

第三章：语法分析

语法树
二义性文法
文法等价变换

内容介绍

- ◆ 语法树
- ◆ 文法的二义性
- ◆ 文法等价变换

1. 1 语言树的定义

- ◆ 语法树：设G是给定的语法，称满足下列条件的树为G的一棵语法树：
 1. 树的每个节点都标有G的一个语法符号，且根节点标有初始符S。
 2. 如果一个非叶节点A按从左到右顺序有n个儿子节点 B_1 、 B_2 、…、 B_n ，则： $A \rightarrow B_1 B_2 \cdots B_n$ 一定是G的一个产生式.

1. 1 语言树的定义

- ◆ 例子：有文法

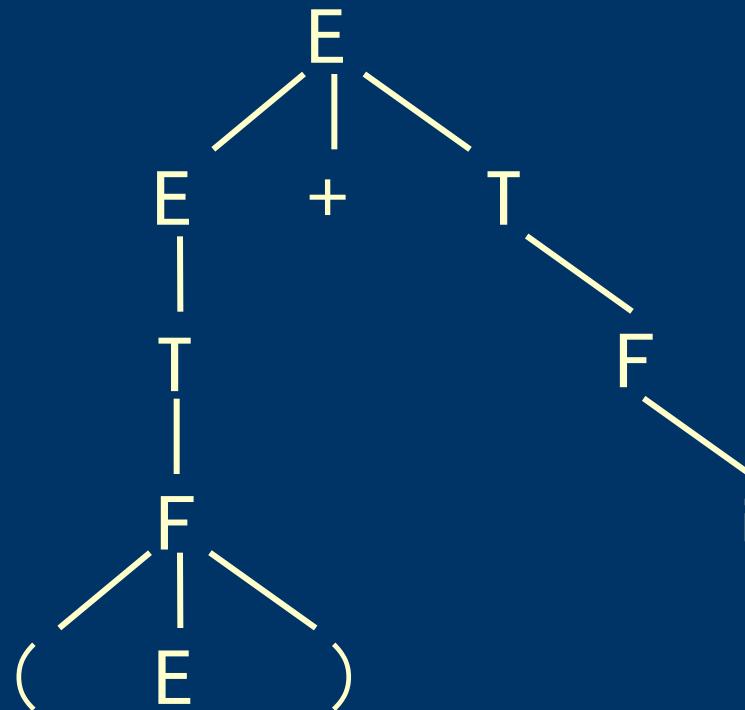
$E \rightarrow E + T$

$E \rightarrow T$

$T \rightarrow T * F$

$T \rightarrow F$

$F \rightarrow (E) \mid i$



1. 2 语法树的作用

- ◆ 作用主要有两点：
- ◆ 语法树反映出我们推导的过程，每一步节点的生长过程都可以对应到我们的一步推导。
- ◆ 语法树反应出串的语法结构。

1.3 语法树和语法概念的关系

- ◆ 子树：某语法树T中的某一节点A和它所有分支组成的树T'，则称T'是T的一颗子树。
- ◆ 简单子树：某一节点A与其子节点（单层节点）组成的树。从严格意义上来说，A应该有且仅有—层子节点。

1. 3 语法树和语法概念的关系

结论

- ◆ 语法树的所有叶节点对应语法符号从左到右连接起来组成的符号串是一个句型。
- ◆ 语法树的子树所有叶节点对应语法符号从左到右连接起来组成的符号串是一个短语。
- ◆ 简单子树所有叶节点对应语法符号从左到右连接起来组成的符号串是一个简单短语。
- ◆ 最左简单子树叶节点对应语法符号从左到右连接起来组成的符号串是句柄。

1. 3 语法树和语法概念的关系

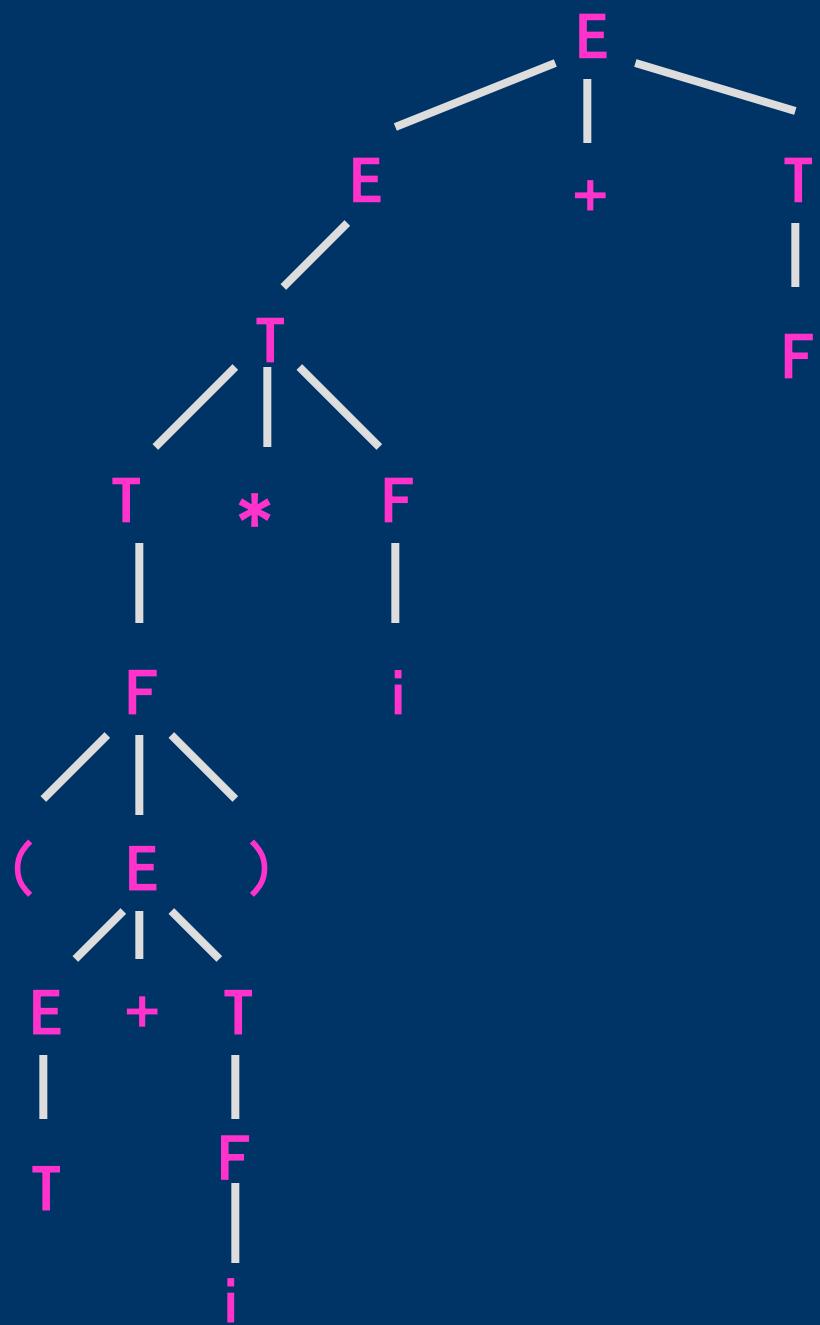
例：已知文法G：

$$E \rightarrow T \mid E + T$$

$$T \rightarrow F \mid T * F$$

$$F \rightarrow (E) \mid i$$

问： $(T+i)^*i+F$ 是否为G的一个句型，若是请列出该句型的所有短语、简单短语和句柄。



例：文法G：

E → T | E + T

$T \rightarrow F \mid T * F$

F → (E) | i

(T+i)* i + F推导过程如下:

$$E \Rightarrow E + T$$

⇒ T+T

⇒ T*F+T

⇒ F*F+T

$\Rightarrow (E) * F + T$

$\Rightarrow (E + T) * E + T$

$\rightarrow (T + T) * F + T$

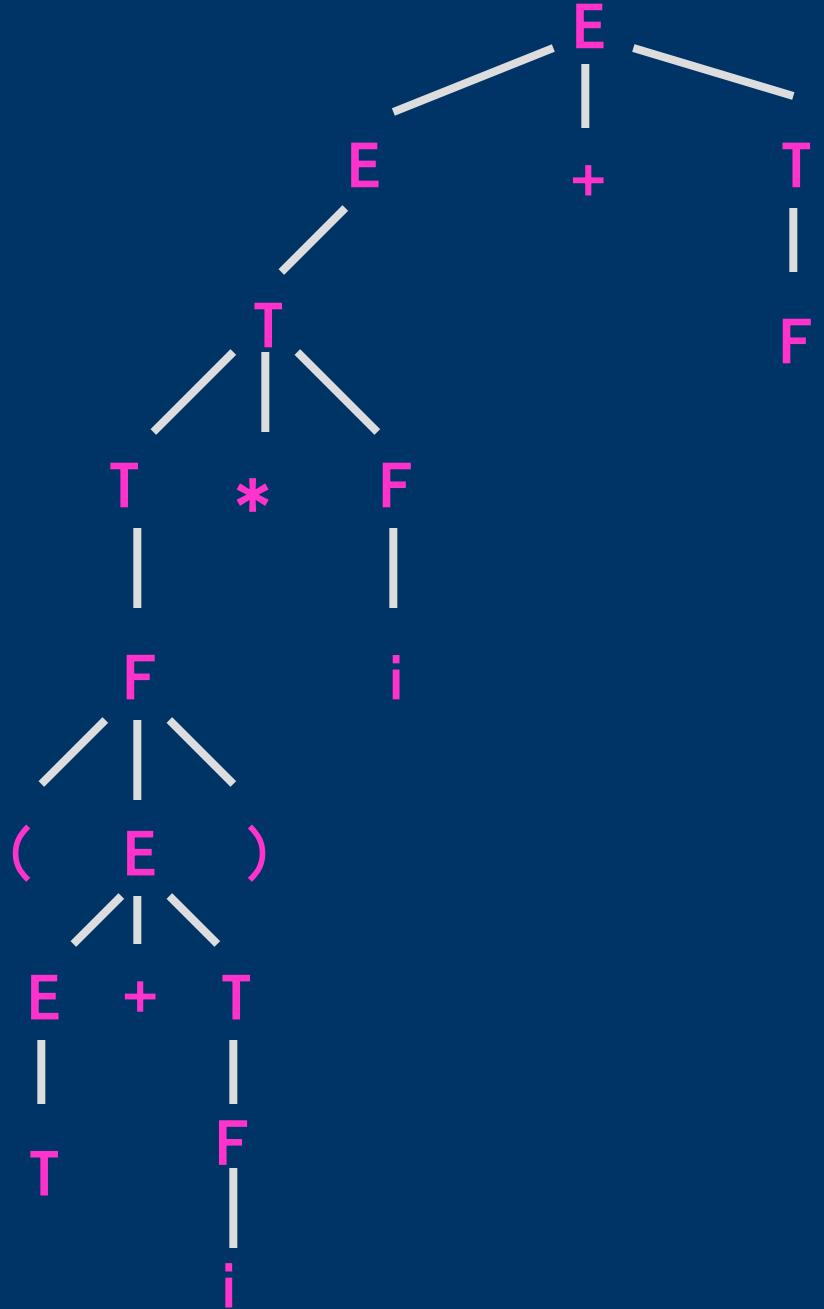
$\Rightarrow (T + F) * F + T$

$\rightarrow (T+i) * F + T$

$\Rightarrow (T + i) * i + T$

$\Rightarrow (T+i)^* i \neq F$

$\rightarrow (\mathbf{1} + \mathbf{1}) = \mathbf{1} + \mathbf{1}$



关于句型 $(T+i)^*i+F$ 的短语、简单短语和句柄：

● 短语有8个：

1. $(T+i)^*i+F$
2. $(T+i)^*$
3. $(T+i)$
4. $T+i$
5. T
6. 第一个*i*
7. 第二个*i*
8. F

● 简单短语有4个：

T , 第一个*i*, 第二个*i*, F

● 句柄：

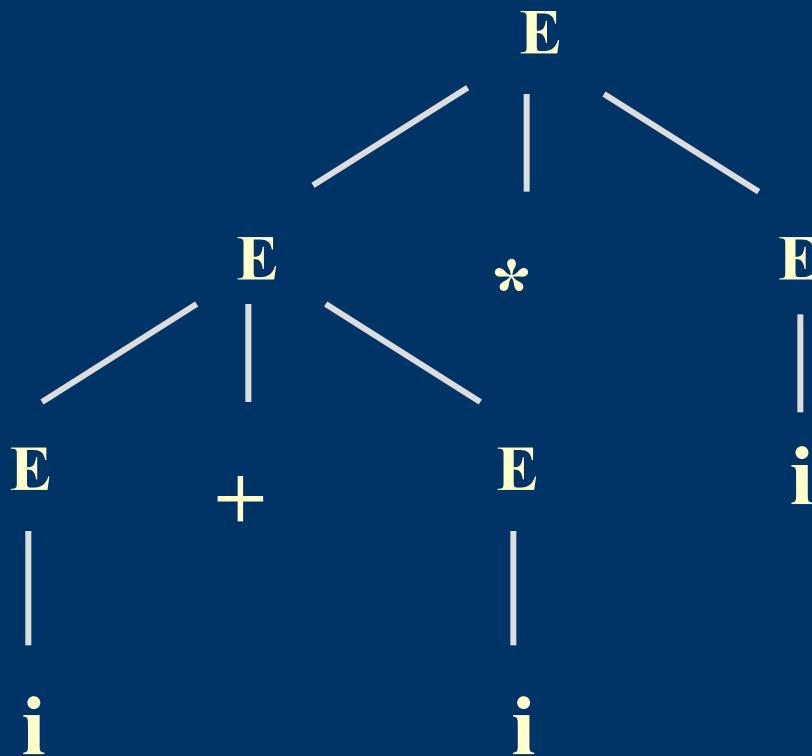
T

2.1 文法的二义性

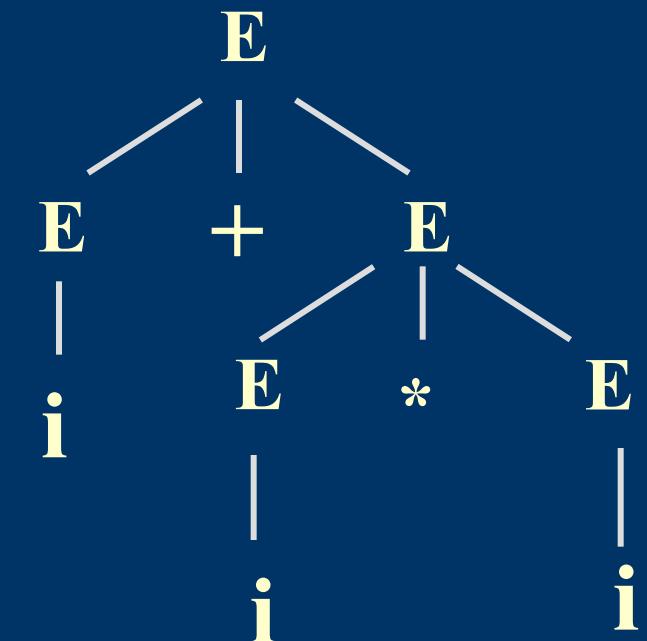
对一个文法 G , 如果至少存在一个句子, 有两棵 (或两棵以上) 不同的语法树, 则称该句子是二义性的. 包含有二义性句子的文法称为二义性文法, 否则, 该文法是无二义性的.

例1: $G(E) : E \rightarrow i \mid E + E \mid E * E \mid (E)$

句型 $i + i * i$



$(i+i)*i$



$i+(i*i)$

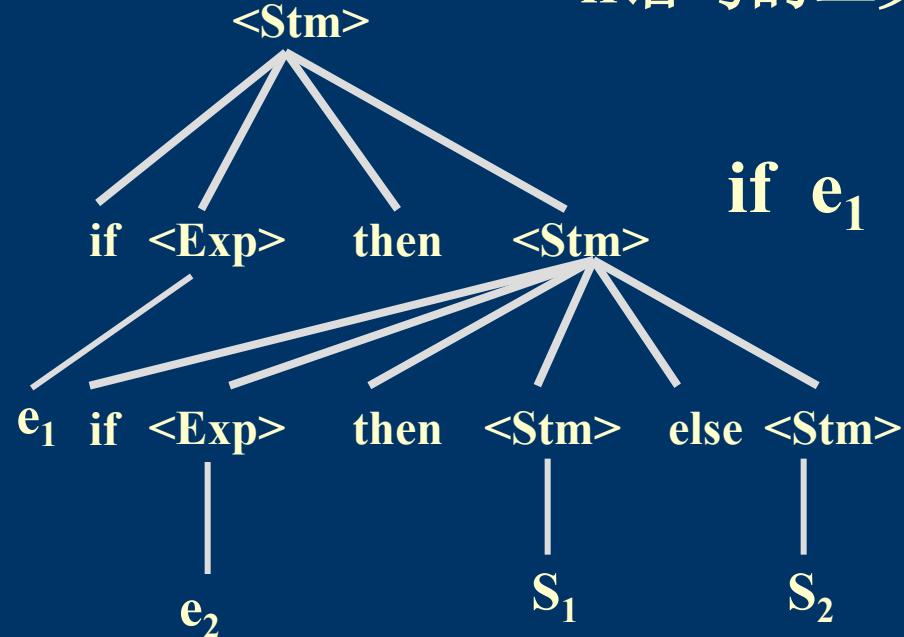
例2: $\langle \text{Stm} \rangle \rightarrow \text{if } \langle \text{Exp} \rangle \text{ then } \langle \text{Stm} \rangle \text{ else } \langle \text{Stm} \rangle$
 $\langle \text{Stm} \rangle \rightarrow \text{if } \langle \text{Exp} \rangle \text{ then } \langle \text{Stm} \rangle$
 $\langle \text{Stm} \rangle \rightarrow \dots \dots$

假设有条件语句:

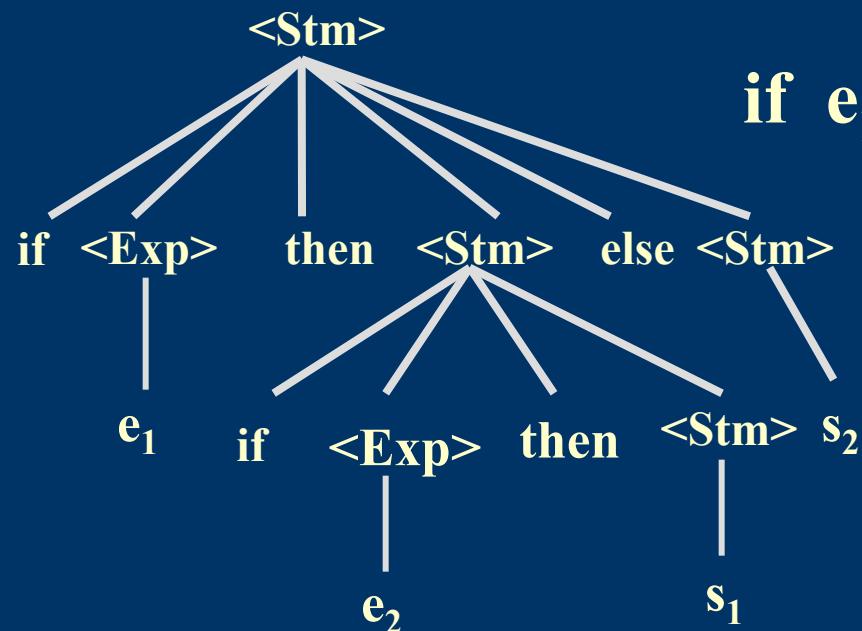
if e_1 then if e_2 then s_1 else s_2

则可有下图所示的两棵不同的语法分析树:

if语句的二义性



if e_1 then if e_2 then s_1 else s_2



if e_1 then if e_2 then s_1 else s_2

2. 2 文法二义性的判定和利用

- ◆ 文法二义性的判断

目前，不存在一个一般性的方法来判断一个现有文法是否为二义性文法。

- ◆ 常用的经验性判断：

$S \rightarrow SS | a$ 可以推导出SSS串，必有二义性

$S \rightarrow S+S$ 可以推导出S+S+S，必有二义性

2. 2 文法二义性的判定和利用

- ◆ 文法二义性的利用

对二义性文法进行修改，消除其二义性会导致文法的复杂程度和符号数目迅速升高。

可以利用二义性文法状态少，分析快的特点，使用二义性文法，对具体问题加入语义规则，约束其二义性即可。

3. 文法等价变化

- ◆ 提取公共前缀
- ◆ 消除直接左递归

3. 1 提取公共前缀

- ◆ 公共前缀: A的不同产生式的右部具有相同的前缀. 这种情形不满足自顶向下的语法分析条件.
- ◆ 可用提取左因子的方法消除产生式的公共前缀. 假定关于 A 的规则是:

$A \rightarrow \delta\beta_1 | \delta\beta_2 | \dots | \delta\beta_n | \gamma_1 | \gamma_2 | \dots | \gamma_m$ (其中每个 γ 不以 δ 开头)
则将这些规则写成:

$$A \rightarrow \delta A' | \gamma_1 | \gamma_2 | \dots | \gamma_m$$

$$A' \rightarrow \beta_1 | \beta_2 | \dots | \beta_n$$

经过反复提取左因子，就能够使每个非终极符的不同产生式的右部具有不同的前缀.

3. 1 提取公共前缀

- ◆ 例子：

$A \rightarrow aB | aC | aD$

$B \rightarrow bB | bD | c$

$D \rightarrow d$

- ◆ 提取公共前缀

$A \rightarrow aA'$

$A' \rightarrow B | C | D$

$B \rightarrow bB' | c$

$B' \rightarrow B | D$

$D \rightarrow d$

3. 2 消除直接左递归

- ◆ 对于简单情形 $A \rightarrow A\alpha \mid \beta$ 即有 $A \Rightarrow \beta\alpha\alpha\dots\alpha$
则转化 $A \rightarrow \beta A'$ $A' \rightarrow \alpha A' \mid \varepsilon$
- ◆ 对于一般情形
$$A \rightarrow A\alpha_1 \mid A\alpha_2 \mid \dots \mid A\alpha_m \mid \beta_1 \mid \beta_2 \mid \dots \mid \beta_n$$

则转化
$$A \rightarrow (\beta_1 \mid \beta_2 \mid \dots \mid \beta_n) A'$$

$$A' \rightarrow (\alpha_1 \mid \alpha_2 \mid \dots \mid \alpha_m) A' \mid \varepsilon$$

3. 2 消除直接左递归

- ◆ 例子

$$E \rightarrow E^+T \mid T$$

$$T \rightarrow T * F \mid F$$

$$F \rightarrow (E) \mid i$$

$$\begin{array}{l} A \rightarrow A\alpha \mid \beta \\ E \rightarrow E^+T \mid T \\ \alpha = +T \quad \beta = T \\ E \rightarrow TE' \\ E' \rightarrow +TE' \mid \epsilon \end{array}$$

$$\begin{array}{l} E \rightarrow TE' \\ E' \rightarrow +TE' \mid \epsilon \\ T \rightarrow FT' \\ T' \rightarrow *FT' \mid \epsilon \\ F \rightarrow (E) \mid i \end{array}$$

若干历年考试题

- ◆ 构造一个文法G，使

$$L(G) = \{a^n b^m c^k \mid m=n+k, n \geq 1, m > 1, k \geq 1\}$$

G[S] : S → AB

A → aAb | ab

B → bBc | bc

若干历年考试题

- ◆ 已知文法G[Z] :

$Z \rightarrow WV$

$W \rightarrow aB \mid aW \mid a$

$B \rightarrow b \mid bB$

$V \rightarrow bV \mid dD$

$D \rightarrow d \mid dD$

- ◆ 判断文法G[Z]是否为二义性文法，如果是请举例

句子abdd有两棵语法树。

若干历年考试题

- ◆ 设有一个文法G[S]：

$S \rightarrow V$

$V \rightarrow T \mid V \cdot T$

$T \rightarrow F \mid T + F$

$F \rightarrow V^* \mid [$

句型 $F+F \cdot [$ 的短语，简单短语和句柄分别为：
首F, $F+F$, $F+F \cdot [$, $[$
首F, $[$;
首F。