

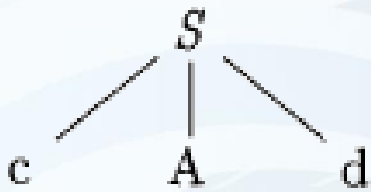
## 4.4 自顶向下的语法分析

- ❖ 自顶向下的语法分析：从分析树的根结点开始，深度优先地创建该树的各个结点。
- ❖ 递归下降分析程序：由一组过程组成，每个非终结符有一个对应的过程。
  - 从开始符对应的过程开始执行，如果这个过程的过程体刚好扫描了整个输入串，就停止执行并宣布语法分析成功。
  - 可能存在回溯。

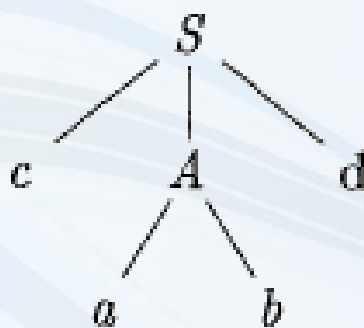
# 回溯问题

❖ 文法:  $S \rightarrow cAd$        $A \rightarrow ab \mid a$

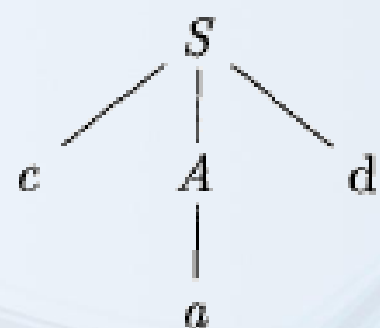
输入:  $cad$



(a)



(b)



(c)

# LL(1)文法

- ❖ 第一个L表示从左向右扫描输入，第二个L表示最左推导，1表示只需向前看一个输入符号
- ❖ 对于LL(1)文法，可以构造出不需要回溯的递归下降语法分析器。

## 文法

- ①  $S \rightarrow aBC$
- ②  $B \rightarrow bC$
- ③  $B \rightarrow dB$
- ④  $B \rightarrow \varepsilon$
- ⑤  $C \rightarrow c$
- ⑥  $C \rightarrow a$

输入:  $abac$

最左推导过程:

$S$   
 $\Rightarrow aBC$   
 $\Rightarrow abCC$   
 $\Rightarrow abaC$   
 $\Rightarrow abac$

# 可空性 (Nullability)

❖ 非终结符A**可空** ( nullable ) 当且仅当:

$$A \xrightarrow{*} \varepsilon$$

注: 终结符不可空

➤ 一个串可空当且仅当由它可导出 $\varepsilon$

❖ 非终结符A可空的判定:

(1) 有规则  $A \rightarrow \varepsilon$ ,

或者

(2) 有规则  $A \rightarrow B_1 B_2 \dots B_n$ ,  
其中  $B_1, B_2, \dots, B_n$  **全部可空**.

# 可空性（例）

文法：

$$E \rightarrow T E'$$

$$E' \rightarrow + T E' \mid \varepsilon$$

$$T \rightarrow F T'$$

$$T' \rightarrow * F T' \mid \varepsilon$$

$$F \rightarrow (E) \mid \text{id} \mid \text{num}$$

$E'$  和  $T'$  可空。

$E$ ,  $T$  和  $F$  不可空

# FIRST函数（FIRST集）

❖ 对于任意文法符号串 $\alpha$ ， $\text{FIRST}(\alpha)$ 是可从 $\alpha$ 推导得到的串的首符号的集合。

➤ 如果 $\alpha \xRightarrow{*} \varepsilon$ ，则 $\varepsilon$ 也在 $\text{FIRST}(\alpha)$ 中（即 $\alpha$ 可空）

➤  $\text{FIRST}(\alpha) = \{t \mid \alpha \xRightarrow{*} t\beta, t \in T\} \cup \{\varepsilon \mid \alpha \xRightarrow{*} \varepsilon\}$

# 计算文法符号X的FIRST(X)

❖ 反复执行以下步骤，直至无法加入新的终结符或  $\epsilon$

1. 若X是终结符, 则 $\text{FIRST}(X) = \{X\}$ .

2. 若X是非终结符且有规则 $X \rightarrow Y_1 Y_2 \dots Y_k$ , 则将 $\text{FIRST}(Y_1 Y_2 \dots Y_k)$ 中的元素加入 $\text{FIRST}(X)$ ;

3. 若 $X \rightarrow \epsilon$ , 则将 $\epsilon$ 加入 $\text{FIRST}(X)$

# 计算文法符号串 $X_1X_2\dots X_n$ 的FIRST集

1. 初始化 $\text{FIRST}(X_1X_2\dots X_n)$ 为空集
2. 将 $\text{FIRST}(X_1)$ 中的所有终结符加入 $\text{FIRST}(X_1X_2\dots X_n)$
3. **for**( $i=1;i<n-1;i++$ )  
    如果 $X_i$ 可空, 将 $\text{FIRST}(X_{i+1})$ 中的所有终结符加入,  
    **否则break;**
4. 如果所有 $X_i$ 可空, 则将  $\varepsilon$  加入 $\text{FIRST}(X_1X_2\dots X_n)$

# FIRST集的计算(例)

❖ 假设 $x_1, x_2, x_3$ 可空,  $y_1, y_2$ 不可空

➤  $\text{FIRST}(y_1x_1y_2x_2)=\text{FIRST}(y_1)$

➤  $\text{FIRST}(x_1y_1x_2y_2)=(\text{FIRST}(x_1)-\{\epsilon\}) \cup \text{FIRST}(y_1)$

➤  $\text{FIRST}(x_1x_2y_1y_2)=(\text{FIRST}(x_1)-\{\epsilon\}) \cup (\text{FIRST}(x_2)-\{\epsilon\}) \cup \text{FIRST}(y_1)$

➤  $\text{FIRST}(x_1x_2x_3)=(\text{FIRST}(x_1)-\{\epsilon\}) \cup (\text{FIRST}(x_2)-\{\epsilon\}) \cup (\text{FIRST}(x_3)-\{\epsilon\}) \cup \{\epsilon\}$   
 $=(\text{FIRST}(x_1)-\{\epsilon\}) \cup (\text{FIRST}(x_2)-\{\epsilon\}) \cup \text{FIRST}(x_3)$

# 计算文法符号X的FIRST集(例)

文法:  $E \rightarrow TE'$        $E' \rightarrow + TE' \mid \varepsilon$        $T \rightarrow FT'$

$T' \rightarrow * FT' \mid \varepsilon$        $F \rightarrow (E) \mid id$

则:

$$\text{FIRST}(E) = \text{FIRST}(TE') = \{ (, id \}$$

$$\text{FIRST}(E') = \text{FIRST}(+TE') \cup \text{FIRST}(\varepsilon) = \{ \varepsilon, + \}$$

$$\text{FIRST}(T) = \text{FIRST}(FT') = \{ (, id \}$$

$$\text{FIRST}(T') = \text{FIRST}(*FT') \cup \text{FIRST}(\varepsilon) = \{ \varepsilon, * \}$$

$$\text{FIRST}(F) = \text{FIRST}((E)) \cup \text{FIRST}(id) = \{ (, id \}$$

# 计算FIRST集（例2）

例：文法G[S]为：

|                             |                    |                             |                   |
|-----------------------------|--------------------|-----------------------------|-------------------|
| $S \rightarrow AB$          | $S \rightarrow bC$ | $A \rightarrow \varepsilon$ | $A \rightarrow b$ |
| $B \rightarrow \varepsilon$ | $B \rightarrow aD$ | $C \rightarrow AD$          | $C \rightarrow b$ |
| $D \rightarrow aS$          | $D \rightarrow c$  |                             |                   |

解：

$$\text{FIRST}(S) = \text{FIRST}(AB) \cup \text{FIRST}(bC) = \{\varepsilon, a, b\}$$

$$\text{FIRST}(A) = \{\varepsilon, b\}$$

$$\text{FIRST}(B) = \{\varepsilon, a\}$$

$$\text{FIRST}(C) = \text{FIRST}(AD) \cup \text{FIRST}(b) = \{a, b, c\}$$

$$\text{FIRST}(D) = \{a, c\}$$

文法

- ①  $S \rightarrow aBC$
- ②  $B \rightarrow bC$
- ③  $B \rightarrow dB$
- ④  $B \rightarrow \varepsilon$
- ⑤  $C \rightarrow c$
- ⑥  $C \rightarrow a$

输入:  $ada$

最左推导过程:

$S$   
 $\Rightarrow aBC$   
 $\Rightarrow adBC$   
 $\Rightarrow adC$   
 $\Rightarrow ada$

$ade$

$S$   
 $\Rightarrow aBC$   
 $\Rightarrow adBC$   
 $\Rightarrow adC$

什么时候使用  $B \rightarrow \varepsilon$

# FOLLOW函数（FOLLOW集）

❖ 对于文法G[S]的非终结符号A:

$t \in \text{FOLLOW}(A)$  当且仅当:

$$S \overset{*}{\Rightarrow} \dots At \dots$$

这里t为**终结符**或输入结束标志**\$**

# 计算非终结符的FOLLOW集

- ❖ 反复执行以下规则,直到无法加入新的终结符和\$
- 1. 如果S是开始符,将\$加入FOLLOW(S),这里\$是输入结束标记.
- 2. 如果有规则  $A \rightarrow \alpha B \beta$ ,则将 FIRST( $\beta$ ) 中的终结符加入 FOLLOW(B).
- 3. 如果有:
  - (1)规则  $A \rightarrow \alpha B$ , 或者
  - (2)规则  $A \rightarrow \alpha B \beta$ , 且 $\beta$ 可空则将FOLLOW(A)加入 FOLLOW(B).

# FOLLOW集的计算（说明）

❖ 如果有规则  $A \rightarrow \alpha B \beta$

- 当 $\beta$ 不可空时， $FOLLOW(B)$ 中加入 $FIRST(\beta)$
- 当 $\beta$ 可空时， $FOLLOW(B)$ 中加入

$(FIRST(\beta) - \{\epsilon\}) \cup FOLLOW(A)$



Helpful  
Tips

**FIRST集看左，**

**FOLLOW集看右**

**FIRST集可能有 $\epsilon$ ，不可能有 $\$$**

**FOLLOW集可能有 $\$$ ，不可能有 $\epsilon$**

# FOLLOW集的计算（例）

$$\text{FOLLOW}(E) = \{\$, \})\}$$

$$\text{FOLLOW}(E') = \text{FOLLOW}(E) = \{\$, \})\}$$

$$\text{FOLLOW}(T) = (\text{FIRST}(E') - \{\epsilon\}) \cup \text{FOLLOW}(E') \cup \text{FOLLOW}(E) = \{\$, +, \})\}$$

$$\text{FOLLOW}(T') = \text{FOLLOW}(T) = \{\$, +, \})\}$$

$$\text{FOLLOW}(F) = (\text{FIRST}(T') - \{\epsilon\}) \cup \text{FOLLOW}(T) \cup \text{FOLLOW}(T') = \{\$, +, *, \})\}$$

文法G:  $E \rightarrow TE'$        $E' \rightarrow + TE'$        $E' \rightarrow \epsilon$        $T \rightarrow FT'$

$T' \rightarrow *FT'$        $T' \rightarrow \epsilon$        $F \rightarrow (E)$        $F \rightarrow id$

$\text{FIRST}(E) = \{(, id\}$        $\text{FIRST}(E') = \{\epsilon, +\}$        $\text{FIRST}(T) = \{(, id\}$

$\text{FIRST}(T') = \{\epsilon, *\}$        $\text{FIRST}(F) = \{(, id\}$

# FOLLOW集的计算 (例2)

$$\text{FOLLOW}(S) = \{\$ \} \cup \text{FOLLOW}(D) = \{\$ \}$$

$$\text{FOLLOW}(A) = (\text{FIRST}(B) - \{\epsilon\}) \cup \text{FOLLOW}(S) \cup \text{FIRST}(D) = \{a, c, \$\}$$

$$\text{FOLLOW}(B) = \text{FOLLOW}(S) = \{\$ \}$$

$$\text{FOLLOW}(C) = \text{FOLLOW}(S) = \{\$ \}$$

$$\text{FOLLOW}(D) = \text{FOLLOW}(B) \cup \text{FOLLOW}(C) = \{\$ \}$$

|                  |                |                |      |     |
|------------------|----------------|----------------|------|-----|
| 文法G[S] : S→AB    | S→bC           | A→ε            | A→b  | B→ε |
| B→aD             | C→AD           | C→b            | D→aS | D→c |
| FIRST(S)={ε,a,b} | FIRST(A)={ε,b} | FIRST(B)={ε,a} |      |     |
| FIRST(C)={a,b,c} | FIRST(D)={a,c} |                |      |     |

文法

- ①  $S \rightarrow aBC$
- ②  $B \rightarrow bC$
- ③  $B \rightarrow dB$
- ④  $B \rightarrow \varepsilon$
- ⑤  $C \rightarrow c$
- ⑥  $C \rightarrow a$

输入:  $ada$

最左推导过程:

$S$   
 $\Rightarrow aBC$   
 $\Rightarrow adBC$   
 $\Rightarrow adC$   
 $\Rightarrow ada$

$ade$

$S$   
 $\Rightarrow aBC$   
 $\Rightarrow adBC$   
报错

$\text{FIRST}(S)=\{a\}$       $\text{FIRST}(B)=\{b,d,\varepsilon\}$       $\text{FIRST}(C)=\{a,c\}$

$\text{FOLLOW}(S)=\{\$\}$       **$\text{FOLLOW}(B)=\{a,c\}$**

$\text{FOLLOW}(C)=\{\$,a,c\}$

# LL(1)文法

❖ 一个文法G是LL(1)的，当且仅当G的任意两个左部相同的产生式 $A \rightarrow \alpha | \beta$ 满足下面条件

➤  $\alpha$ 和 $\beta$ 不会导出以相同终结符开头的串

➤  $\alpha$ 和 $\beta$ 中最多只有一个可空

➤ 如果 $\beta$ 可空，那么 $\alpha$ 不能推导出以FOLLOW(A)中某个符号开头的串

# 产生式的可选集

- ❖ 产生式 $A \rightarrow \beta$ 的可选集是指可以选用该产生式时，对应的输入符号的集合，记为 $SELECT(A \rightarrow \beta)$ .

$$SELECT(A \rightarrow \beta) = FIRST( FIRST(\beta) FOLLOW(A) )$$

- $\epsilon \notin FIRST(\beta)$ 时， $SELECT(A \rightarrow \beta) = FIRST(\beta)$
- $\epsilon \in FIRST(\beta)$ 时， $SELECT(A \rightarrow \beta) = FOLLOW(A) \cup (FIRST(\beta) - \{\epsilon\})$
- 也称预测集(predict set)。

对于文法:  $S \rightarrow cT|RT$     $T \rightarrow De|\epsilon$     $R \rightarrow dR|\epsilon$     $D \rightarrow a|bS$

(1) 计算各非终结符的FIRST集和FOLLOW集

(2) 计算各规则的SELECT集

解:

$FIRST(S) = \{c\} \cup FIRST(RT) = \{c\} \cup (FIRST(R) - \{\epsilon\}) \cup FIRST(T) = \{a, b, c, d, \epsilon\}$

$FIRST(T) = FIRST(De) \cup \{\epsilon\} = FIRST(D) \cup \{\epsilon\} = \{a, b, \epsilon\}$

$FIRST(R) = \{\epsilon, d\}$        $FIRST(D) = \{a, b\}$

$FOLLOW(S) = \{\$ \}$     $FOLLOW(D) = \{e, \$ \}$

$FOLLOW(T) = FOLLOW(S) = \{e, \$ \}$

$FOLLOW(R) = (FIRST(T) - \{\epsilon\}) \cup FOLLOW(S) = \{a, b, e, \$ \}$

$FOLLOW(D) = \{e\}$

$S \rightarrow cT | RT$     $T \rightarrow De | \epsilon$     $R \rightarrow dR | \epsilon$     $D \rightarrow a | bS$

$FIRST(S) = \{a, b, c, d, \epsilon\}$     $FIRST(T) = \{a, b, \epsilon\}$     $FIRST(R) = \{\epsilon, d\}$     $FIRST(D) = \{a, b\}$

$FOLLOW(S) = \{e, \$\}$     $FOLLOW(T) = \{e, \$\}$

$FOLLOW(R) = \{a, b, e, \$\}$     $FOLLOW(D) = \{e\}$

---

$SELECT(S \rightarrow cT) = \{c\}$

$SELECT(S \rightarrow RT) = (FIRST(RT) - \{\epsilon\}) \cup FOLLOW(S) = \{a, b, d, e, \$\}$

$SELECT(T \rightarrow De) = FIRST(De) = \{a, b\}$

$SELECT(T \rightarrow \epsilon) = FOLLOW(T) = \{e, \$\}$

$SELECT(R \rightarrow dR) = \{d\}$

$SELECT(R \rightarrow \epsilon) = FOLLOW(R) = \{a, b, e, \$\}$

$SELECT(D \rightarrow a) = \{a\}$

$SELECT(D \rightarrow bS) = \{b\}$

# 构造预测分析表 (LL(1)分析表)

表的每一行对应一个非终结符，每一列对应一个输入符号（**终结符**和输入结束标志\$）

1. 为每条规则 $A \rightarrow \alpha$ ，重复步骤 2、3
2. 终结符 $a$ 在 $FIRST(\alpha)$ 中时，将 $A \rightarrow \alpha$ 加入 $M[A, a]$
3. 若 $\alpha$ 可空：对于 $b \in Follow(A)$ ,将 $A \rightarrow \alpha$ 加入 $M[A, b]$ 。
4. 分析表中**未定义项目**均为错误。
5. 若表中存在**多重定义项目**,则本文法非LL(1)。

# 构造预测分析表 (LL(1)分析表)

表的每一行对应一个非终结符，每一列对应一个输入符号（终结符和输入结束标志\$）

1. 为每条规则 $A \rightarrow \alpha$ ,

若终结符 $a$ 或 $\$$ 在 $SELECT(A \rightarrow \alpha)$ 中时，则将 $A \rightarrow \alpha$ 加入 $M[A, a]$ 或 $M[A, \$]$ 中。

2. 分析表中未定义项目均为错误。

3. 若表中存在多重定义项目,则本文法非LL(1)。

基于SELECT集描述

# 构造预测分析表（例）

| NON -<br>TERMINAL | INPUT SYMBOL              |                           |                       |                     |                           |                           |
|-------------------|---------------------------|---------------------------|-----------------------|---------------------|---------------------------|---------------------------|
|                   | id                        | +                         | *                     | (                   | )                         | \$                        |
| $E$               | $E \rightarrow TE'$       |                           |                       | $E \rightarrow TE'$ |                           |                           |
| $E'$              |                           | $E' \rightarrow +TE'$     |                       |                     | $E' \rightarrow \epsilon$ | $E' \rightarrow \epsilon$ |
| $T$               | $T \rightarrow FT'$       |                           |                       | $T \rightarrow FT'$ |                           |                           |
| $T'$              |                           | $T' \rightarrow \epsilon$ | $T' \rightarrow *FT'$ |                     | $T' \rightarrow \epsilon$ | $T' \rightarrow \epsilon$ |
| $F$               | $F \rightarrow \text{id}$ |                           |                       | $F \rightarrow (E)$ |                           |                           |

$E \rightarrow TE'$

$E' \rightarrow +TE'$

$E' \rightarrow \epsilon$

$T \rightarrow FT'$

$T' \rightarrow *FT'$

$T' \rightarrow \epsilon$

$F \rightarrow (E)$

$F \rightarrow \text{id}$

$\text{FOLLOW}(E) = \{\$, \}$

$\text{FOLLOW}(E') = \{\$, \}$

$\text{FOLLOW}(T) = \{\$, +, \}$

$\text{FOLLOW}(T') = \{\$, +, \}$

$\text{FOLLOW}(F) = \{\$, +, *, \}$

$\text{FIRST}(E) = \{(\, \text{id}\}$

$\text{FIRST}(E') = \{\epsilon, +\}$

$\text{FIRST}(T) = \{(\, \text{id}\}$

$\text{FIRST}(T') = \{\epsilon, *\}$

$\text{FIRST}(F) = \{(\, \text{id}\}$

对于上下文无关文法：

$S \rightarrow a S b B \mid g A$        $A \rightarrow C B \mid c A h$

$B \rightarrow d \mid f C \mid C$        $C \rightarrow \varepsilon$

(1) 写出各非终结符的FIRST集和FOLLOW集。

(2) 构建LL(1)分析表, 本文法是LL(1)文法吗, 为什么?

解:  $FIRST(S)=\{a,g\}$     $FIRST(A)=\{c,d,f,\varepsilon\}$     $FIRST(B)=\{d,f,\varepsilon\}$     $FIRST(C)=\{\varepsilon\}$

$FOLLOW(S)=\{b,\$ \}$     $FOLLOW(A)=\{b,h,\$ \}$

$FOLLOW(B)=\{b,h,\$ \}$     $FOLLOW(C)=\{b,h,d,f,\$ \}$

可构建LL(1)分析表如下:

|   | a                    | b                           | c                   | d                           | f                           | g                  | h                           | \$                          |
|---|----------------------|-----------------------------|---------------------|-----------------------------|-----------------------------|--------------------|-----------------------------|-----------------------------|
| S | $S \rightarrow aSbB$ |                             |                     |                             |                             | $S \rightarrow gA$ |                             |                             |
| A |                      | $A \rightarrow CB$          | $A \rightarrow cAh$ | $A \rightarrow CB$          | $A \rightarrow CB$          |                    | $A \rightarrow CB$          | $A \rightarrow CB$          |
| B |                      | $B \rightarrow C$           |                     | $B \rightarrow d$           | $B \rightarrow fC$          |                    | $B \rightarrow C$           | $B \rightarrow C$           |
| C |                      | $C \rightarrow \varepsilon$ |                     | $C \rightarrow \varepsilon$ | $C \rightarrow \varepsilon$ |                    | $C \rightarrow \varepsilon$ | $C \rightarrow \varepsilon$ |

分析表无多重定义项目, 故本文法LL(1)

改造以下文法，使其可能用于无回溯的自顶向下分析。（回顾）

$$S \rightarrow SS+ \mid SS^* \mid a$$

解：（1）提取左公共因子SS得：

$$S \rightarrow SSA \mid a \quad A \rightarrow +|^*$$

（2）设定文法符号顺序S, A。

仅S存在直接左递归，消除左递归得： $S \rightarrow aR$      $R \rightarrow SAR \mid \varepsilon$

最终文法为：

$$S \rightarrow aR \quad R \rightarrow SAR \mid \varepsilon \quad A \rightarrow +|^*$$

是不是LL(1)文法？