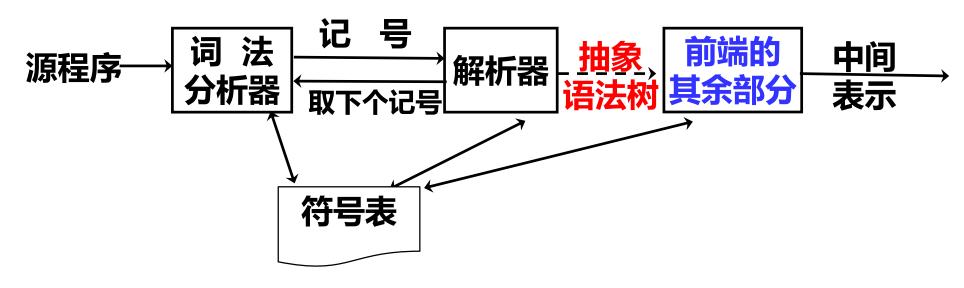


语法制导的翻译I

《编译原理和技术(H)》

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



- □ 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、注释解析树等
- □ 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



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
- □ S属性定义、L属性定义

□ 语法制导的定义(Syntax-Directed Definition)

- 基础的上下文无关文法
- 用来表示语法成分 对应的语义信息

描述语义属性值 的计算规则

- 每个文法符号有一组属性~
- 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 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
 - \checkmark 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
 - -- 由子结点计算得来, 自下而上传递
 - ✓ b是继承属性(inherited attribute): 如果b是产生式右部某个文法符号X的 属性, $c_1, c_2, ..., c_k$ 是A的属性或右部文法符号的属性
 - -- 由兄弟结点、父结点和自己的属性值来计算,同层或自上而下传递



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

 $S' \rightarrow S$

print(S.val)

 $S \rightarrow D_1 S_1 D_2$

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

 $S \rightarrow 2$

S.val = true

 $D \rightarrow 0$

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

 $D \rightarrow 1$

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

S.val: S对应的串是否是回文数

对S和D加下标

以区分同类语法

结构的不同实例

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

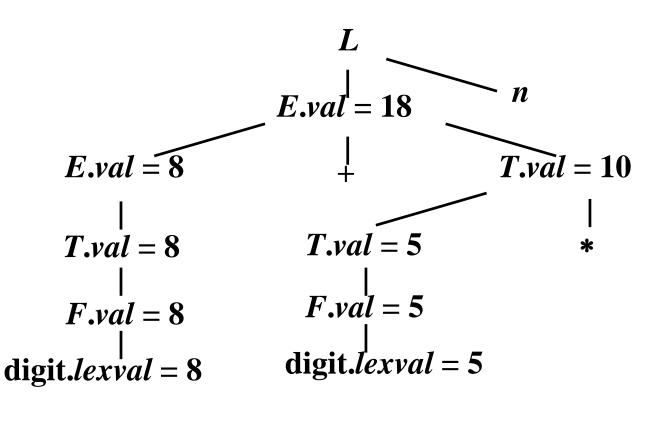
各文法符号的属性均是**综合属性**的语法制导定义——S **属性定义**

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

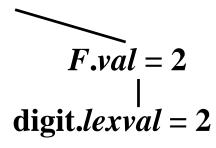
注释解析树(annotated parse tree)

□ 结点的属性值都标注出来的解析树

8+5*2 n (n为换行符)的注释解析树, 在树的根结点打印18



结点的综合属性值 可由子结点的属性 值计算得到





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)	

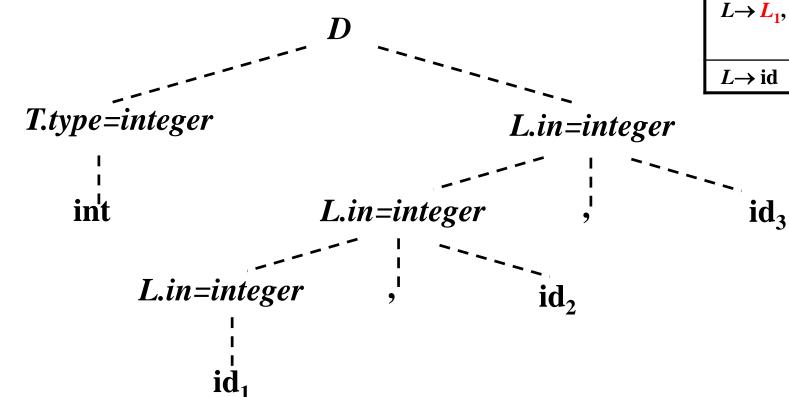
type-T的综合属性

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

University of Science and Technology of China

int id₁, id₂, id₃

解析树(虚线),注释了部分属性



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)	



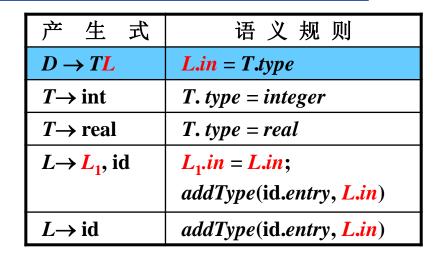
属性依赖图(dependence graph)

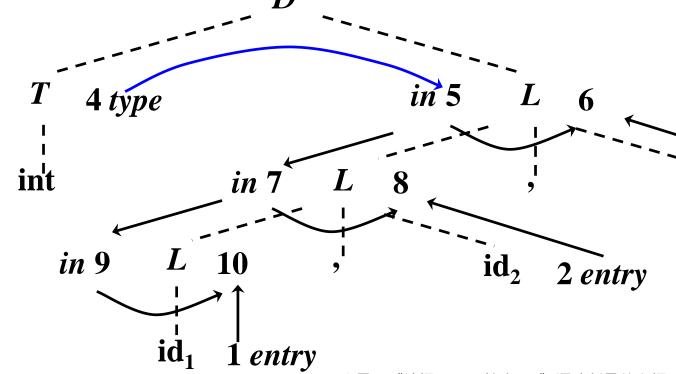


University of Science and Technology of China

□ 属性依赖图

■ 描述解析树中结点的属性间依赖关系的有向图 int id_1 , id_2 , id_3 的依赖图(实线)





顶点为属性: 对应解析树中每个文法 符号X 的每个属性a

3 entry

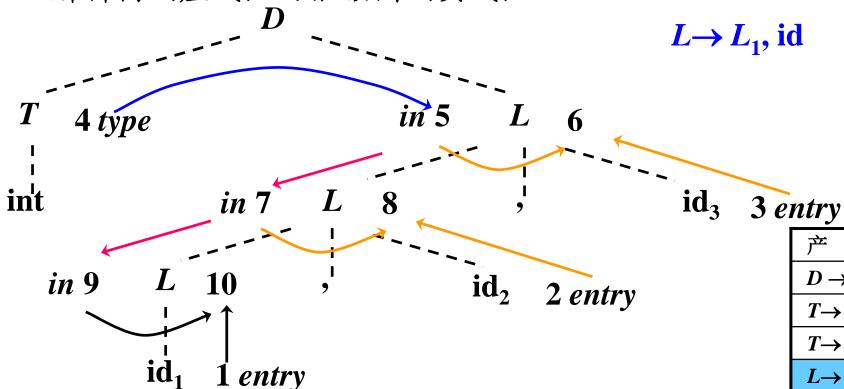
弧为属性间依赖关系:如果属性*X.a* 的值依赖于属性*Y.b*的值,则存在从 *Y.b*的顶点指向*X.a*的顶点的弧



属性依赖图(dependence graph)

int id₁, id₂, id₃

解析树(虚线)的依赖图(实线)

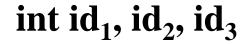


 $L_1.in = L.in;$ addType (id.entry, L.in)

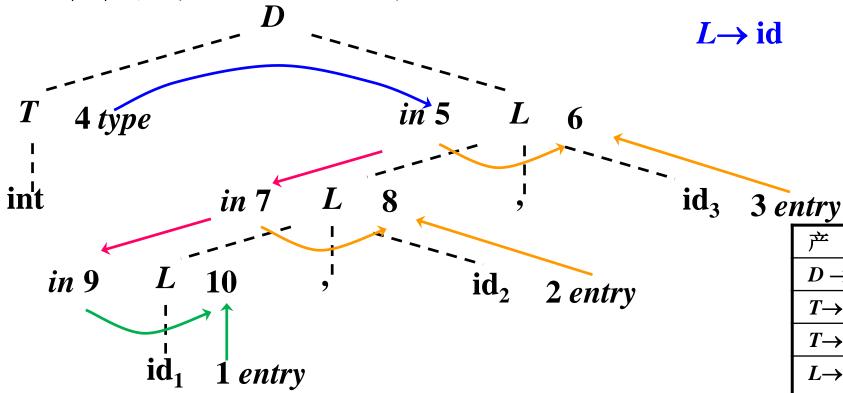
产生式	语 义 规 则	
$D \rightarrow TL$	L.in = T.type	
$T \rightarrow int$	T. type = integer	
$T \rightarrow \text{real}$	T. type = real	
$L \rightarrow L_1$, id	L_1 .in = L .in;	
	addType(id.entry, L.in)	
$L \rightarrow id$	addType(id.entry, L.in)	



属性依赖图(dependence graph)



解析树(虚线)的依赖图(实线)



addType (id.entry, L.in)

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		



S 属性定义和 L 属性定义

□ S属性定义

仅含综合属性的语法制导定义

□ L属性定义(属性信息从左流向右)

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

可以按边解析边翻译的方式计算继承属性

□ S属性定义是L属性定义

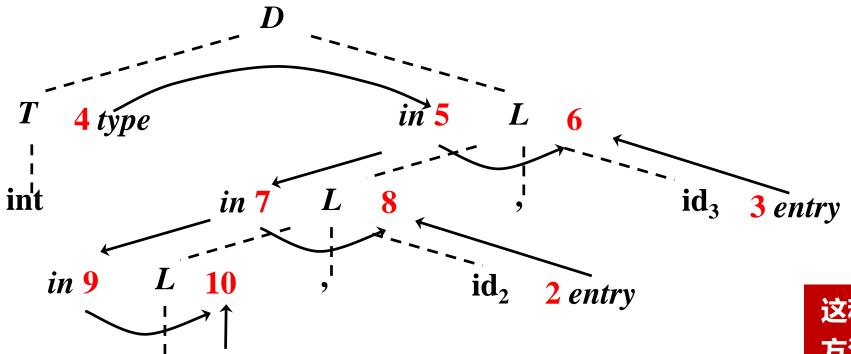


■ 拓扑排序(topological sort): 是DAG的结点的一种排序 $m_1,...,m_k$, 若有 m_i 到 m_j

的边,则在排序中 m_i 先于 m_j

entry

例 拓扑排序为结点1, 2, 3, 4, 5, 6, 7, 8, 9, 10 -->



序号为 n 的结点的属性写成 a_n a_4 = integer; $a_5 = a_4$; $addType(id_3.entry, a_5)$; $a_7 = a_5$; $addType(id_2.entry, a_7)$; $a_9 = a_7$; $addType(id_1.entry, a_9)$;

这种语义规则的计算 方法称为解析树方法



■ 属性计算次序

- 1) 构造输入的解析树
- 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

语义规则的计算方法

- □ 解析树方法
 - 刚才介绍的方法, 动态确定输入的属性计算次序, 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(如边解析边计算),(语言设计者提供的)语义规则必须符合所选解析方法的限制

——适用于自动生成的方法



4.2 S属性定义的自下而上计算

- □ 语法树及其构造(文法对构造的影响)
 - 语法制导定义 vs. 翻译方案
- □ S属性定义的自下而上计算

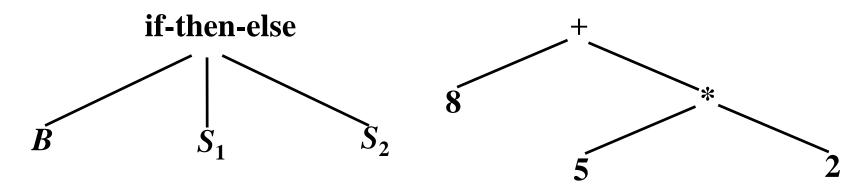
语法树(syntax tree)

□ 语法树是解析树的浓缩表示

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解析树, 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	
$F o \mathrm{id}$	F.nptr = mkLeaf (id, id.entry)	
$F \rightarrow \text{num}$	F.nptr = mkLeaf (num, num.val)	

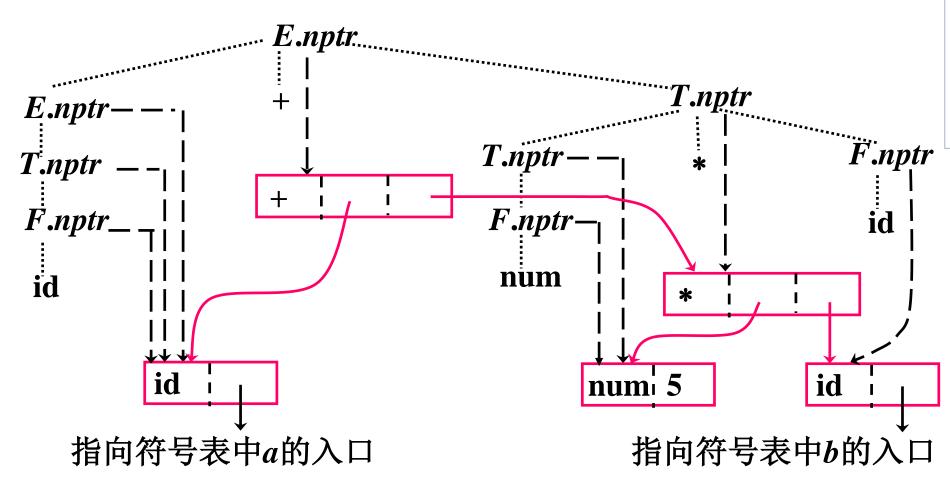
参见: <u>bison-examples.tar.gz</u> 中的config/asgn2ast.y, asgn.lex



语法树的构造

University of Science and Technology of China

a+5*b的语法树的构造



实际执行的函数调用序列:

 $p_1 = mkleaf(id, entrya);$

 $p_2 = mkleaf(\mathbf{num}, 5);$

 $p_3 = mkleaf(id, entryb);$

 $p_4 = mknode('*', p_2, p_3);$

 $p_5 = mknode('+', p_1, p_4);$

□ 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 \rightarrow \text{num}
                  {F.nptr = mkLeaf (num, num.val);}
```

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

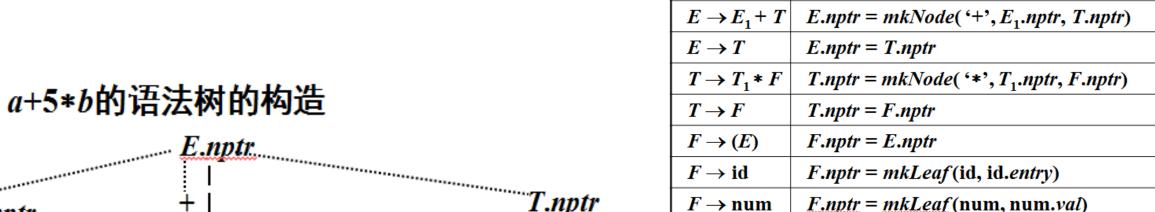


4.2 S属性定义的自下而上计算

- □ 语法树及其构造(文法对构造的影响)
 - 语法制导定义 vs. 翻译方案
- □ S属性定义的自下而上计算

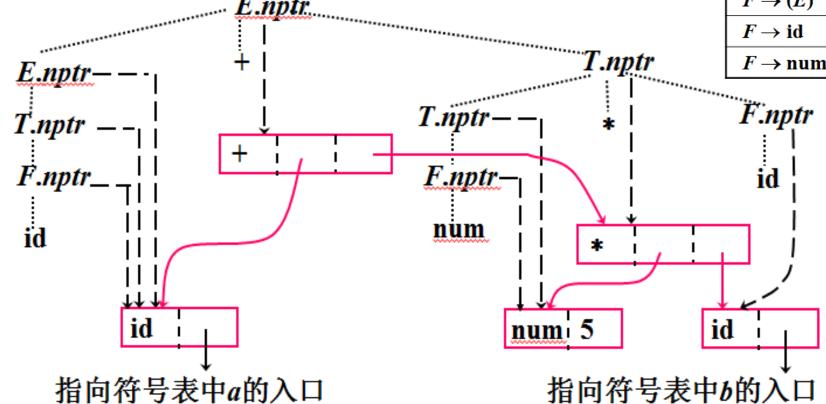
语义规则

University of Science and Technology of China



生

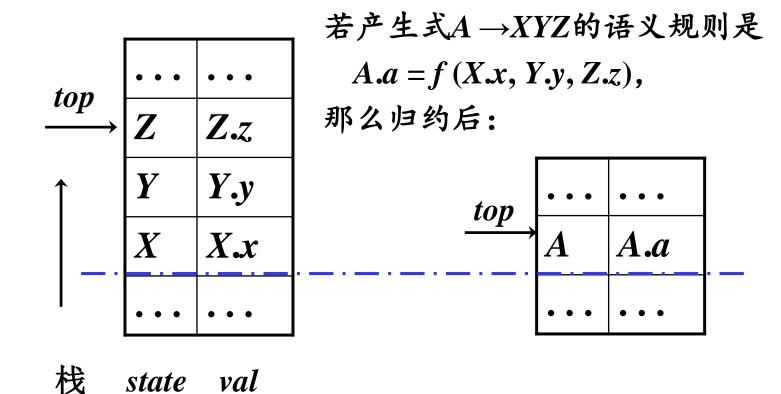
太



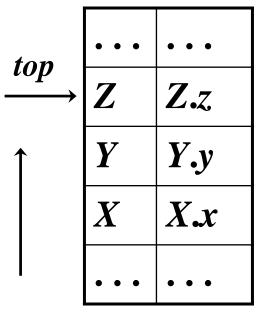
S属性的自下而上计算

□ 边解析边计算

LR解析器的栈增加一个域来保存综合属性值







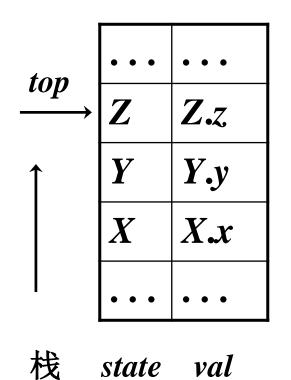
栈 state val

产生式	语 义 规 则
$L \rightarrow E$ n	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
$F \rightarrow \text{digit}$	F.val = digit.lexval

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



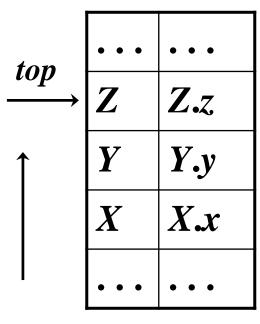




产生式	代 码 段
$L \rightarrow E$ n	print (E.val)
$E \rightarrow E_1 + T$	$E.val = E_1.val + T.val$
$E \rightarrow T$	E.val = T.val
$T \rightarrow T_1 * F$	$T.val = T_1.val * F.val$
$T \rightarrow F$	T.val = F.val
$F \rightarrow (E)$	F.val = E.val
$F \rightarrow \text{digit}$	F.val = digit.lexval







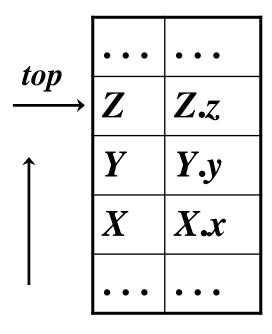
.		
栈	state	val

产生式	代 码 段
$L \rightarrow E$ n	print (val[top-1])
$E \rightarrow E_1 + T$	$E.val = E_1.val + T.val$
$E \rightarrow T$	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
$F \rightarrow \text{digit}$	F.val = digit.lexval

注: 栈顶位置指示器top的修改由原来的解析程序在语义动作执行后去做







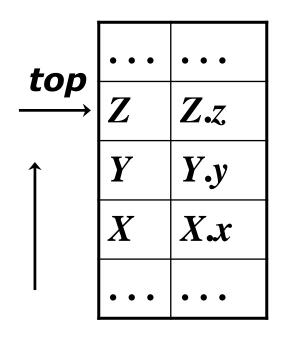
栈	state	val
亿	state	vai

\ b bs	
产生式	代码段
$L \rightarrow E$ n	<pre>print (val[top-1]);</pre>
$E \rightarrow E_1 + T$	val[top-2] = val[top-2] + val[top];
$E \rightarrow T$	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
$F \rightarrow \text{digit}$	F.val = digit.lexval

注: 栈顶位置指示器top的修改由原来的解析程序在语义动作执行后去做







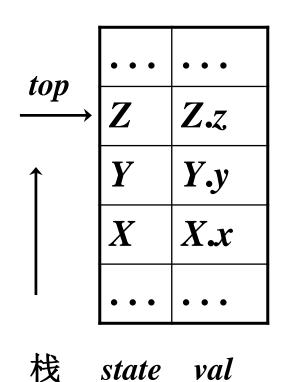
栈 state val

产生式	代 码 段
$L \rightarrow E$ n	<pre>print (val[top-1]);</pre>
$E \rightarrow E_1 + T$	val[top-2] = val[top-2] + val[top];
$E \rightarrow T$	
$T \rightarrow T_1 * F$	$T.val = T_1.val * F.val$
$T \rightarrow F$	T.val = F.val
$F \rightarrow (E)$	F.val = E.val
$F \rightarrow \text{digit}$	F.val = digit.lexval

注: 栈顶位置指示器top的修改由原来的解析程序在语义动作执行后去做







state

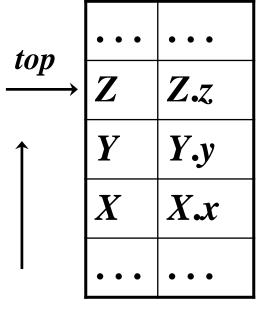
val

产生式	代 码 段
$L \rightarrow E$ n	<pre>print (val[top-1]);</pre>
$E \to E_1 + T$	val[top-2] = val[top-2] + val[top];
$E \rightarrow T$	
$T \rightarrow T_1 * F$	$val[top-2] = val[top-2] \times val[top];$
$T \rightarrow F$	T.val = F.val
$F \rightarrow (E)$	F.val = E.val
$F o ext{digit}$	F.val = digit.lexval

注: 栈顶位置指示器top的修改由原来的解析程序在语义动作执行后去做







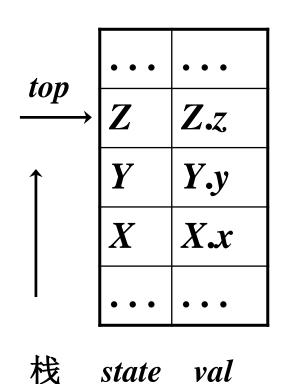
栈	state	val

产生式	代 码 段
$L \rightarrow E$ n	<pre>print (val[top-1]);</pre>
$E \rightarrow E_1 + T$	val[top-2] = val[top-2] + val[top];
$E \rightarrow T$	
$T \rightarrow T_1 * F$	$val[top-2] = val[top-2] \times val[top];$
$T \rightarrow F$	
$F \rightarrow (E)$	F.val = E.val
$F \rightarrow \text{digit}$	F.val = digit.lexval

注: 栈顶位置指示器top的修改由原来的解析程序在语义动作执行后去做







state

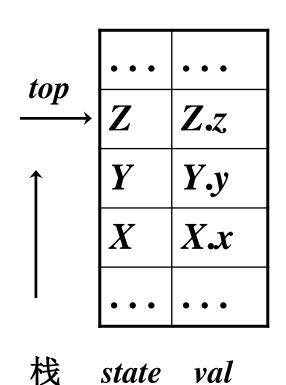
val

产生式	代 码 段
$L \rightarrow E$ n	<pre>print (val[top-1]);</pre>
$E \rightarrow E_1 + T$	val[top-2] = val[top-2] + val[top];
$E \rightarrow T$	
$T \rightarrow T_1 * F$	$val[top-2] = val[top-2] \times val[top];$
$T \rightarrow F$	
$F \rightarrow (E)$	val[top-2] = val[top-1];
$F \rightarrow \text{digit}$	F.val = digit.lexval

注: 栈顶位置指示器top的修改由原来的解析程序在语义动作执行后去做







产生式	代 码 段
$L \rightarrow E$ n	<pre>print (val[top-1]);</pre>
$E \to E_1 + T$	val[top-2] = val[top-2] + val[top];
$E \rightarrow T$	
$T \rightarrow T_1 * F$	$val[top-2] = val[top-2] \times val[top];$
$T \rightarrow F$	
$F \rightarrow (E)$	val[top-2] = val[top-1];
$F \rightarrow \text{digit}$	

注: 栈顶位置指示器top的修改由原来的解析程序在语义动作执行后去做



Bison举例bison-examples: config/expr.y



```
%{
#include <stdio.h>
#include <math.h>
%}
               各种语义值类型和域
                 置于共用体中
%union
 float val;
%token NUMBER
%token PLUS MINUS MULT DIV EXPON
%left MINUS PLUS
%left MULT DIV
                 声明exp、Number
%right EXPON
                 的语义值是在val域
%type <val> exp NUMBER
%%
```

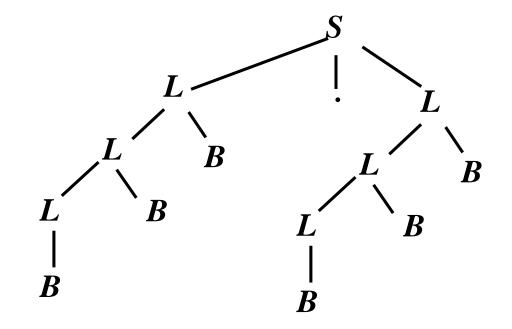
```
input:
             | input line
      : NUMBER
                         { $$ = $1;
exp
     \{ \exp PLUS \ \exp \ \{ \$\$ = \$1 + \$3; \} \}
     \{ \$ = \$1 - \$3; \}
     \{ \exp MULT \ \exp \ \{ \$\$ = \$1 * \$3; \} \}
     \{ \exp DIV \ \exp \ \{ \$\$ = \$1 / \$3; \} \}
     | MINUS exp %prec MINUS { $$ = -$2; }
     \{ \exp EXPON \exp \{ \$\$ = pow(\$1,\$3); \} \}
    | LB exp RB \{ \$\$ = \$2; \}
%%
yyerror(char *message)
{ printf("%s\n",message);}
int main(int argc, char *argv[])
 yyparse(); return(0);}
```

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

若按2²+0+2⁰+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

$$S \rightarrow L \cdot L \mid L$$

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

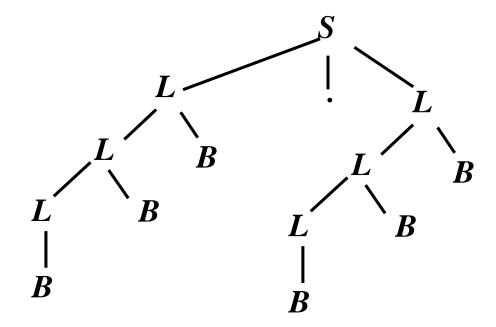


例如, 输入101.101时, S.val = 5.625 (可以修改文法)

若小数点左边按 $(1 \times 2 + 0) \times 2 + 1$ 计算。该办法不能直接用于小数点右边,需改成 $((1 \times 2 + 0) \times 2 + 1)/2^3$,这时需要综合属性来统计B的个数

$$S \rightarrow L \cdot L \mid L$$

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



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

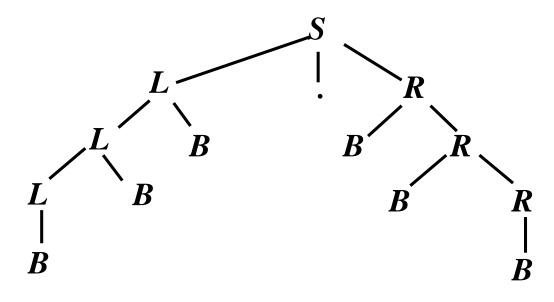
更清楚的办法是将文法改成下面的形式

$$S \to L \cdot {\overset{}{\scriptstyle{R}}} \mid L$$

$$L \rightarrow L B \mid B$$

$$R \rightarrow B R \mid B$$

$$B \rightarrow 0 \mid 1$$





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

$$S \rightarrow L \cdot R$$

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

$$S \rightarrow L$$

$$S. val = L. val$$

$$L \rightarrow L_1 B$$

$$\mathbf{L.} \ val = \mathbf{L_1.} \ val \times 2 + \mathbf{B.} \ val$$

$$L \rightarrow B$$

$$L. val = B. val$$

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

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

$$R \rightarrow B$$

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

$$B \rightarrow 0$$

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

$$B \rightarrow 1$$

$$B. val = 1$$

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'
ightarrow E \quad print \ (E.\ code\)$$
 $E
ightarrow E_1 + T$

if $T.\ op == plus \ then$
 $E.\ code = E_1.\ code || "+" || "("||T.\ code||")"$

else

 $E.\ code = E_1.\ code \ || "+" || T.\ code;$
 $E.\ op = plus$
 $E
ightarrow T \quad E.\ code = T.\ code; \ E.\ op = T.\ op$

□ 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
$$T.\ code = \text{``('' || } T_1.\ code \mid| \text{``)'' || } \text{``**' || } \text{``('' || } F.\ code \mid| \text{``)''}$$
 else
$$T.\ code = T_1.\ code \mid| \text{``*'' || } \text{``('' || } F.\ code \mid| \text{``)''}$$
 else if $T_1.\ op = plus$ then
$$T.\ code = \text{``('' || } T_1.\ code \mid| \text{``)'' || } \text{``**'' || } F.\ code$$
 else
$$T.\ code = T_1.\ code \mid| \text{``*'' || } F.\ code;$$

$$T.\ op = times$$

□ 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

- \blacksquare 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



$$S' \rightarrow E$$
 $E. \ left_op = 0; E. \ right_op = 0; print (E. \ code)$
 $E \rightarrow E_1 + T$
 $E_1. \ left_op = E. \ left_op; E_1. \ right_op = 1;$
 $T. \ left_op = 1; T. \ right_op = E. \ right_op;$
 $E. \ code = E_1. \ code / "+" / T. \ code ; E. \ self_op = 1;$
 $E \rightarrow T$
 $T. \ left_op = E. \ left_op;$
 $T. \ right_op = E. \ right_op;$

 $E.\ code = T.\ code;\ E.\ self_op = T.\ self_op$





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

$$F \rightarrow id$$

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





$$F \rightarrow (E)$$

 $E. left_op = 0; E. right_op = 0;$

 $F. \ self_op =$ 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 $(E.\ self_op >= F.\ right_op)$ then $E.\ code$ else "(" $\parallel E.\ code \parallel$ ")"



下期预告: L属性定义的语义计算

- □ L属性定义的自上而下计算
- □ L属性定义的自下而上计算