

# Translation

- So far we have talked about parsing<sup>a</sup> of a language. But our main goal is translation.
- Semantic actions to translate the source language program to a target language program often go hand-in-hand with parsing. It is called syntax-directed translation.

<sup>&</sup>lt;sup>a</sup>Even the parsing cannot be completed without some analysis beyond the context-free structure.

# Translation

- To perform semantic actions along with parsing actions (e.g. reduction), we associate computation with the production rules.
  Computed information is propagated as attributes of non-terminals.
- An alternative is to be built the parse tree explicitly, and perform semantic actions by traversing the tree.

# Example

Consider the following production rule of the classic expression grammar:  $E \rightarrow E_1 + T^a$ . We consider three different translations:

- implementation of a simple calculator,
- conversion of an infix expression to a postfix expression,
- general purpose code generation.

<sup>a</sup>We have used subscript to differentiate between two instances of E.

#### Example: Calculator

- We have already seen that the only attribute of E and T required for this translation are values expressions corresponding to the sub-trees of E and T.
- Let us call the attribute to be val.
- The semantic action associated with the given production rule is,

$$E \to E_1 + T \{ E \cdot \operatorname{val} = E_1 \cdot \operatorname{val} + T \cdot \operatorname{val} \}^{\mathrm{a}}.$$

<sup>a</sup>In bison this gets translated to \$ = \$1 + \$3.



- The action may take place when  $E_1 + T$  is reduced to E. The computed value is saved as the attribute of E.
- Alternatively, it may take place during the postorder traversal of the syntax tree.
- There is no other side-effect of the semantic action.



- But if we want to keep a provision to store a value as a named object (variable), we need a symbol table where the variables names and their values are stored.
- In that case the semantic action of
   ES → id := E will changes the state of the
   symbol-table data structure (side effect) by
   entering the E·val corresponding to id.name.





- Here the problem is to convert an infix arithmetic expression to an equivalent postfix expression.
- Both the input and output are strings of characters.
- Let the attribute associated to each non-terminal be exp of type char \*. The semantic action is the following.



## Example: Code Generation

- The main difference of translation for code generation with two previous translations is that there is no data value corresponding to E<sub>1</sub> and T available at compilation time.
- Both  $E_1$  and T corresponds to two sequences of translated codes that will compute values of expressions corresponding to  $E_1$  and Twhen they are executed.

## Example: Code Generation

- The translation for to the rule  $E \rightarrow E_1 + T$ generates code so that the computed values of  $E_1$  and T are added to generate and store the value of expression for E.
- The computed values of the expressions  $E_1$ and T are stored in compiler defined temporary locations.

## Example: Code Generation

- Compiler creates temporary variables<sup>a</sup> where the intermediate values of sub-expressions are stored. These variable names are also entered in the symbol table.
- The main attribute of a non-terminal like *E* or *T* is the index of the symbol table corresponding to its temporary variable name.

<sup>&</sup>lt;sup>a</sup>They are also called virtual registers.





- This action has side-effects, it makes an entry of the new location in the symbol table. And the generated code is added in a data structure.
- As an alternative *E* and *T* may store their code sequences as their attributes.

## Associating Information

- Information is associated to syntactic categories by attaching attributes to the corresponding non-terminals.
- Computation of these attributes are associated with the production rules.
- Initial attribute values are supplied by the scanner.

# Definition

- A syntax-directed definition is a context-free grammar where attributes are associated with the grammar symbols. Rules for computing the attributes are associated with the production rules. There should not be any circularity in the definition.
- These are called attribute grammars when the definition does not have any side-effect.



- A syntax-directed translation is an executable specification of SDD. Fragments of programs are associated to different points in the production rules.
- The order of execution of the code is important in this case.



 $A \rightarrow \{Action_1\} B \{Action_2\} C \{Action_3\}$ <u>Action\_1</u>: takes place before parsing of the input corresponding to the non-terminal B. <u>Action\_2</u>: takes place after consuming the input for B, but before consuming the input for C. <u>Action\_3</u>: takes place at the time of reduction of BC to A or after consuming the input corresponding to BC.



- Embedded action may create some problem in a parser generator like Bison.
- Bison replaces the embedded action in a production rule by an ε-production and associates the embedded action with the new rule.



- But this may change the nature of the grammar. As an example, the grammar
   S → A | B, A → aba, B → abb is LALR.
- An embedded action is introduced as shown,  $S \rightarrow A \mid B, A \rightarrow a \{action\} ba, B \rightarrow abb.$





# A Simple Example

Consider the following grammar of signed binary numerals. We wish to translate it to decimal number. Compiler Design

 $0: S' \rightarrow N$  $1: N \rightarrow SL$  $2: S \rightarrow +$  $3: S \rightarrow 4: L \rightarrow LB$  $5: L \rightarrow B$  $6: B \rightarrow 0$  $7: B \rightarrow 1$ 



- We first construct the LR(0) automaton of the grammar and find that the grammar is SLR.
- We associate attributes to the non-terminals.
- We also associate SDDs to the production rules.





# SLR Parsing Table

S	Action				Goto				
	+		0	1	\$	N	S	L	B
0	$s_3$	$s_4$				1	2		
1					Acc				
2			$S_7$	$s_8$				5	6
3			$r_6$	$r_6$					
4			$r_7$	$r_7$					
5			$S_7$	$s_8$	$r_1$				9

# SLR Parsing Table

S	Action				Ga	oto			
	+	_	0	1	\$	N	S	L	В
6			$r_5$	$r_5$	$r_5$				
7			$r_6$	$r_6$	$r_6$				
8			$r_7$	$r_7$	$r_7$				

#### Attributes of Non-Terminals

Following are the attributes of different non-terminals:

Non-terminal	Attribute	Type
N	val	int
S	sign	char
L	val	int
В	val	int

SDI  $0: S' \rightarrow N$  print N.val 1:  $N \rightarrow SL$  if (S.sign == '-') N.val= - L.val; else N.val = L.val; 2:  $S \rightarrow +$  S.sign = '+';  $3: S \rightarrow -$  S.sign = '-'; 4:  $L \rightarrow L_1 B$  L.val = 2\*L1.val+B.val; 5:  $L \rightarrow B$  L.val = B.val;  $6: B \rightarrow 0$  B.val = 0; 7:  $B \rightarrow 1$  B.val = 1;

Туре	$\mathrm{Stack} \rightarrow$	Input $\rightarrow$	Action/Value
Parsing	\$0	+101\$	shift
Value	\$		
Parsing	\$03	101\$	reduce
Value	\$+		
Parsing	\$02	101\$	shift
Value	\$S		S.sign='+'
Parsing	\$028	01\$	reduce
Value	\$S1		

Type	$\mathrm{Stack} \rightarrow$	Input $\rightarrow$	Action/Value
Parsing	\$026	01\$	reduce
Value	\$SB		B.val = 1
Parsing	\$025	01\$	shift
Value	\$SL		L.val = B.val
Parsing	\$0257	1\$	reduce
Value	\$SLO		
Parsing	\$0259	1\$	reduce
Value	\$SLB		B.val = 0

Type	$\mathrm{Stack} \rightarrow$	$\mathrm{Input}^{\rightarrow}$	Action/Value
Parsing	\$025	1\$	shift
Value	\$SL		L.val = 2*L1.val + B.val
Parsing	\$0258	\$	reduce
Value	\$SL1		
Parsing	\$0259	\$	reduce
Value	\$SLB		B.val=1
Parsing	\$025	\$	reduce
Value	\$SL		L.val = 2*L1.val + B.val

Goutam Biswas

Type	$\mathrm{Stack} \rightarrow$	Input $\rightarrow$	Action/Value
Parsing	\$01	\$	Accept
Value	\$N		N.val = +L.val


### Synthesized Attribute

- In this example the value of an attribute of a non-terminal is either coming from the scanner<sup>a</sup> or it is computed from the attributes of its children.
- This type of attribute is known as a synthesized attribute.

<sup>a</sup>Attribute of a terminal comes from the scanner.

### S-Attributed

- An attributed grammar is called S-attributed if every attribute is synthesized.
- Attributes in such a grammar can be easily computed during a bottom-up parsing.

### Another Set of Attributes

Non-terminal	Attribute	Type
N	val	int
S	sign	char
	val, pos	int, int
В	val, pos	int, int





5:  $L \rightarrow B$  B.pos = L.pos; L.val = B.val; 6:  $B \rightarrow 0$  B.val = 0; 7:  $B \rightarrow 1$  B.val =  $2^{B.pos}$ ;





- Attributes of a non-terminal depends on the nature of translation. But it may also depend on the nature of the grammar.
- Following is a grammar of integers in 2's complement numerals. It is to be translated to a signed decimal numeral.



Associate appropriate attributes to the non-terminals and give rules of semantic actions. Write bison specification for the grammar.



Compiler Design

 $0: S' \to N$  $1: N \rightarrow SL$  $2: S \rightarrow +$  $3: S \rightarrow 4: L \rightarrow BL$  $5: L \rightarrow B$  $6: B \rightarrow 0$  $7: B \rightarrow 1$ 

### Attributes of Non-Terminals

We need a new attribute of L to remember the bit position:

Non-terminal	Attribute	Type
N	val	int
S	$\operatorname{sign}$	char
L	val	int
	$\operatorname{pos}$	int
B	val	int

Action for Rules

$$0: S' \to N$$
 print N.val

1:  $N \rightarrow SL$  if (S.sign == '-') N.val=- L.val;

else N.val = L.val;

- $2: S \rightarrow + S.sign = '+';$
- $3: S \rightarrow -$  S.sign = '-';

4:  $L \rightarrow BL_1$  if(B.val)

L.val=pow(2,L1.pos)+L1.val;

```
else L.val=L1.val;
```

L.pos=L1.pos+1;



5: 
$$L \rightarrow B$$
 L.val = B.val; L.pos = 1  
6:  $B \rightarrow 0$  B.val = 0;

7: 
$$B \rightarrow 1$$
 B.val = 1;

# Example

Consider the following grammar for variable declaration:









When an id is reduced to the non-terminal L, it is inserted in the symbol table along with its type information<sup>a</sup>. The type information is not available from any subtree rooted at L. It has to be inherited from T via the root D.

<sup>a</sup>The type information is important for space allocation, representation, operations, correctness and other purposes.

### SDDefinition

- 1:  $D \rightarrow TL$ ; L.type = T.type
- $2: T \rightarrow int$  T.type = INT
- $3: T \rightarrow \text{double} \text{ T.type} = \text{DOUBLE}$
- $4: L \rightarrow L_1, ext{id}$  L1.type = L.type
  - addSym(id.name, L.type)
- 5:  $L \rightarrow id$  addSym(id.name, L.type)

## Inherited Attribute

- Let B be a non-terminal of a parse tree node.
- An inherited attribute *B.i* is defined in terms of the attributes of the parent and sibling nodes of *B*.
- In the previous example the non-terminal L gets the attribute from T as an inherited attribute.



- The synthesized attribute *B.s* of a non-terminal *B* is defined by the attributes of its children.
- The attribute of a leaf-node comes from the scanner.

### S-Attributed Definitions

An SDD is S-attributed if every attribute is synthesized. This may be called an S-attributed

#### grammar.

This definition can be implemented in a LR-parser during a reduction as the traversal on the parse-tree is postorder.

### *L*-Attributed Definitions

An SDD is called *L*-attributed ('L' for left) if each attribute is either synthesized, or inherited with the following restrictions. Let  $A \to \alpha_1 \alpha_2 \cdots \alpha_n$  be a production rule, and  $\alpha_k$  has an inherited attribute 'a'.

### L-Attributed Definition

The value of  $\alpha_k a$  is computed using

- the inherited attribute of A (parent),
- the inherited or synthesized attributes of  $\alpha_1, \alpha_2, \cdots, \alpha_{k-1}$  (symbols to the left of  $\alpha_k$ ),
- attributes of  $\alpha_k$ , provided no dependency cycle<sup>a</sup> is formed.

<sup>a</sup>
$$A \rightarrow B \{ A.s = B.i; B.i = A.s + k \}.$$

Rules
-------

The type definition mentioned earlier is L-attributed.

 $1: D \rightarrow TL;$ L.type = T.type $2: T \rightarrow int$ T.type = INT $3: T \rightarrow double$ T.type = DOUBLE $4: L \rightarrow L_1, id$ L1.type = L.typeaddSym(id.name, L.type) $5: L \rightarrow id$ addSym(id.name, L.type)



The question is how to propagate the type information in a parser generated by bison The non-terminal T gets the value of synthesized type attribute when a T-production rule is reduced. But that cannot be propagated as an attribute of the non-terminal L directly as this non-terminal is not present in the stack.



# Solution I

An ad hoc solution is to use a global variable to hold the type value.

 $T \rightarrow int$  type = INT  $T \rightarrow double$  type = DOUBLE  $L \rightarrow L_1, id$  addSym(id.name, type)  $L \rightarrow id$  addSym(id.name, type)



We introduce a different attribute of L, a list of symbol table entries corresponding to different identifiers, and initialize their types at the end.

Compiler Design

1:  $D \rightarrow TL$ ; initType(L.list, T.type)  $2: T \rightarrow int$  T.type = INT  $3: T \rightarrow \text{double T.type} = \text{DOUBLE}$ 4:  $L \rightarrow L_1$ , id L.list = L\_1.list + mklist(addSym(id.name))  $5: L \rightarrow id$ L.list =mklist(addSym(id.name)) Read + as append in the list.

Lect 9

# Solution III

We can device another solution from the value stack. For that we consider the states of LR(0) automaton of the grammar.

# LR(0) Automaton



### Example: Parsing & Value Stack

Type	$\mathrm{Stack} \rightarrow$	$Input \rightarrow$	Action/Value
Par	<b>\$</b> 0	<pre>int id, id;\$</pre>	$\operatorname{shift}$
Val	\$		
Par	\$03	id, id;\$	reduce
Val	\$int		
Par	\$02	id, id;\$	$\operatorname{shift}$
Val	\$T		T.type=INT
Par	\$026	, id;\$	reduce
Val	\$T id		

Example: Parsing &	Value Stack
--------------------	-------------

Type	$\mathrm{Stack} \rightarrow$	$\mathrm{Input} \rightarrow$	Action/Value
Par	\$025	, id;\$	reduce
Val	\$T L		addSym(id.name,L.type)

How does L gets the type information. Note that in bison  $L \equiv$ \$ and id  $\equiv$ \$1. But the type information is available in T in the stack, below the handle.

Type	$\mathrm{Stack} \rightarrow$	$\mathrm{Input} {\rightarrow}$	Action/Value
Par	\$0257	id;\$	shift
Val	\$T L ,		

Goutam Biswas

### Example: Parsing & Value Stack

Type	$\mathrm{Stack} \rightarrow$	$\mathrm{Input} \rightarrow$	Action/Value
Par	\$02578	;\$	reduce
Val	\$T L , id		
Par	\$025	;\$	
Val	\$T L		addSym(id.name,L.type)

Again the type information is available just below the handle.



In Bison the attribute below the handle can be accessed. In this case the non-terminal T corresponds to \$0 and its type attribute is \$0.type.



- Often a natural grammar is transformed to make it suitable for parsing.
- But the new parse tree no longer match with the abstract syntax tree of the language.
- As an example the left-recursion is removed from the grammar for LL(1) parsing.
- The original S-attributed grammar gets modified after this transformation.
## S-Attributed Expression Grammar

- $S \rightarrow E$  { print E.val }
- $E \rightarrow E_1 + T$  { E.val = E1.val + T.val}
- $E \rightarrow T$  { E.val = T.val}
- $T \rightarrow T_1 * F$  { T.val = T1.val \* F.val}
- $T \rightarrow F$  { T.val = F.val}
- $F \rightarrow (F) \{ F.val = E.val \}$
- $F \rightarrow ic$  { F.val = ic.val}

Compiler Design









# Note

- Two arguments of '+' are in different subtrees. It is necessary to pass the value of T<sub>3</sub>.s to the subtree of E'<sub>1</sub>.
- It is also necessary for left-associativity of '+', to propagate the computed value down the tree say from E'<sub>1</sub> to E'<sub>2</sub>.
- We achieve this by inherited attributes E'.iand T'.i of the non-terminals E' and T'.



But it is also necessary to propagate the computed value towards the root. This is done through the synthesized attributes of E' and T' i.e. E'.s, T'.s.



$$\begin{array}{rcl} E & \rightarrow & T \left\{ \text{ E'.ival = T.sval } \right\} E' \\ & \left\{ \text{ E.sval = E'.sval } \right\} \\ E' & \rightarrow & +T \left\{ \text{ E1'.ival = E'.ival + T.sval } \right\} E'_1 \\ & \left\{ \text{ E'.sval = E1'.sval } \right\} \\ E' & \rightarrow & \varepsilon \left\{ \text{ E'.sval = E'.ival } \right\} \end{array}$$



$$T \rightarrow F \{ \text{ T'.ival} = \text{F.sval} \} T'$$

$$\{ \text{ T.sval} = \text{T'.sval} \}$$

$$T' \rightarrow *F \{ \text{ T1'.ival} = \text{T'.ival} * \text{F.sval} \} T'_1$$

$$\{ \text{ T'.sval} = \text{T1'.sval} \}$$

$$T' \rightarrow \varepsilon \{ \text{ T'.sval} = \text{T'.ival} \}$$





### Another Example

L-attributed grammars come naturally with flow-control statements. Following is an example with if-then-else statement.

IS  $\rightarrow$  if BE then  $S_1$  else  $S_2$ .

### Attributes of Statement

- Every statement has a natural synthesized attribute, S.code, holding the code corresponding to S.
- Also a statement S has a continuation, the next instruction to be executed after execution of S. This may be handled as a jump target (label). But this label is an inherited attribute of S, S.next, propagated in the subtree of S.



- The boolean expression also has a synthesized attribute BE.code.
- But it has two inherited attributes, BE.true, a jump target (label) where the control is transferred if the boolean expression is evaluated to true. This is the beginning of  $S_1$ .

Similarly there is BE.false, a label at the beginning of  $S_2$ .

#### SDD for if-then-else

$$\begin{split} \text{IS} &\to \text{if BE} \quad |1=\text{newLabel}(), 12=\text{newLabel}() \\ & \texttt{then } S_1 \quad \text{BE.true} = 11, \text{ BE.false}=12 \\ & \texttt{else } S_2. \quad \text{S}_1.\text{next} = \text{S}_2.\text{next} = \text{IS.next} \\ & \text{IS.code} = \text{BE.code} + 11'\text{i'} + \\ & \text{S}_1.\text{code} + 12'\text{i'} + \text{S}_2.\text{code} \end{split}$$





Afterward we shall see how this is managed in an actual implementation using back-patching.





