

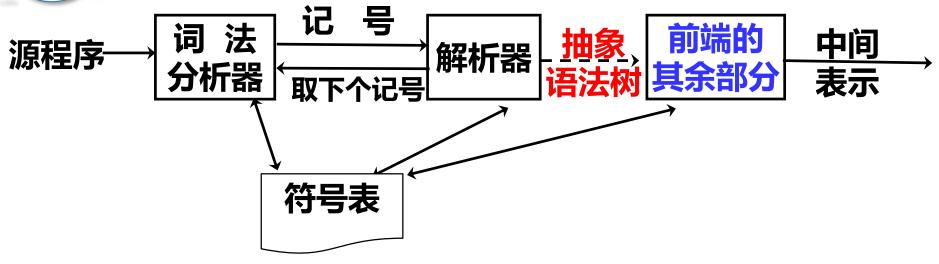
# 语法制导的翻译I

《编译原理和技术》

#### 张昱

0551-63603804, yuzhang@ustc.edu.cn 中国科学技术大学 计算机科学与技术学院

# 本章内容



- □ 语义的描述: 语法制导的定义、翻译方案
  - 语法制导: syntax-directed 按语法结构来指导语义的定义和计算
  - 抽象语法树、注释分析树等
- □ 语法制导翻译的实现方法: 自上而下、自下而上
  - 边语法分析边翻译

# 本草内容

□ 语义描述举例:输出表达式的后缀形式

中缀式→后缀式

■ 语法制导的定义(syntax-directed definition)

$$E \rightarrow E_1 + T$$

 $E.code = E_1.code \parallel T.code \parallel '+'$ 

E或T的code属性表示E或T对应的后缀形式

语法制导定义可读性好, 更适于描述语义规范

■ 翻译方案(translation scheme)

$$E \rightarrow E_1 + T$$

{ print '+' }

翻译方案陈述了实现细节(如语义规则的计算时机)

- □ 语法制导翻译技术可以用于
  - 构建抽象语法树、语义分析、中间代码生成等



## 4.1 语法制导的定义

- □语法制导的定义
- □综合属性、继承属性
- □属性依赖图与属性的计算次序



## 语法制导的定义

#### □ 语法制导的定义(Syntax-Directed Definition)

- ■基础的上下文无关文法
- ■每个文法符号有一组属性 对应的语义信息
- 每个文法产生式 $A \rightarrow \alpha$ 有一组形式为 $b = f(c_1, c_2, ..., c_k)$ 的 语义规则 描述语义属性值的计算规则 其中f 是函数, $b \rightarrow c_1, c_2, ..., c_k$  是该产生式文法符号的属性

用来表示语法成分

- b是综合属性(synthesized attribute): 如果b是A的属性,  $c_1, c_2, ..., c_k$  是产生式右部文法符号的属性或A的其它属性
- b是继承属性(inherited attribute): 如果b是右部某文法符号X的属性



下面是产生字母表 $\Sigma = \{0, 1, 2\}$ 上数字串的一个文法:

$$S \rightarrow D S D \mid 2$$

$$D \rightarrow 0 \mid 1$$

写一个语法制导定义, 判断它接受的句子是否为回文数

以区分同类语法

、结构的不同实例

$$S \rightarrow D_1 S_1 D_2$$

 $S.val = (D_1.val) = D_2.val$  and  $S_1.val$ 

$$S \rightarrow 2$$

 $S' \rightarrow S$ 

$$S.val = true$$

print(S.val)

$$D \rightarrow 0$$

$$\mathbf{D}.val = \mathbf{0}$$

S.val: S对应的串是否是回文数

$$D \rightarrow 1$$

$$\mathbf{D}.val = \mathbf{1}$$

D.val: D对应的串的数值

#### 各文法符号的属性均是综合属性的语法制导定义——S 属性定义



## 简单计算器的语法制导定义

产生式	语义规则	L的匿名 属性
$L \rightarrow E$ n 换行标记	print (E.val)	
$E \rightarrow E_1 + T$	$E.val = E_1.val + T.val$	
$E \rightarrow T$	E.val = T.val	
$T \rightarrow T_1 * F$	$T.val = T_1.val * F.val$	
$T \rightarrow F$	T.val = F.val	
$F \rightarrow (E)$	F.val = E.val	由词法
$F \rightarrow \text{digit}$	F.val = digit.lexval	分析给出

#### 各文法符号的属性均是**综合属性**的语法制导定义——S 属性定义

参见: <u>bison-examples.tar.gz</u> 中的config/expr1.y, expr.lex



### 注释分析树(annotated parse tree)

#### □ 结点的属性值都标注出来的分析树

8+5\*2 n (n为换行符)的注释分析树

E.val = 18E.val = 8T.val = 10T.val = 8T.val = 5digit.lexval = 2F.val = 5F.val = 8digit.lexval = 5digit.lexval = 8

结点的综合属性值 可由子结点的属性 值计算得到

F.val = 2



## S属性定义的局限

#### □ 考虑消除左递归后,表达式语言的 LL文法

产生式	语 义 规 则
$T \rightarrow T_1 * F$	$T.val = T_1.val * F.val$
$T \rightarrow F$	T.val = F.val
$F \rightarrow \text{digit}$	F.val = digit.lexval

消除 左递归

产生式	语义规则
$T \rightarrow FW$	T.val = F.val?
$W \rightarrow *FW_1$	W.val = ? *F.val
$W \rightarrow \varepsilon$	
$F \rightarrow \text{digit}$	F.val = digit.lexval

- T 对应的项中,第1个运算分量是F,而运算符\*和其第2个运算分量在W的子结构中
- $W \rightarrow * FW_1$ : 在分析W期间,\*的左运算分量不在W的子结构中



## 左递归的消除引起继承属性

- 继承属性(inherited attribute): 如果b是右部某文法符号X的属性,  $c_1,c_2,...,c_k$ 是产生式右部文法符号的属性或A的其它属性
- 为W引入继承属性i,继承运算符\*的左运算分量的属性为W引入综合属性s,它表示最终的计算结果

继承W左边的文法符号F的属性

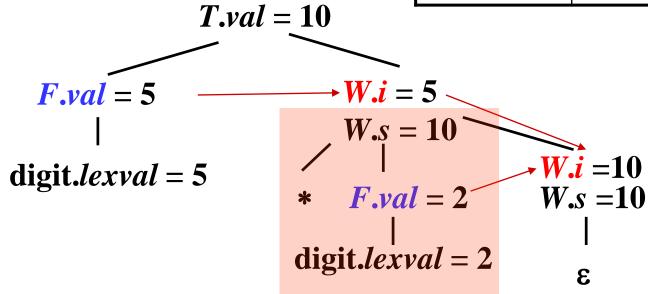
产生式	语义规则
$T \rightarrow FW$	W.i = F.val ; T.val = W.s
$W \rightarrow *FW_1$	$W_{1}.i = W.i * F.val; W.s = W_{1}.s$
$W \rightarrow \varepsilon$	W.s = W.i
•••	•••



## 含继承属性的注释分析树

- □ 5\*2 n (n为换行符)的注释分析树
- □ 不能自下而上由子结 点的属性值计算出父 结点的属性值

产生式	语 义 规 则
$T \rightarrow FW$	W.i = F.val ; T.val = W.s
$W \rightarrow *FW_1$	$W_1.i = W.i * F.val; W.s = W_1.s$
$W \rightarrow \varepsilon$	W.s = W.i
	•••



张昱:《编译原理和技术》课程信息



#### □ 属性依赖图

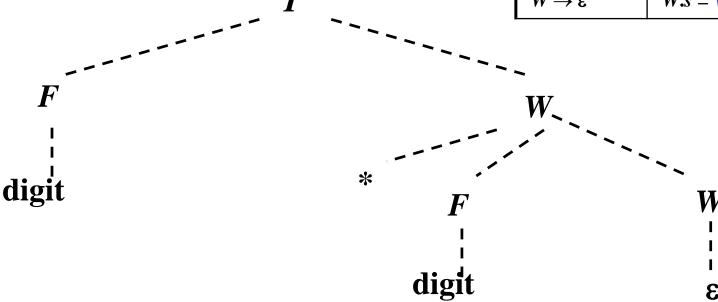
- 描述分析树中结点的属性间依赖关系的有向图
- 顶点为属性:对应分析树中每个文法符号X的每个属性a
- 弧为属性间依赖关系:如果属性X.a的值依赖于属性Y.b的值,则存在从Y.b的顶点指向X.a的顶点的弧



- □ 属性依赖图
  - 描述分析树中结点的属性间依赖关系的有向图
- □ 5\*2 n (n为换行符)

分析树(虚线)

产生式	语 义 规 则
$T \rightarrow FW$	W.i = F.val; $T.val = W.s$
$W \rightarrow *FW_1$	$W_{1}.i = W.i * F.val; W.s = W_{1}.s$
$W \rightarrow \varepsilon$	W.s = W.i





- □ 属性依赖图
  - 描述分析树中结点的属性间依赖关系的有向图
- □ 5\*2 n (n为换行符)

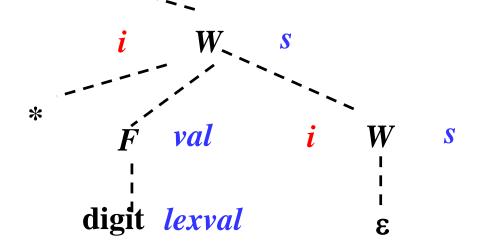
分析树(虚线)的依赖图

T val

产生式	语 义 规 则
$T \rightarrow FW$	W.i = F.val; $T.val = W.s$
$W \rightarrow *FW_1$	$W_{1}.i = W.i * F.val; W.s = W_{1}.s$
$W \rightarrow \varepsilon$	W.s = W.i

F val
digit lexval

顶点:属性





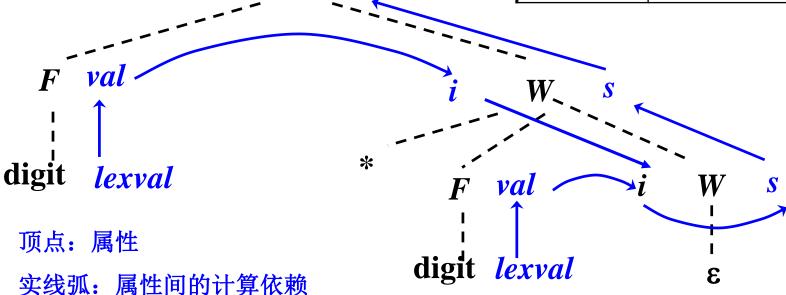
- □ 属性依赖图
  - 描述分析树中结点的属性间依赖关系的有向图

val

□ 5\*2 n (n为换行符)

分析树(虚线)的依赖图

产生式	语 义 规 则
$T \rightarrow FW$	W.i = F.val; $T.val = W.s$
$W \rightarrow *FW_1$	$W_{1}.i = W.i * F.val; W.s = W_{1}.s$
$W \rightarrow \varepsilon$	W.s = W.i



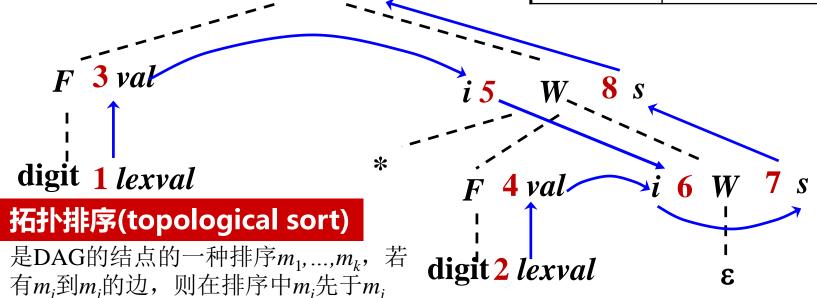


- □ 属性依赖图
  - 描述分析树中结点的属性间依赖关系的有向图
- □ 5\*2 n (n为换行符)

分析树(虚线)的依赖图(实线)

T 9 val

产生式	语义规则
$T \rightarrow FW$	W.i = F.val; $T.val = W.s$
$W \rightarrow *FW_1$	$W_1.i = W.i * F.val; W.s = W_1.s$
$W \rightarrow \varepsilon$	W.s = W.i





## 属性计算次序

- 属性计算次序
- 1) 构造输入的分析树
- 2) 构造属性依赖图
- 3) 对结点进行拓扑排序
- 4) 按拓扑排序的次序计算属性



## S 属性定义和 L 属性定义

□ S属性定义

仅使用综合属性的语法制导定义

□ L属性定义 (属性信息自左向右流动)

如果每个产生式 $A \rightarrow X_1 ... X_{j-1} X_j ... X_n$ 的每条语义规则计算的属性是A的综合属性,或者是 $X_i$ 的继承属性,但它仅依赖:

- 该产生式中 $X_j$  左边符号 $X_1, X_2, ..., X_{j-1}$ 的属性;
- A的继承属性

可以按边分析边翻译的方式计算继承属性

□ S属性定义是L属性定义



## 语义规则的计算方法

□ 分析树方法

刚才介绍的方法, 动态确定计算次序, 效率低

——概念上的一般方法

□基于规则的方法

(编译器实现者)静态确定(语言设计者提供的)语义规则的计算次序——适用于手工构造的方法

□忽略规则的方法

(编译器实现者)事先确定属性的计算策略(如边分析边计算),(语言设计者提供的)语义规则必须符合所选分析方法的限制——适用于自动生成的方法



## L 属性定义的另一个例子

int id, id, id

产生式	语 义 规 则
$D \rightarrow TL$	L.in = T.type
$T \rightarrow int$	T. type = integer
$T \rightarrow \text{real}$	T. type = real
$L \rightarrow L_1$ , id	$L_1.in = L.in;$
	addType(id.entry, L.in)
$L \rightarrow id$	addType(id.entry, L.in)

type-T的综合属性

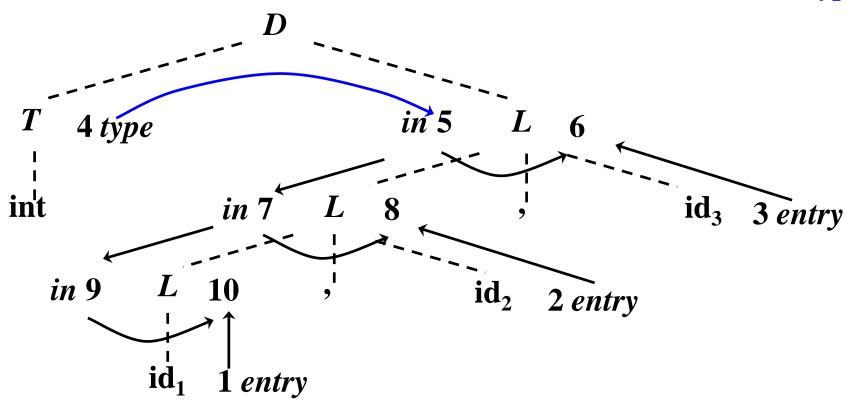
in-L的继承属性,把声明的类型传递给标识符列表 addType-把类型信息加到符号表中的标识符条目里



int id<sub>1</sub>, id<sub>2</sub>, id<sub>3</sub>

分析树(虚线)的依赖图(实线)

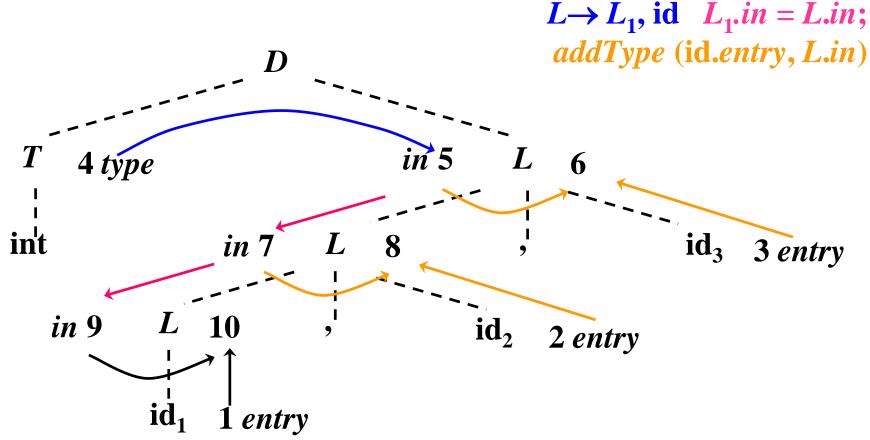
$$D \rightarrow TL$$
  $L.in = T.type$ 





int id<sub>1</sub>, id<sub>2</sub>, id<sub>3</sub>

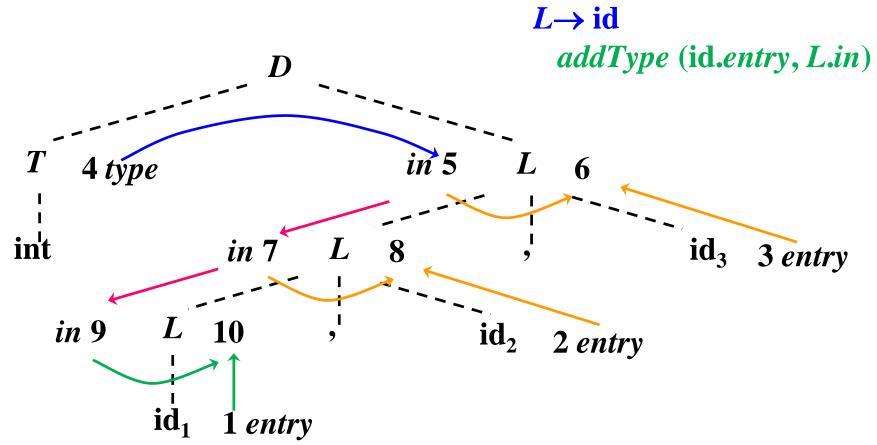
分析树(虚线)的依赖图(实线)





int id<sub>1</sub>, id<sub>2</sub>, id<sub>3</sub>

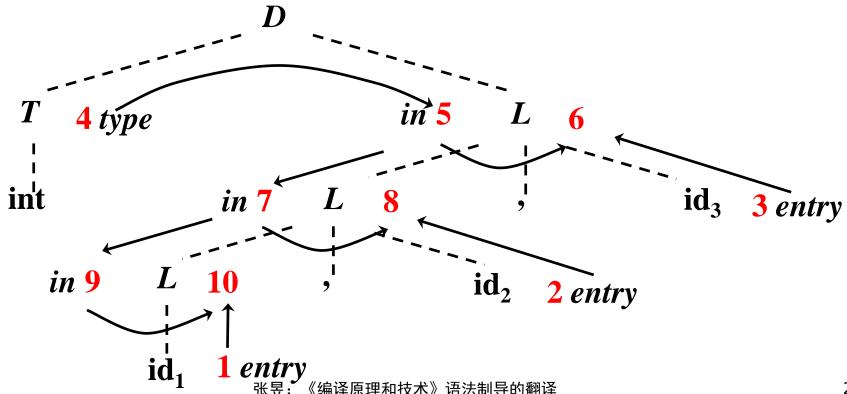
分析树(虚线)的依赖图(实线)

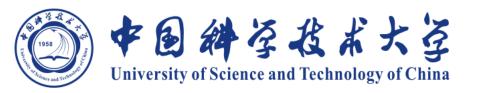




## 属性计算次序

■ 拓扑排序(topological sort): 是DAG的结点的一种排序 m<sub>1</sub>,...,m<sub>k</sub>, 若有m<sub>i</sub>到m<sub>j</sub>的边,则在排序中m<sub>i</sub>先于m<sub>j</sub>
 例 1, 2, 3, 4, 5, 6, 7, 8, 9, 10





## 4.2 语法树及其构造

- □语法树
- □ 语法树的构造 (文法对构造的影响)
  - 语法制导定义 vs. 翻译方案
  - 自上而下计算 vs. 自下而上计算



## 语法树(syntax tree)

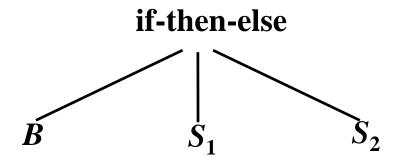
#### □ 语法树是分析树的浓缩表示

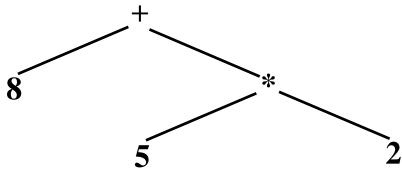
每个结点表示一个语法构造, 算符和关键字是语法树中的内部结点

举例:

if B then  $S_1$  else  $S_2$ 

$$8 + 5 * 2$$





语法制导翻译可以基于分析树,也可以基于语法树



## 语法制导定义: 构造语法树

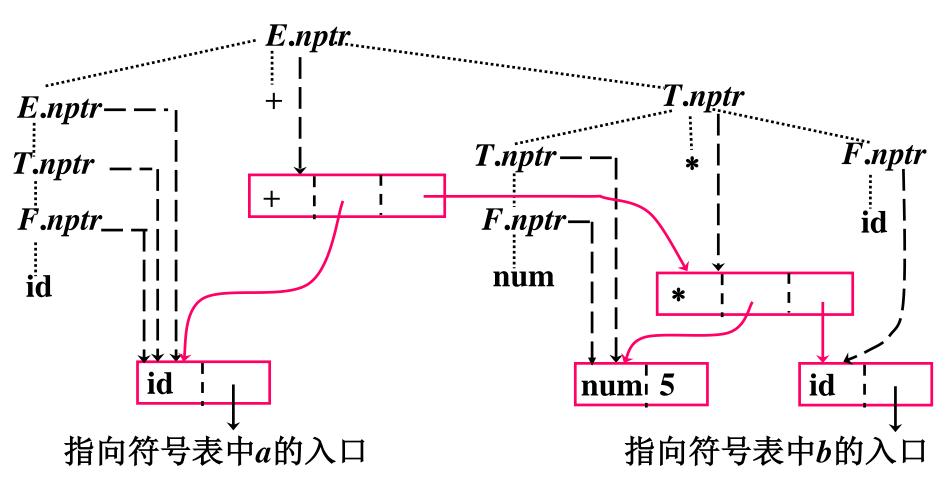
产生式	语 义 规 则
$E \rightarrow E_1 + T$	$E.nptr = mkNode('+', E_1.nptr, T.nptr)$
$E \rightarrow T$	E.nptr = T.nptr
$T \rightarrow T_1 * F$	$T.nptr = mkNode(`*, T_1.nptr, F.nptr)$
$T \rightarrow F$	T.nptr = F.nptr
$F \rightarrow (E)$	F.nptr = E.nptr
$F  ightarrow \mathrm{id}$	F.nptr = mkLeaf (id, id.entry)
$F \rightarrow \text{num}$	F.nptr = mkLeaf (num, num.val)

参见: <u>bison-examples.tar.gz</u> 中的config/asgn2ast.y, asgn.lex



## 语法树的构造

#### a+5\*b的语法树的构造



## 制泽方案 The state of the state o

### □ 构造语法树的翻译方案 (左递归文法)

语义动作, 置于花括号中

```
E \rightarrow E_1 + T
                   {E.nptr = mkNode(`+', E_1.nptr, T.nptr)}
E \rightarrow T
                   \{E.nptr = T.nptr\}
T \rightarrow T_1 * F
                   \{T.nptr = mkNode(`*, T_1.nptr, F.nptr)\}
T \rightarrow F
                   \{T.nptr = F.nptr\}
F \rightarrow (E)
                   \{F.nptr = E.nptr\}
                   \{F.nptr = mkLeaf (id, id.entry) \}
F \rightarrow id
                   \{F.nptr = mkLeaf (num, num.val) \}
F \rightarrow \text{num}
```

综合属性的计算规则置于产生式右部的右边,表示识别出右部后计算



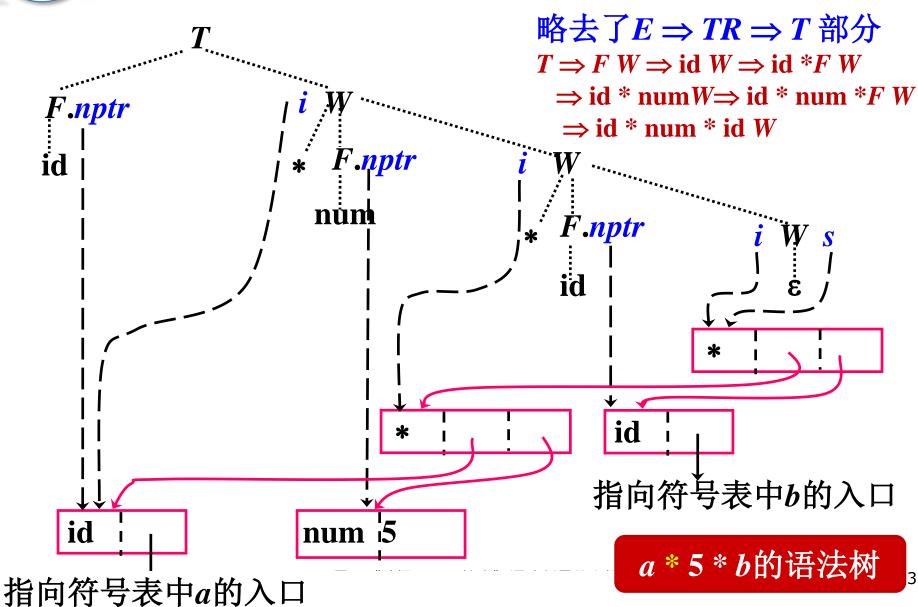
## 左递归的消除引起继承属性

#### 表达式语言的 LL文法

产生式	语 义 规 则
$E \rightarrow T R$	R.i = T.nptr ; E.nptr = R.s
$R \rightarrow + TR_1$	$R_{1}.i = mkNode ('+', R.i, T.nptr); R.s = R_{1}.s$
$R \rightarrow \varepsilon$	R.s = R.i
$T \rightarrow FW$	W.i = F.nptr ; T.nptr = W.s
$W \rightarrow *FW_1$	$W_{1}.i = mkNode ('+', W.i, F.nptr); W.s = W_{1}.s$
$W \rightarrow \varepsilon$	W.s = W.i
•••	•••



## 语法树的构造 (LL文法)



# **翻译方案**

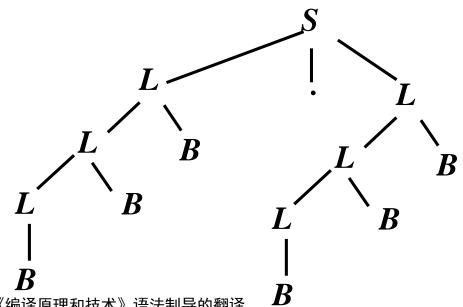
```
E \rightarrow T
                                              T + T + T + \dots
               \{R.i = T.nptr\}
               \{E.nptr = R.s\}
       R
R \rightarrow +
               \{R_1.i = mkNode ('+', R.i, T.nptr)\}
               \{R.s = R_1.s\}
                                         继承属性的计算嵌在产
               \{R.s = R.i\}
R \rightarrow \epsilon
                                         生式右部的某文法符号
T \rightarrow F
               \{W.i = F.nptr\}
                                         之前,表示在分析该文
               \{T.nptr = W.s\}
       \boldsymbol{W}
                                         法符号之前计算
W \rightarrow *
               \{W_1.i = mkNode (`*, W.i, F.nptr)\}
       \boldsymbol{F}
       W_1 \quad \{W_s = W_{1} \cdot s\}
               \{W.s = W.i\}
W \rightarrow \epsilon
   F 的产生式部分不再给出
```

为下面文法写一个语法制导的定义,用S的综合属性val 给 出下面文法中S产生的二进制数的值。

例如, 输入101.101时, S.val = 5.625 (可以修改文法)

若按 $2^2+0+2^0+2^{-1}+0+2^{-3}$ 来计算,该文法对小数点 左边部分的计算不利, 因为需要继承属性来确定每个B 离开小数点的距离

$$S \rightarrow L \cdot L \mid L$$
  
 $L \rightarrow L \mid B \mid B$   
 $B \rightarrow 0 \mid 1$ 

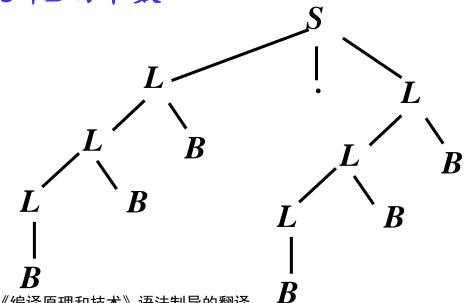


为下面文法写一个语法制导的定义,用S的综合属性val给出下面文法中S产生的二进制数的值。

例如,输入101.101时,S.val = 5.625(可以修改文法)

若小数点左边按(1×2+0)×2+1计算。该办法不能直接用于小数点右边,需改成((1×2+0)×2+1)/2³, 这时需要综合属性来统计B的个数

$$S \rightarrow L \cdot L \mid L$$
  
 $L \rightarrow L \mid B \mid B$   
 $B \rightarrow 0 \mid 1$ 



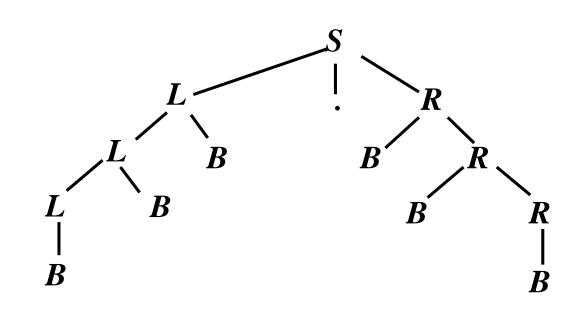


为下面文法写一个语法制导的定义,用S的综合属性val 给出下面文法中S产生的二进制数的值。

例如,输入101.101时,S.val = 5.625 (可以修改文法)

更清楚的办法是将文法改成下面的形式

$$S \rightarrow L \cdot R \mid L$$
 $L \rightarrow L \mid B \mid B$ 
 $R \rightarrow B \mid R \mid B$ 
 $B \rightarrow 0 \mid 1$ 





为下面文法写一个语法制导的定义,用S的综合属性val给出下面文法中S产生的二进制数的值。

例如,输入101.101时, S.val = 5.625 (可以修改文法)

$$S \rightarrow L \cdot R$$

$$S. val = L. val + R. val$$

$$S \rightarrow L$$

$$S. val = L. val$$

$$L \rightarrow L_1 B$$

$$L. val = L_1. val \times 2 + B. val$$

$$L \rightarrow B$$

$$L. val = B. val$$

$$\mathbf{R} \to \mathbf{B} \; \mathbf{R}_1$$

$$R. val = R_1. val / 2 + B. val / 2$$

$$R \rightarrow B$$

$$\mathbf{R.} \ val = \mathbf{B.} \ val / 2$$

$$B \rightarrow 0$$

$$B. val = 0$$

$$B \rightarrow 1$$

$$B. val = 1$$



给出把中缀表达式翻译成没有冗余括号的中缀表达式的语 法制导定义。例如, 因为+和\*是左结合,

((a\*(b+c))\*(d))可以重写成a\*(b+c)\*d

#### 两种方法:

- 先把括号都去掉, 然后在必要的地方再加括号
- 去掉表达式中的冗余括号,保留必要的括号



#### □ 先把括号都去掉,然后在必要的地方再加括号

$$S' \rightarrow E \qquad print \ (E.\ code\ )$$

$$E \rightarrow E_1 + T$$

$$\text{if } T.\ op == plus \ \text{then}$$

$$E.\ code = E_1.\ code \parallel ``+" \parallel ``(" \parallel T.\ code \parallel ``)"$$

$$\text{else}$$

$$E.\ code = E_1.\ code \parallel ``+" \parallel T.\ code;$$

$$E.\ op = plus$$

$$E \rightarrow T \qquad E.\ code = T.\ code; \ E.\ op = T.\ op$$



#### □ 先把括号都去掉,然后在必要的地方再加括号

$$T \to T_1 * F$$
 if  $(F. op == plus)$  or  $(F. op == times)$  then if  $T_1. op == plus$  then  $T. code = "(" || T_1. code || ")" || "*" || "(" || F. code || ")"$  else 
$$T. code = T_1. code || "*" || "(" || F. code || ")"$$
 else if  $T_1. op = plus$  then 
$$T. code = "(" || T_1. code || ")" || "*" || F. code$$
 else 
$$T. code = T_1. code || "*" || F. code;$$
 
$$T. op = times$$



#### □ 先把括号都去掉,然后在必要的地方再加括号

```
T \rightarrow F
T. code = F. code; T. op = F. op
F \rightarrow id
F. code = id. lexeme; F. op = id
F \rightarrow (E)
F. code = E. code; F. op = E. op
```



#### □ 去掉表达式中的冗余括号,保留必要的括号

- 给E, T和F两个继承属性left\_op和right\_op分别表示左右两 侧算符的优先级
- 给它们一个综合属性self\_op表示自身主算符的优先级
- 再给一个综合属性code表示没有冗余括号的代码
- 分别用1和2表示加和乘的优先级,用3表示id和(E)的优先级,用0表示左侧或右侧没有运算对象的情况



$$S' \rightarrow E$$
 $E. \ left\_op = 0; E. \ right\_op = 0; print (E. \ code)$ 
 $E \rightarrow E_1 + T$ 
 $E_1. \ left\_op = E. \ left\_op; E_1. \ right\_op = 1;$ 
 $T. \ left\_op = 1; T. \ right\_op = E. \ right\_op;$ 
 $E. \ code = E_1. \ code || "+" || T. \ code ; E. \ self\_op = 1;$ 
 $E \rightarrow T$ 
 $T. \ left\_op = E. \ left\_op;$ 
 $T. \ right\_op = E. \ right\_op;$ 
 $E. \ code = T. \ code; E. \ self\_op = T. \ self\_op$ 



$$T \rightarrow T_1 * F$$
 ...
 $T \rightarrow F$  ...

$$F \rightarrow \mathrm{id}$$

 $F. code = id. lexeme; F. self_op = 3$ 



```
F \rightarrow (E)
   E. left\_op = 0; E. right\_op = 0;
   F. self_op =
    if (F. left\_op < E. self\_op) and
                             (E. self\_op >= F. right\_op)
             then E. self_op else 3
   F. code =
    if (F. left\_op < E. self\_op) and
                             (E. self\_op >= F. right\_op)
                   E. code else "(" || E. code || ")"
             then
```



### Lab: ParseTree =>AST

#### □ AST的定义

Node

**Public Member Functions** 

```
virtual void accept(Visitor &visitor) = 0;
```

**Public Attributes** 

```
Position loc;
```

- □ 访问者Visitor
  - Visitor

```
struct Node:
    struct InitVal: Node:
    struct Assembly : Node;
    struct GlobalDef : virtual Node:
        struct FuncDef : GlobalDef;
        struct VarDef : Stmt, GlobalDef;
    struct Stmt : virtual Node:
        struct VarDef : Stmt, GlobalDef;
        struct AssignStmt : Stmt;
        struct ReturnStmt : Stmt:
        struct BlockStmt : Stmt;
        struct EmptyStmt : Stmt;
        struct ExprStmt : Stmt;
        struct IfStmt : Stmt:
        struct WhileStmt : Stmt;
        struct BreakStmt : Stmt;
        struct ContinueStmt : Stmt;
    struct Expr : Node:
        struct CondExpr : Expr;
            struct UnaryCondExpr : CondExpr;
            struct BinaryCondExpr : CondExpr;
        struct AddExpr : Expr;
            struct BinaryExpr : AddExpr;
            struct UnaryExpr : AddExpr;
            struct LVal : AddExpr;
            struct Literal : AddExpr;
            struct FuncCallStmt : AddExpr;
    struct FuncParam : Node:
    struct FuncFParamList : Node;
```

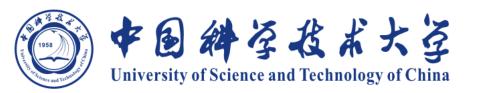


### Lab: ParseTree =>AST

#### □ 访问者Visitor

```
class Visitor
public:
    virtual void visit(Assembly &node) = 0;
    virtual void visit(FuncDef &node) = 0:
    virtual void visit(BinaryExpr &node) = 0;
    virtual void visit(UnaryExpr &node) = 0;
    virtual void visit(LVal &node) = 0;
    virtual void visit(Literal &node) = 0;
    virtual void visit(ReturnStmt &node) = 0;
    virtual void visit(VarDef &node) = 0:
    virtual void visit(AssignStmt &node) = 0;
    virtual void visit(FuncCallStmt &node) = 0;
    virtual void visit(BlockStmt &node) = 0;
    virtual void visit(EmptyStmt &node) = 0;
    virtual void visit(ExprStmt &node) = 0;
    virtual void visit(FuncParam &node) = 0;
    virtual void visit(FuncFParamList &node) = 0;
    virtual void visit(IfStmt &node) = 0;
    virtual void visit(WhileStmt &node) = 0;
    virtual void visit(BreakStmt &node) = 0;
    virtual void visit(ContinueStmt &node) = 0:
    virtual void visit(UnaryCondExpr &node) = 0;
   virtual void visit(BinaryCondExpr &node) = 0;
    virtual void visit(InitVal &node) = 0:
};
```

```
struct Node;
    struct InitVal: Node:
    struct Assembly : Node;
    struct GlobalDef : virtual Node;
        struct FuncDef : GlobalDef:
        struct VarDef : Stmt, GlobalDef;
    struct Stmt : virtual Node:
        struct VarDef : Stmt, GlobalDef;
        struct AssignStmt : Stmt;
        struct ReturnStmt : Stmt:
        struct BlockStmt : Stmt:
        struct EmptyStmt : Stmt;
        struct ExprStmt : Stmt;
        struct IfStmt : Stmt;
        struct WhileStmt : Stmt;
        struct BreakStmt : Stmt:
        struct ContinueStmt : Stmt;
    struct Expr : Node;
        struct CondExpr : Expr;
            struct UnaryCondExpr : CondExpr;
            struct BinaryCondExpr : CondExpr;
        struct AddExpr : Expr;
            struct BinaryExpr : AddExpr;
            struct UnaryExpr : AddExpr;
            struct LVal : AddExpr;
            struct Literal : AddExpr;
            struct FuncCallStmt : AddExpr;
    struct FuncParam : Node:
    struct FuncFParamList : Node;
```



## 下期预告: 语义计算

- □自上而下计算
- □自下而上计算