen('goto' S.next)$ $\quad \parallel label(B.false) \parallel S_2.code$
$S \rightarrow id = E;$	



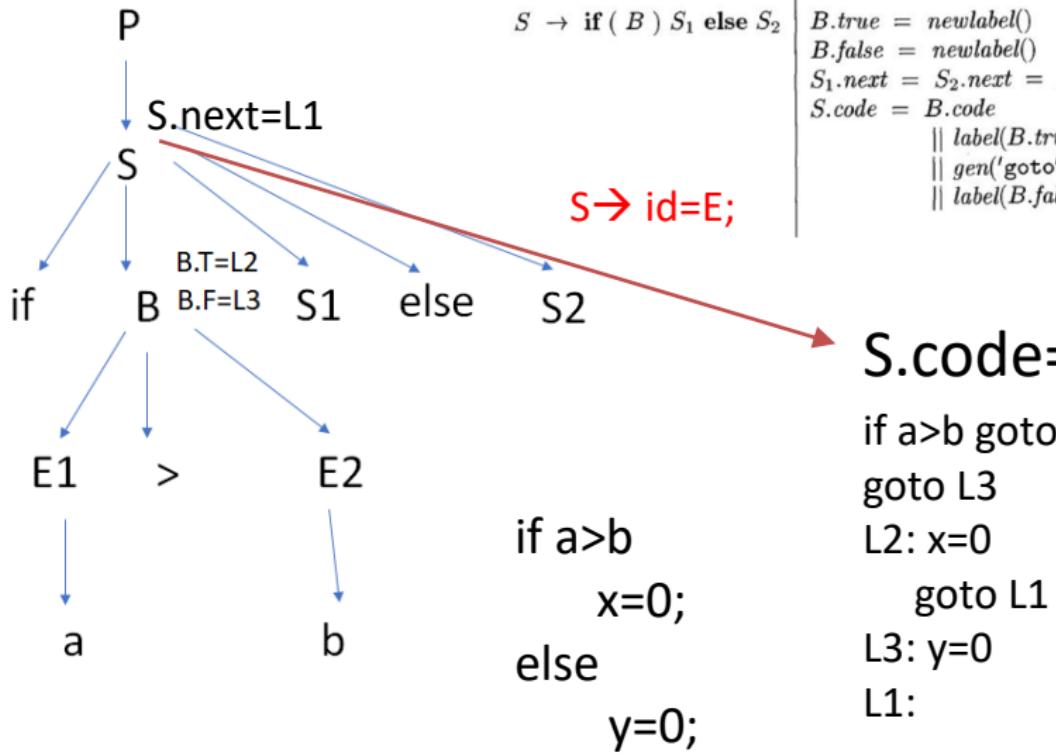
B.code=

if $a > b$ goto L2

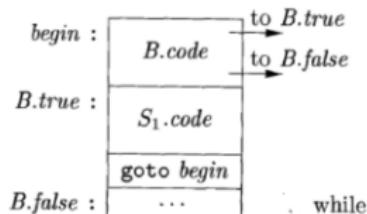
goto L3

$$B.\text{code} = E_1.\text{code} \parallel E_2.\text{code} \\ \parallel \text{gen}'\text{'if' } E_1.\text{addr rel.op } E_2.\text{addr 'goto' } B.\text{true}) \\ \parallel \text{gen}'\text{'goto' } B.\text{false})$$

PRODUCTION	SEMANTIC RULES
$P \rightarrow S$	$S.next = newlabel()$ $P.code = S.code \parallel label(S.next)$
$S \rightarrow \text{if} (B) S_1 \text{ else } S_2$	$B.true = newlabel()$ $B.false = newlabel()$ $S_1.next = S_2.next = S.next$ $S.code = B.code$ $\parallel label(B.true) \parallel S_1.code$ $\parallel \text{gen('goto' } S.next)$ $\parallel label(B.false) \parallel S_2.code$



Translation of while statement



$S \rightarrow \text{while} (B) S_1$

```
begin = newlabel()
B.true = newlabel()
B.false = S.next
S.next = begin
S.code = label(begin) || B.code
         || label(B.true) || S1.code
         || gen('goto' begin)
```

- We use a **local variable begin** to hold a new label attached to the **first instruction** for this while-statement,
 - which is also the first instruction for B.
 - Variable begin is local to the semantic rules for this production.
- The **inherited label S.next** marks the instruction that control must flow to if B is false; hence, **B.false is set to be S.next**.
- A **new label B.true is attached to the first instruction for S1**;
 - the code for B generates a jump to this label if B is true.
- **After the code for S1 we place the instruction goto begin**, which causes a **jump back to the beginning of the code** for the boolean expression.
- Note that S1.next is set to this label begin

Avoiding Redundant Gotos

```
if x > 200 goto L4  
goto L1
```

L4: ...

Free fall to L4 when x>200
Jumps to L1 only when x<=200

Instead, consider the instruction:

```
ifFalse x > 200 goto L1  
L4: ...
```

The code for statement S1
immediately follows the code for the
boolean expression B

if

B

L1 S1
L2

$S \rightarrow \text{if } (B) S_1$

$B.\text{true} = \text{fall}$
 $B.\text{false} = S_1.\text{next} = S.\text{next}$
 $S.\text{code} = B.\text{code} \parallel S_1.\text{code}$

$S \rightarrow \text{if } (B) S_1$

$B.\text{true} = \text{newlabel}()$ **L1**
 $B.\text{false} = S_1.\text{next} = S.\text{next}$
 $S.\text{code} = B.\text{code} \parallel \text{label}(B.\text{true}) \parallel S_1.\text{code}$

Avoiding Redundant Gotos

$B \rightarrow E_1 \text{ rel } E_2$

$test = E_1.\text{addr rel.op } E_2.\text{addr}$

$s = \text{if } B.\text{true} \neq \text{fall and } B.\text{false} \neq \text{fall then}$

$\quad \text{gen('if' test 'goto' } B.\text{true}) \parallel \text{gen('goto' } B.\text{false})$

$\quad \text{else if } B.\text{true} \neq \text{fall then gen('if' test 'goto' } B.\text{true})$

Jump when true

$\quad \text{else if } B.\text{false} \neq \text{fall then gen('iffalse' test 'goto' } B.\text{false})$

Jump when false

$\quad \text{else ''}$

$B.\text{code} = E_1.\text{code} \parallel E_2.\text{code} \parallel s$

$B \rightarrow E_1 \text{ rel } E_2$

$B.\text{code} = E_1.\text{code} \parallel E_2.\text{code}$
 $\parallel \text{gen('if' } E_1.\text{addr rel.op } E_2.\text{addr 'goto' } B.\text{true})$
 $\parallel \text{gen('goto' } B.\text{false})$

Avoiding Redundant Gotos

Semantic rules for $B \rightarrow B_1 \parallel B_2$

PRODUCTION	SEMANTIC RULES
$B \rightarrow B_1 \parallel B_2$	$B_1.\text{true} = B.\text{true}$
	$B_1.\text{false} = \text{newlabel}()$
	$B_2.\text{true} = B.\text{true}$
	$B_2.\text{false} = B.\text{false}$
	$B.\text{code} = B_1.\text{code} \parallel \text{label}(B_1.\text{false}) \parallel B_2.\text{code}$

$B_1.\text{true} = \text{if } B.\text{true} \neq \text{fall} \text{ then } B.\text{true} \text{ else newlabel}()$

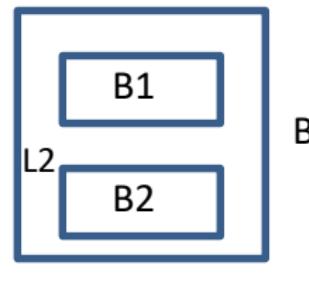
$B_1.\text{false} = \text{fall}$

$B_2.\text{true} = B.\text{true}$

$B_2.\text{false} = B.\text{false}$

$B.\text{code} = \text{if } B.\text{true} \neq \text{fall} \text{ then } B_1.\text{code} \parallel B_2.\text{code}$
 $\text{else } B_1.\text{code} \parallel B_2.\text{code} \parallel \text{label}(B_1.\text{true})$

- Meaning of label fall for B is different from its meaning for B1.
- Suppose B.true is fall; i.e, control falls through B, if B evaluates to true.
- Although B evaluates to true if B1 does, B1.true must ensure that control jumps over the code for B2 to get to the next instruction after B.



Avoiding Redundant Gotos

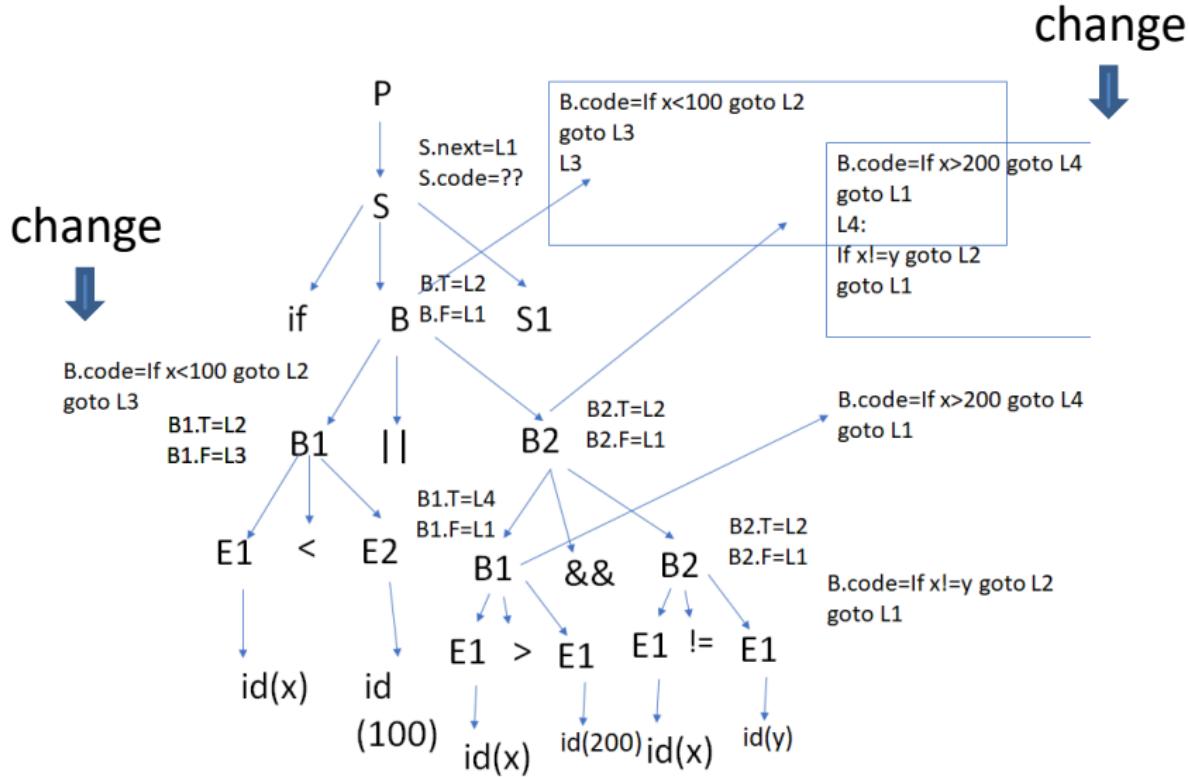
```
if( x < 100 || x > 200 && x != y ) x = 0;
```

```
if x < 100 goto L2
ifFalse x > 200 goto L1
ifFalse x != y goto L1
L2: x = 0
L1:
```

```
if x < 100 goto L2
goto L3
L3: if x > 200 goto L4
      goto L1
L4: if x != y goto L2
      goto L1
L2: x = 0
L1:
```

Avoiding Redundant Gotos

```
if( x < 100 || x > 200 && x != y ) x = 0;
```



Backpatching

- Key problem when generating code for **boolean expressions and flow-of-control statements** is that
- Matching a **jump instruction** with the **target of the jump**.
- For example, the translation of the **boolean expression B** in $\text{if} (B) S$ contains a **jump**, for when B is false, **to the instruction following the code for S**.
- In a one-pass translation, **B must be translated before S is examined**.
- What then is the **target of the goto that jumps over the code for S**?

if



$S \rightarrow \text{if} (B) S_1$



$B.\text{true} = \text{newlabel}()$
 $B.\text{false} = S_1.\text{next} = S.\text{next}$
 $S.\text{code} = B.\text{code} \parallel \text{label}(B.\text{true}) \parallel S_1.\text{code}$

We managed with **inherited** attributes

Backpatching

- Lists of jumps are passed as **synthesized attributes**.
- When a **jump is generated**, the **target** of the jump is temporarily left **unspecified**.
- Each such **jump** is put on a **list of jumps** whose **labels are to be filled in when the proper label** can be determined.
- All of the jumps on a list have the **same target label**.

- Synthesized attributes **truelist and falselist** of nonterminal **B** are used to manage labels.
- **B.truelist** will be a **list of jump or conditional jump instructions** into which, we must insert the label to which **control goes if B is true**.
- **B.falselist** likewise is the **list of instructions** that eventually get the **label** to which **control goes when B is false**.
- **Code is generated for B**, jumps to the true and false exits are left incomplete, with the **label field unfilled**.
- These **incomplete jumps** are placed on lists pointed to by **B.truelist and B.falselist**, as appropriate.

Backpatching

- We generate instructions into an **instruction array**,
- **Labels** will be **indices** into this array.
- To manipulate lists of jumps, we use **three functions**:

1. **makelist(i)** creates a new list containing an index i into the array of instructions; makelist returns a pointer to the newly created list.
2. **merge(p1,p2)** concatenates the lists pointed to by p1 and p2, and returns a pointer to the concatenated list.
3. **backpatch(p,i)** inserts i as the target label for each of the instructions on the list pointed to by p.

Backpatching for Boolean Expressions

- We now construct a translation scheme suitable for generating code for Boolean expressions during bottom-up parsing.
- A **marker nonterminal M** causes a semantic action to pick up, at appropriate times, the **index of the next instruction** to be generated.

1) $B \rightarrow B_1 \mid\mid M B_2 \quad \{ \text{backpatch}(B_1.\textit{falselist}, M.\textit{instr});$
 $B.\textit{truelist} = \text{merge}(B_1.\textit{truelist}, B_2.\textit{truelist});$
 $B.\textit{falselist} = B_2.\textit{falselist}; \}$

- If B_1 is true, then B is also true, so the jumps on **$B_1.\textit{truelist}$ become part of $B.\textit{truelist}$** .
- If **B_1 is false**, however, we must next **test B_2** ,
- So the **target for the jumps $B_1.\textit{falselist}$ must be the beginning of the code generated for B_2** .
- **This target** is obtained using the **marker nonterminal M**.
- That nonterminal M produces, as a **synthesized attribute $M.\textit{instr}$** , the index of the next instruction, **just before B_2 code starts being generated**.

Backpatching for Boolean Expressions

To obtain that instruction index, we associate with the production $M \rightarrow \epsilon$ the semantic action

$$\{ M.instr = nextinstr; \}$$

The variable $nextinstr$ holds the index of the next instruction to follow. This value will be backpatched onto the $B_1.falselist$ (i.e., each instruction on the list $B_1.falselist$ will receive $M.instr$ as its target label) when we have seen the remainder of the production $B \rightarrow B_1 \mid\mid M B_2$.

- 1) $B \rightarrow B_1 \text{ || } M \text{ } B_2$ { *backpatch*($B_1.\textit{falselist}$, $M.\textit{instr}$);
 $B.\textit{truelist} = \text{merge}(B_1.\textit{truelist}, B_2.\textit{truelist})$;
 $B.\textit{falselist} = B_2.\textit{falselist}$; }

- 2) $B \rightarrow B_1 \text{ && } M \text{ } B_2$ { *backpatch*($B_1.\textit{truelist}$, $M.\textit{instr}$);
 $B.\textit{truelist} = B_2.\textit{truelist}$;
 $B.\textit{falselist} = \text{merge}(B_1.\textit{falselist}, B_2.\textit{falselist})$; }

- 3) $B \rightarrow ! \text{ } B_1$ { $B.\textit{truelist} = B_1.\textit{falselist}$;
 $B.\textit{falselist} = B_1.\textit{truelist}$; }

- 4) $B \rightarrow (\text{ } B_1 \text{ })$ { $B.\textit{truelist} = B_1.\textit{truelist}$;
 $B.\textit{falselist} = B_1.\textit{falselist}$; }

- 5) $B \rightarrow E_1 \text{ rel } E_2$ { $B.\textit{truelist} = \text{makelist}(nextinstr)$;
 $B.\textit{falselist} = \text{makelist}(nextinstr + 1)$;
 $\text{emit('if' } E_1.\textit{addr rel.op } E_2.\textit{addr 'goto } '_{ })$;
 $\text{emit('goto } '_{ })$; }

- 6) $B \rightarrow \text{true}$ { $B.\textit{truelist} = \text{makelist}(nextinstr)$;
 $\text{emit('goto } '_{ })$; }

- 7) $B \rightarrow \text{false}$ { $B.\textit{falselist} = \text{makelist}(nextinstr)$;
 $\text{emit('goto } '_{ })$; }

- 8) $M \rightarrow \epsilon$ { $M.\textit{instr} = nextinstr$; }

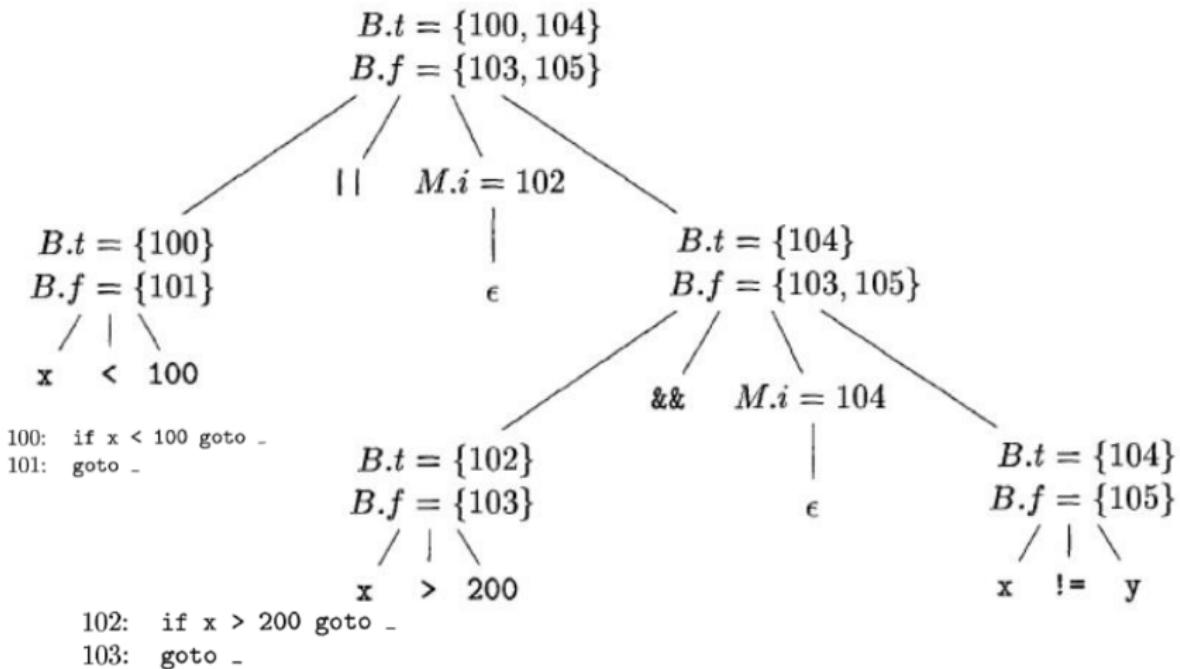
```

 $B \rightarrow B_1 \sqcup M B_2$  { backpatch( $B_1.falselist, M.instr$ );
 $B.truelist = merge(B_1.truelist, B_2.truelist)$ ;
 $B.falselist = B_2.falselist$ ; }

```

Consider again the expression

$$x < 100 \sqcup x > 200 \And x \neq y$$



```
100: if x < 100 goto -
101: goto -
102: if x > 200 goto 104
103: goto -
104: if x != y goto -
105: goto -
```

Consider again the expression

$$x < 100 \text{ || } x > 200 \text{ && } x \neq y$$



(a) After backpatching 104 into instruction 102.

```
100: if x < 100 goto -
101: goto 102
102: if y > 200 goto 104
103: goto -
104: if x != y goto -
105: goto -
```



(b) After backpatching 102 into instruction 101.

The entire expression is true if and only if the gotos of instructions 100 or 104 are reached, and is false if and only if the gotos of instructions 103 or 105 are reached. These instructions will have their targets filled in later in the compilation, when it is seen what must be done depending on the truth or falsehood of the expression.

Flow-of-Control Statements

Boolean expressions generated by nonterminal B have two lists of jumps, **B.truelist** and **B.falselist**, corresponding to the true and false exits from the **code for B**

Statements generated by **nonterminals S and L** have a list of **unfilled jumps**

S.nextlist is a list of **all conditional and unconditional jumps** to the instruction following the code for **statement S** in execution order.
L.nextlist is defined similarly.

Flow-of-Control Statements

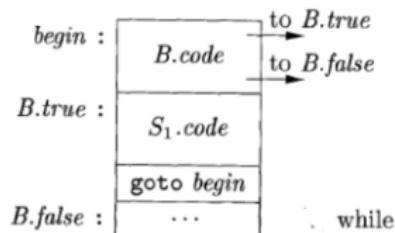
- 1) $S \rightarrow \text{if}(B) M S_1 \{ \text{backpatch}(B.\text{truelist}, M.\text{instr}); S.\text{nextlist} = \text{merge}(B.\text{falselist}, S_1.\text{nextlist}); \}$
- 2) $S \rightarrow \text{if}(B) M_1 S_1 N \text{ else } M_2 S_2 \{ \text{backpatch}(B.\text{truelist}, M_1.\text{instr}); \text{backpatch}(B.\text{falselist}, M_2.\text{instr}); \text{temp} = \text{merge}(S_1.\text{nextlist}, N.\text{nextlist}); S.\text{nextlist} = \text{merge}(\text{temp}, S_2.\text{nextlist}); \}$
- 3) $S \rightarrow \text{while } M_1 (B) M_2 S_1 \{ \text{backpatch}(S_1.\text{nextlist}, M_1.\text{instr}); \text{backpatch}(B.\text{truelist}, M_2.\text{instr}); S.\text{nextlist} = B.\text{falselist}; \text{emit}'\text{goto}' M_1.\text{instr}); \}$
- 4) $S \rightarrow \{ L \} \{ S.\text{nextlist} = L.\text{nextlist}; \}$
- 5) $S \rightarrow A ; \{ S.\text{nextlist} = \text{null}; \}$
- 6) $M \rightarrow \epsilon \{ M.\text{instr} = \text{nextinstr}; \}$
- 7) $N \rightarrow \epsilon \{ N.\text{nextlist} = \text{makelist}(\text{nextinstr}); \text{emit}'\text{goto } \text{`}'; \}$
- 8) $L \rightarrow L_1 M S \{ \text{backpatch}(L_1.\text{nextlist}, M.\text{instr}); L.\text{nextlist} = S.\text{nextlist}; \}$
- 9) $L \rightarrow S \{ L.\text{nextlist} = S.\text{nextlist}; \}$

Flow-of-Control Statements (while)

$S \rightarrow \text{while } (B) S_1$

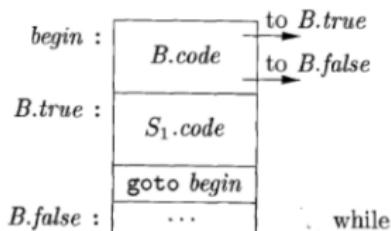
$S \rightarrow \text{while } M_1 (B) M_2 S_1$

```
{ backpatch( $S_1.\text{nextlist}$ ,  $M_1.\text{instr}$ );
  backpatch( $B.\text{truelist}$ ,  $M_2.\text{instr}$ );
   $S.\text{nextlist} = B.\text{falselist};$ 
  emit('goto'  $M_1.\text{instr}$ ); }
```



- The two occurrences of the **marker nonterminal M** record the instruction numbers of the
- M1--beginning of the code for B (**begin**) and the
- M2--beginning of the code for S1 (**B.true**).

Flow-of-Control Statements (while)



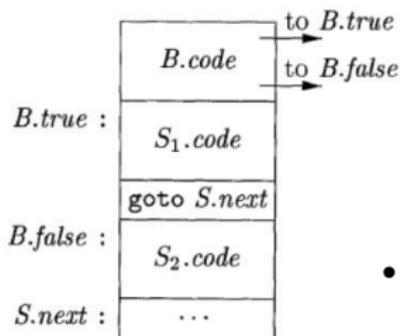
$S \rightarrow \text{while } M_1(B) M_2 S_1$

```
{ backpatch(S1.nextlist, M1.instr);  
backpatch(B.truelist, M2.instr);  
S.nextlist = B.falselist;  
emit('goto' M1.instr); }
```

- Attribute **M.instr** points to the number of the **next instruction**.
- After the **body S₁ of the while-statement is executed**,
 - Control flows to the beginning.
 - We backpatch **S₁.nextlist** to make all targets on that list be **M₁.instr**.
- An **explicit jump** to the beginning of the code for B is appended after the code for S₁
- B. truelist** is backpatched to go to the **beginning of S₁** ‘ by making jumps on **B.truelist** go to **M₂.instr**

Flow-of-Control Statements (if-else)

if (B) S_1 else S_2



$S \rightarrow \text{if } (B) M_1 S_1 N \text{ else } M_2 S_2$

{
 backpatch($B.\text{truelist}$, $M_1.\text{instr}$);
 backpatch($B.\text{falselist}$, $M_2.\text{instr}$);
 temp = merge($S_1.\text{nextlist}$, $N.\text{nextlist}$);
 $S.\text{nextlist} = \text{merge}(\text{temp}, S_2.\text{nextlist})$;
}

$N \rightarrow \epsilon$

{
 $N.\text{nextlist} = \text{makelist}(\text{nextinstr})$;
 emit('goto _');
}

$B.\text{true} :$

$S_1.\text{code}$

goto $S.\text{next}$

$B.\text{false} :$

$S_2.\text{code}$

$S.\text{next} :$

...

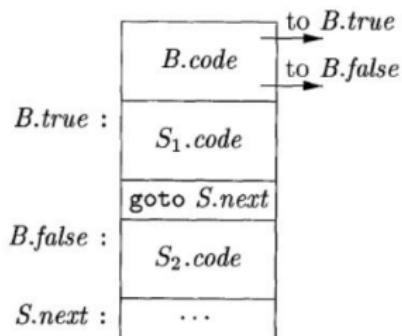
- **S1 is an assignment statement**, we must include at the end of the code for S1 a **jump over the code for S2**
- N has attribute **N.nextlist**, which will be a list consisting of the instruction number of the **jump goto _**

Flow-of-Control Statements (if-else)

if (B) S_1 else S_2

$S \rightarrow \text{if } (B) M_1 S_1 N \text{ else } M_2 S_2$

```
{ backpatch(B.truelist, M1.instr);  
  backpatch(B.falselist, M2.instr);  
  temp = merge(S1.nextlist, N.nextlist);  
  S.nextlist = merge(temp, S2.nextlist); }
```



$N \rightarrow \epsilon$ { $N.\text{nextlist} = \text{makelist}(\text{nextinstr});$
 $\text{emit}(\text{'goto -'})$; }

We backpatch the jumps when B is true to the instruction $M_1.\text{instr}$, the latter is the beginning of the code for S_1 . Similarly, we backpatch jumps when B is false to go to the beginning of the code for S_2 . The list $S.\text{nextlist}$ includes all jumps out of S_1 and S_2 , as well as the jump generated by N . (Variable temp is a temporary that is used only for merging lists.)

Translation of if-else statement

	PRODUCTION	SEMANTIC RULES								
	$P \rightarrow S$	$S.next = newlabel()$ $P.code = S.code \parallel label(S.next)$								
	$S \rightarrow \text{if} (B) S_1 \text{ else } S_2$	$B.true = newlabel()$ $B.false = newlabel()$ $S_1.next = S_2.next = S.next$ $S.code = B.code$								
$B.true :$	<table border="1"><tr><td>$B.code$</td><td>to $B.true$</td></tr><tr><td></td><td>to $B.false$</td></tr><tr><td>$S_1.code$</td><td></td></tr><tr><td>$\text{goto } S.next$</td><td></td></tr></table>	$B.code$	to $B.true$		to $B.false$	$S_1.code$		$\text{goto } S.next$		$\parallel label(B.true) \parallel S_1.code$ $\parallel \text{gen('goto' } S.next)$ $\parallel label(B.false) \parallel S_2.code$
$B.code$	to $B.true$									
	to $B.false$									
$S_1.code$										
$\text{goto } S.next$										
$B.false :$	<table border="1"><tr><td>$S_2.code$</td></tr></table>	$S_2.code$								
$S_2.code$										
$S.next :$	<table border="1"><tr><td>...</td></tr></table>	...								
...										

(b) if-else

Record or Structure data type

$T \rightarrow \text{record } \{'D'\}'$

The fields in this record type are specified by the sequence of declarations generated by D . The approach of Fig. 6.17 can be used to determine the types and relative addresses of fields, provided we are careful about two things:

- The field names within a record must be distinct; that is, a name may appear at most once in the declarations generated by D .
- The offset or relative address for a field name is relative to the data area for that record.

```
float x;  
record { float x; float y; } p;  
record { int tag; float x; float y; } q;
```

; $x = p.x + q.x;$

Record or Structure data type

$T \rightarrow \text{record } \{ 'D' \}'$

$T \rightarrow \text{record } \{ ' \quad \{ Env.push(top); top = new Env(); }$
 $Stack.push(offset); offset = 0; \}$

$D \}' \quad \{ T.type = record(top); T.width = offset;$
 $top = Env.pop(); offset = Stack.pop(); \}$

For convenience, record types will encode both the types and relative addresses of their fields, using a symbol table for the record type. A record type has the form *record*(*t*), where *record* is the type constructor, and *t* is a symbol-table object that holds information about the fields of this record type.

The embedded action before *D* saves the existing symbol table, denoted by *top* and sets *top* to a fresh symbol table. It also saves the current *offset*, and sets *offset* to 0. The declarations generated by *D* will result in types and relative addresses being put in the fresh symbol table. The action after *D* creates a record type using *top*, before restoring the saved symbol table and *offset*.

Record or Structure data type

$T \rightarrow \text{record } \{ D \}$

$T \rightarrow \text{record } \{ \begin{array}{l} \text{Env.push}(top); top = \text{new Env}(); \\ \text{Stack.push}(offset); offset = 0; \end{array} \}$

$D \} \quad \{ T.type = \text{record}(top); T.width = offset; \\ top = \text{Env.pop}(); offset = \text{Stack.pop}(); \}$

Let class *Env* implement symbol tables. The call *Env.push*(*top*) pushes the current symbol table denoted by *top* onto a stack. Variable *top* is then set to a new symbol table. Similarly, *offset* is pushed onto a stack called *Stack*. Variable *offset* is then set to 0.

After the declarations in *D* have been translated, the symbol table *top* holds the types and relative addresses of the fields in this record. Further, *offset* gives the storage needed for all the fields. The second action sets *T.type* to *record*(*top*) and *T.width* to *offset*. Variables *top* and *offset* are then restored to their pushed values to complete the translation of this record type.

Intermediate Code for Procedures

$D \rightarrow \text{define } T \text{ id } (F) \{ S \}$

$F \rightarrow \epsilon \mid T \text{ id }, F$

$S \rightarrow \text{return } E ;$

$E \rightarrow \text{id } (A)$

$A \rightarrow \epsilon \mid E , A$

- Nonterminals **D** and **T** generate **function definition and types**, respectively
- A **function definition generated by D** consists of **keyword define**, a return type, the function name, formal parameters in parentheses and a function body consisting of a statement.
- **Nonterminal F generates zero or more formal parameters**, where a formal parameter consists of a type followed by an identifier.
- **Nonterminals S and E generate statements and expressions**, respectively. The production for **S adds a statement that returns** the value of an expression.
- The **production for E adds function calls**, with actual parameters generated by A. An actual parameter is an expression,

Intermediate Code for Procedures

`n = f(a[i]);`

following three-address code:

- 1) `t1 = i * 4`
- 2) `t2 = a [t1]`
- 3) `param t2`
- 4) `t3 = call f, 1`
- 5) `n = t3`

Function calls.

- When generating three-address instructions for a function call `id(E, E, ..., E)`,
- It is sufficient to generate the **three-address instructions for evaluating or reducing the parameters E** to addresses (`E.addr`),
- Followed by a **param instruction** for each parameter.

Intermediate Code for Procedures

`n = f(a[i]);`

following three-address code:

- 1) `t1 = i * 4`
- 2) `t2 = a [t1]`
- 3) `param t2`
- 4) `t3 = call f, 1`
- 5) `n = t3`

Symbol tables.

Let **s** be the top symbol table when the function definition is reached.

The function name is entered into **s** for use in the rest of the program.

The formal parameters of a function can be handled in analogy with field names in a record

In the production for D, after seeing define and the function name, we push s and set up a new symbol table

`Env.push(top)]`

`t = new Env();`

The new symbol table, **t**.

The new table **t** is used to translate the function body.

We revert to the previous symbol table **s** after the function body is translated.