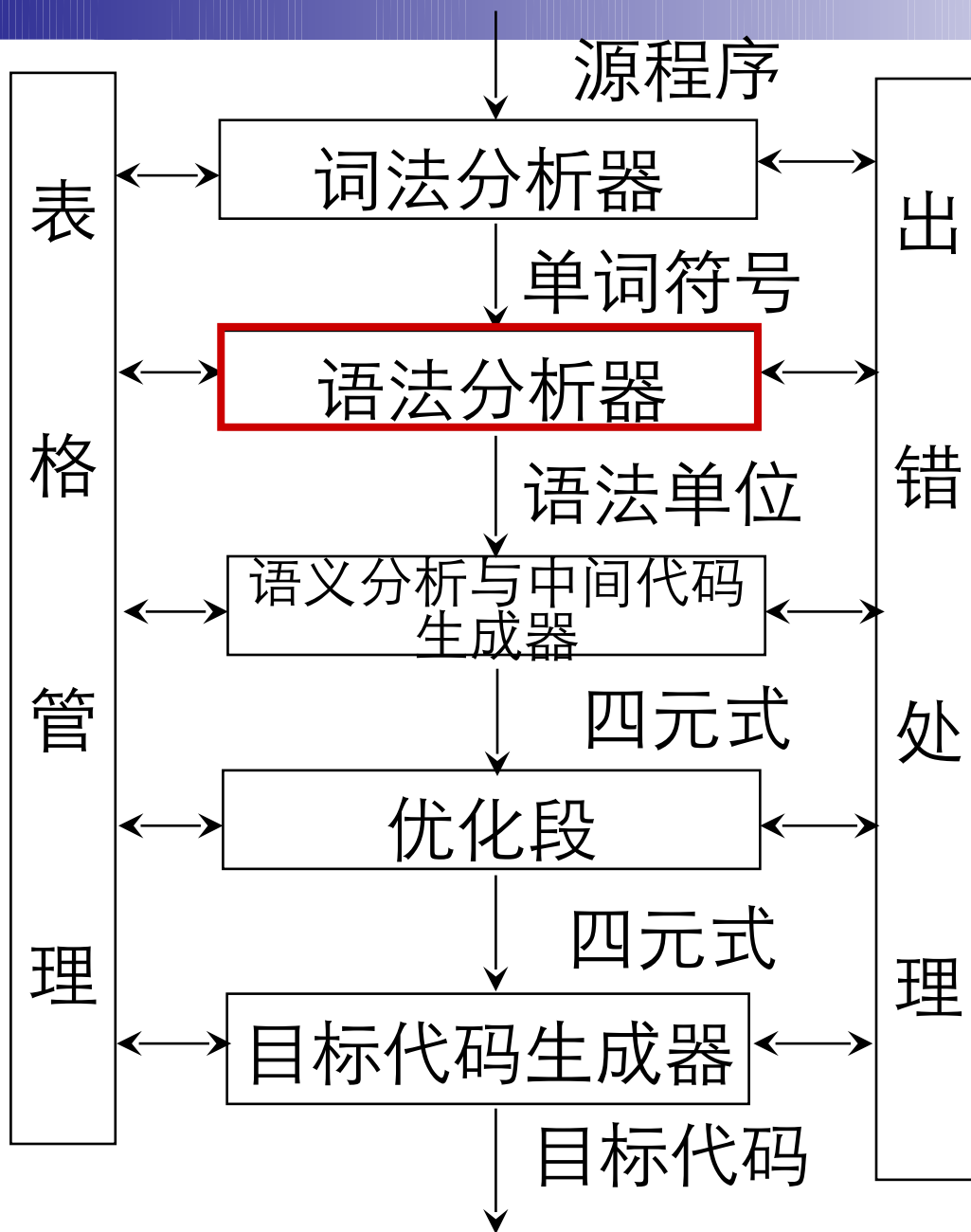




# 编译原理

## 第四章 语法分析——自上而下分析

# 编译程序总框



# 第四章 语法分析—自上而下分析

- 语法分析器的功能
- 自上而下分析面临的问题
- LL(1) 分析法
- 递归下降分析程序构造
- 预测分析程序

# 第四章 语法分析—自上而下分析

- 本章主要介绍语法分析的处理
- 语法分析的前提
  - 对语言的语法结构进行描述
    - 采用**正规式**和**有限自动机**可以描述和识别语言的**单词符号**
    - 用**上下文无关文法**来描述**语法规则**

# 上下文无关文法

- 一个上下文无关文法  $G$  是一个四元式

$G=(V_T, V_N, S, P)$ ，其中

- $V_T$ ：终结符集合（非空）
- $V_N$ ：非终结符集合（非空），且  $V_T \cap V_N = \emptyset$
- $S$ ：文法的开始符号， $S \in V_N$
- $P$ ：产生式集合（有限），每个产生式形式为
  - $P \rightarrow \alpha$ ， $P \in V_N$ ， $\alpha \in (V_T \cup V_N)^*$
- 开始符  $S$  至少必须在某个产生式的左部出现一次

# 上下文无关文法

- 定义：称  $\alpha A\beta$  **直接推出**  $\alpha\gamma\beta$ ，即

$$\alpha A\beta \Rightarrow \alpha\gamma\beta$$

仅当  $A \rightarrow \gamma$  是一个产生式，

且  $\alpha, \beta \in (V_T \cup V_N)^*$ 。

- 如果  $\alpha_1 \Rightarrow \alpha_2 \Rightarrow \dots \Rightarrow \alpha_n$ ，则我们称这个序列是从  $\alpha_1$  到  $\alpha_n$  的一个**推导**。若存在一个从  $\alpha_1$  到  $\alpha_n$  的推导，则称  $\alpha_1$  可以**推导出**  $\alpha_n$

# 上下文无关文法

- 定义：假定  $G$  是一个文法， $S$  是它的开始符号。如果  $S \xRightarrow{*} \alpha$  则  $\alpha$  称是一个句型。
- 仅含终结符号的句型是一个句子。
- 文法  $G$  所产生的句子的全体是一个语言，将它记为  $L(G)$ 。

$$L(G) = \{\alpha \mid S \xRightarrow{+} \alpha, \alpha \in V_T^*\}$$

# 第四章 语法分析—自上而下分析

- 语法分析器的功能
- 自上而下分析面临的问题
- LL(1) 分析法
- 递归下降分析程序构造
- 预测分析程序



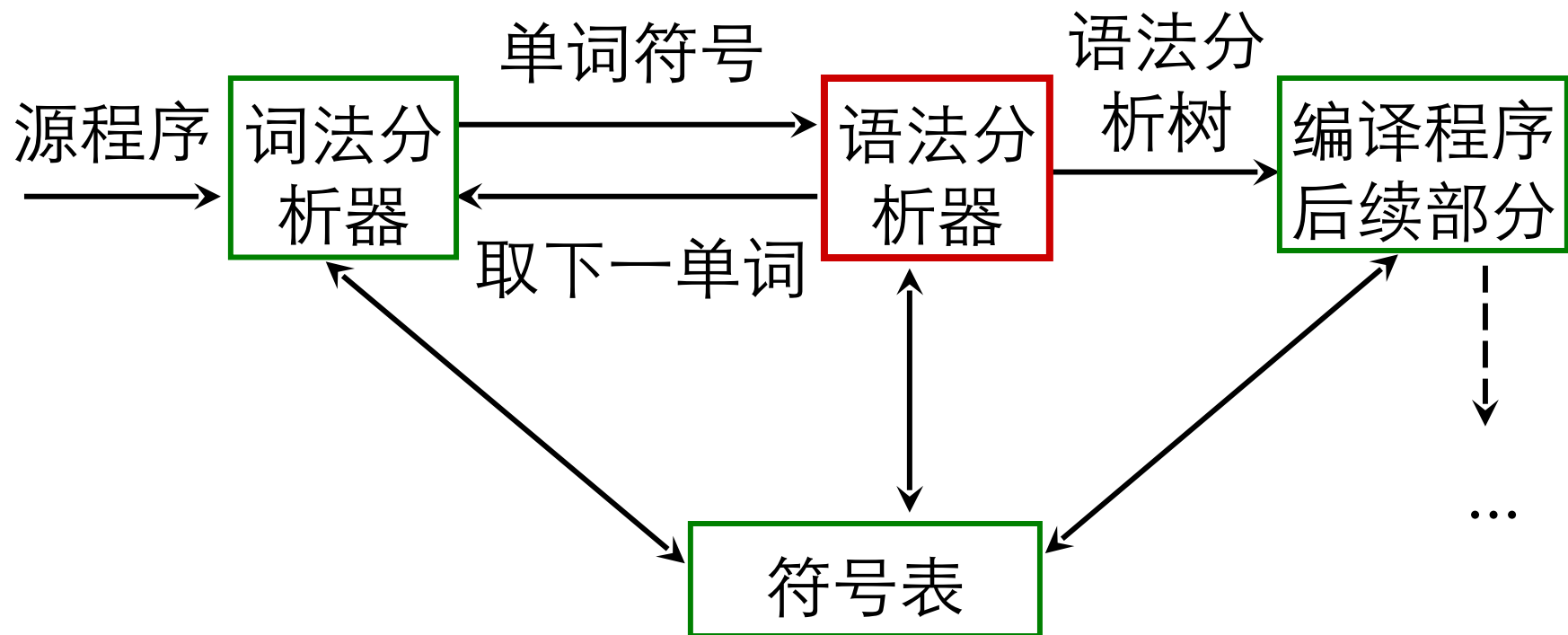
# 4.1 语法分析器的功能

- 语法分析的任务

- 分析一个文法的句子结构

- 语法分析器的功能

- 按照文法的产生式（语言的语法规则），识别输入符号串是否为一个句子（合式程序）



# 语法分析的方法

## ■ 自下而上分析法 (Bottom-up)

### □ 基本思想

- 从输入串开始，逐步进行归约，直到文法的开始符号
- 归约：根据文法的产生式规则，把产生式的右部替换成左部符号
- 从树末端开始，构造语法树

### □ 算符优先分析法

- 按照终结符的优先关系和结合性质进行语法分析
- 适合分析表达式

### □ LR 分析法

- 规范归约

# 语法分析的方法

- 自下而上分析法 (Bottom-up)

- 自上而下分析法 (Top-down)

- 基本思想

- 它从文法的开始符号出发，反复使用各种产生式，寻找"匹配"的**推导**

- **递归下降分析法**

- 对每一语法变量（非终结符）构造一个相应的子程序，每个子程序识别一定的语法单位
    - 通过子程序间的相互调用实现对输入串的认识

- **预测分析程序**

- 非递归实现
    - 直观、简单

# 第四章 语法分析—自上而下分析

- 语法分析器的功能
- 自上而下分析面临的问题
- LL(1) 分析法
- 递归下降分析程序构造
- 预测分析程序

## 4.2 自上而下分析面临的问题

- 自上而下就是从文法的开始符号出发，向下**推导**，推出句子
- 自上而下分析的主旨
  - 针对输入串，试图用一切可能的办法，从文法开始符号（根结点）出发，自上而下地为输入串建立一棵语法树
  - 为输入串寻找一个最左推导

- 例 3.4.1 假定有文法  $G(S)$ :

$$(1) S \rightarrow xAy$$

$$(2) A \rightarrow ** \mid *$$

分析输入串  $x^*y$  ( 记为  $\alpha$  ) 。

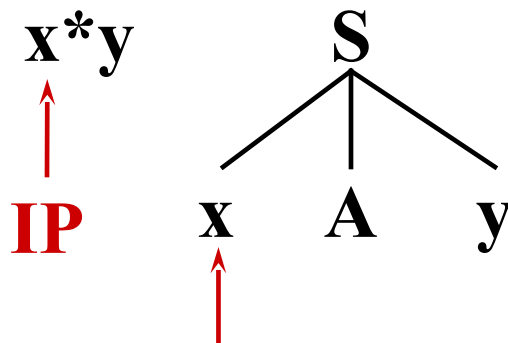
$x^*y$        $S$   
↑  
**IP**

■ 例 3.4.1 假定有文法  $G(S)$ :

(1)  $S \rightarrow xAy$

(2)  $A \rightarrow ** \mid *$

分析输入串  $x^*y$  ( 记为  $\alpha$  ) 。



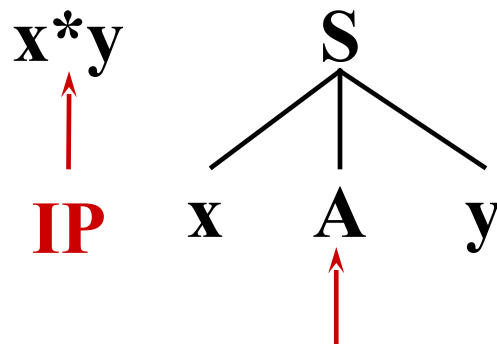


■ 例 3.4.1 假定有文法  $G(S)$ :

(1)  $S \rightarrow xAy$

(2)  $A \rightarrow ** \mid *$

分析输入串  $x^*y$  ( 记为  $\alpha$  ) 。

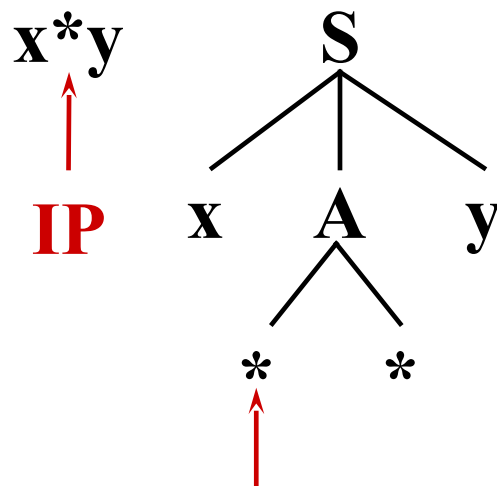


■ 例 3.4.1 假定有文法  $G(S)$ :

(1)  $S \rightarrow xAy$

(2)  $A \rightarrow ** \mid *$

分析输入串  $x^*y$  ( 记为  $\alpha$  ) 。

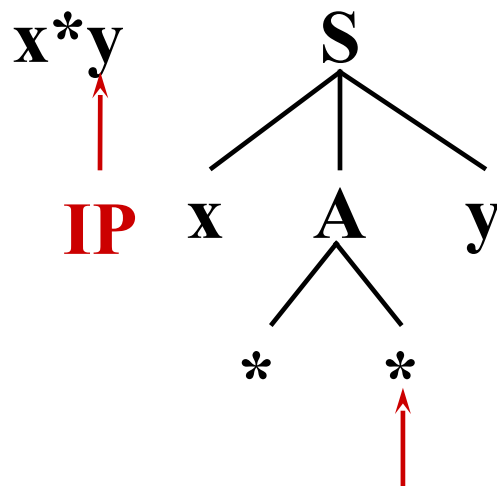


■ 例 3.4.1 假定有文法  $G(S)$ :

(1)  $S \rightarrow xAy$

(2)  $A \rightarrow ** \mid *$

分析输入串  $x^*y$  ( 记为  $\alpha$  ) 。

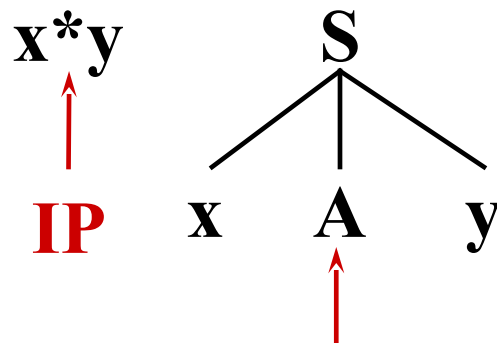


■ 例 3.4.1 假定有文法  $G(S)$ :

(1)  $S \rightarrow xAy$

(2)  $A \rightarrow ** \mid *$

分析输入串  $x^*y$  ( 记为  $\alpha$  ) 。

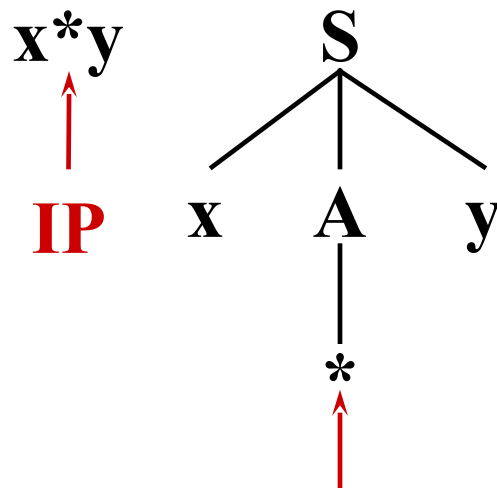


- 例 3.4.1 假定有文法  $G(S)$ :

(1)  $S \rightarrow xAy$

(2)  $A \rightarrow ** \mid *$

分析输入串  $x^*y$  ( 记为  $\alpha$  ) 。

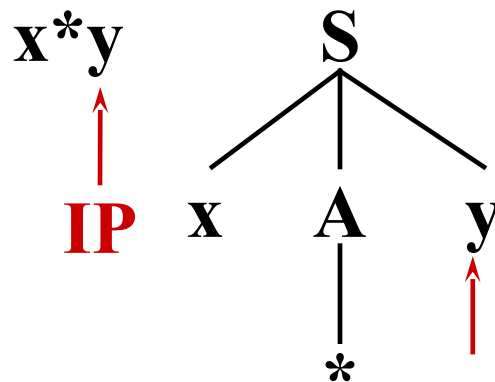


■ 例 3.4.1 假定有文法  $G(S)$ :

(1)  $S \rightarrow xAy$

(2)  $A \rightarrow ** \mid *$

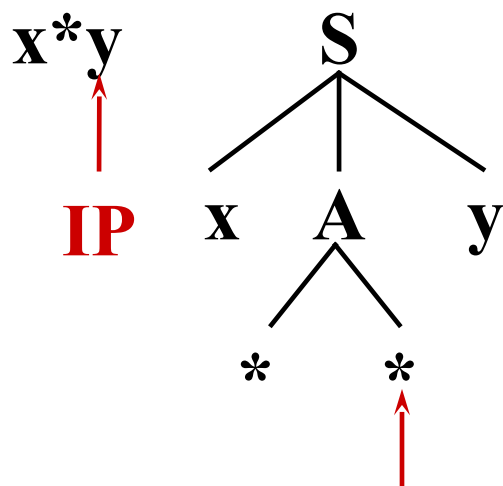
分析输入串  $x^*y$  ( 记为  $\alpha$  ) 。



# 多个产生式候选带来的问题

## ■ 回溯

- 分析过程中，当一个非终结符用某一个候选匹配成功时，这种匹配可能是暂时的
- 出错时，不得不“回溯”



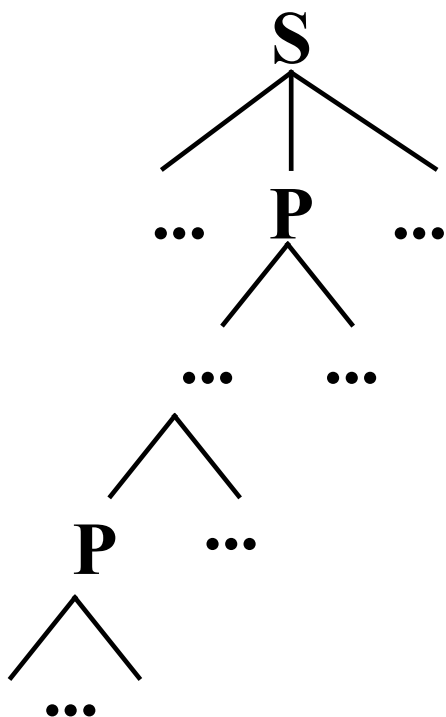
$A \rightarrow ** \mid *$

# 多个产生式候选带来的问题

## ■ 文法左递归问题

- 一个文法是含有左递归的，如果存在非终结符  $P$

$$P \xRightarrow{+} P\alpha$$



...a...

含有左递归的文法将使自上而下的分析陷入无限循环！



# 第四章 语法分析—自上而下分析

- 语法分析器的功能
- 自上而下分析面临的问题
- LL(1) 分析法
- 递归下降分析程序构造
- 预测分析程序

## 4.3 LL(1) 分析法

- 构造不带回溯的自上而下分析算法
  - 要消除文法的左递归性
  - 克服回溯

# 左递归的消除

$$\begin{aligned} P &\Rightarrow P \alpha \\ &\Rightarrow P \alpha \alpha \\ &\dots\dots \\ &\Rightarrow P \alpha \dots \alpha \\ &\Rightarrow \beta \alpha \dots \alpha \end{aligned}$$

- 直接消除见诸于产生式中的左递归：假定关于非终结符  $P$  的规则为

$$P \rightarrow P\alpha \mid \beta$$

其中  $\beta$  不以  $P$  开头。

我们可以把  $P$  的规则等价地改写为如下的非直接左递归形式：

$$\begin{aligned} P &\rightarrow \beta P' \\ P' &\rightarrow \alpha P' \mid \varepsilon \end{aligned}$$

左递归变  
右递归

$$\begin{aligned} P &\Rightarrow \beta P' \\ &\Rightarrow \beta \alpha P' \\ &\Rightarrow \beta \alpha \alpha P' \\ &\dots\dots \\ &\Rightarrow \beta \alpha \dots \alpha P' \\ &\Rightarrow \beta \alpha \dots \alpha \end{aligned}$$

# 左递归的消除

- 一般而言，假定  $P$  关于的全部产生式是

$$P \rightarrow P\alpha_1 \mid P\alpha_2 \mid \cdots \mid P\alpha_m \mid \beta_1 \mid \beta_2 \mid \cdots \mid \beta_n$$

其中，每个  $\alpha$  都不等于  $\varepsilon$ ，每个  $\beta$  都不以  $P$  开头

那么，消除  $P$  的直接左递归性就是把这些规则改写成：

$$P \rightarrow \beta_1 P' \mid \beta_2 P' \mid \cdots \mid \beta_n P'$$

$$P' \rightarrow \alpha_1 P' \mid \alpha_2 P' \mid \cdots \mid \alpha_m P' \mid \varepsilon$$

左递归变  
右递归

$$P \rightarrow P\alpha_1 \mid P\alpha_2 \mid \cdots \mid P\alpha_m \mid \beta_1 \mid \beta_2 \mid \cdots \mid \beta_n$$

变换成:

$$P \rightarrow \beta_1 P' \mid \beta_2 P' \mid \cdots \mid \beta_n P'$$

$$P' \rightarrow \alpha_1 P' \mid \alpha_2 P' \mid \cdots \mid \alpha_m P' \mid \varepsilon$$

■ 例 文法  $G(E)$ :

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

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

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

经消去直接左递归后变成:

$$E \rightarrow TE'$$

$$E' \rightarrow +TE' \mid \varepsilon$$

$$T \rightarrow FT'$$

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

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

(4.2)

■ 例 文法  $G(S)$ :

$$S \rightarrow Qc \mid c$$

$$Q \rightarrow Rb \mid b$$

$$R \rightarrow Sa \mid a \quad (4.3)$$

虽没有直接左递归，但  $S$ 、 $Q$ 、 $R$  都是左递归的

$$S \Rightarrow Qc \Rightarrow Rbc \Rightarrow Sabc$$

■ 一个文法消除左递归的条件

□ 不含以  $\varepsilon$  为右部的产生式

□ 不含回路

$$P \overset{+}{\Rightarrow} P$$

■ 例 文法  $G(S)$ :

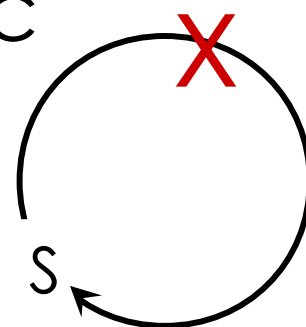
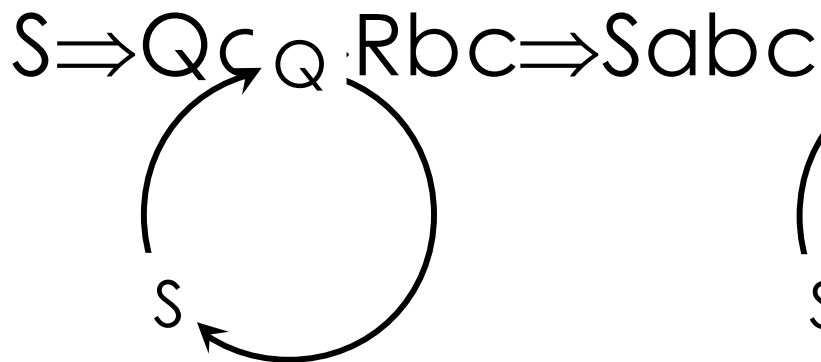
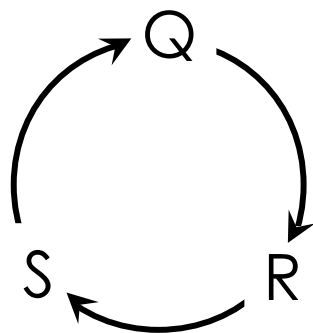
$S \rightarrow Qc \mid c$

$Q \rightarrow Rb \mid b$

$R \rightarrow Sa \mid a$

(4.3)

虽没有直接左递归，但  $S$ 、 $Q$ 、 $R$  都是左递归的



$S \rightarrow Qc \mid c$   
 $Q \rightarrow Rb \mid b$   
 $R \rightarrow Sa \mid a$

$S \rightarrow Qc \mid c$   
 $Q \rightarrow Sab \mid ab \mid b$   
 $R \rightarrow Sa \mid a$

$S \rightarrow Sabc \mid abc \mid bc \mid c$   
 $Q \rightarrow Sab \mid ab \mid b$   
 $R \rightarrow Sa \mid a$

# 消除左递归的算法

1. 把文法  $G$  的所有非终结符按任一种顺序排列成  $P_1, P_2, \dots, P_n$ ; 按此顺序执行:

2. FOR  $i:=1$  TO  $n$  DO

BEGIN

FOR  $j:=1$  TO  $i-1$  DO

把形如  $P_i \rightarrow P_j \gamma$  的规则改写成

$P_i \rightarrow \delta_1 \gamma \mid \delta_2 \gamma \mid \dots \mid \delta_k \gamma$ ;

( 其中  $P_j \rightarrow \delta_1 \mid \delta_2 \mid \dots \mid \delta_k$  是关于  $P_j$  的所有规则 )

消除关于  $P_i$  规则的直接左递归性

END

3. 化简由 2 所得的文法。去除那些从开始符号出发



■ 例 考虑文法  $G(S)$

$S \rightarrow Qc \mid c$

$Q \rightarrow Rb \mid b$

$R \rightarrow Sa \mid a$

■ 令它的非终结符的排序为  $R$ 、 $Q$ 、 $S$

■ 对于  $R$ ，不存在直接左递归

■ 把  $R$  代入到  $Q$  的有关候选后，把  $Q$  的规则变为

$Q \rightarrow Sab \mid ab \mid b$

$S \rightarrow Qc \mid c$

$Q \rightarrow Sab \mid ab \mid b$

$R \rightarrow Sa \mid a$

- 例 考虑文法  $G(S)$

$S \rightarrow Qc \mid c$

$Q \rightarrow Rb \mid b$

$R \rightarrow Sa \mid a$

$S \rightarrow Sabc \mid abc \mid bc \mid c$

$Q \rightarrow Sab \mid ab \mid b$

$R \rightarrow Sa \mid a$

- 令它的非终结符的排序为  $R$ 、 $Q$ 、 $S$

- $Q$  的规则变为

$Q \rightarrow Sab \mid ab \mid b$

- 现在的  $Q$  不含直接左递归，把它代入到  $S$  的有关候选后， $S$  变成

$S \rightarrow Sabc \mid abc \mid bc \mid c$

■ 例 考虑文法  $G(S)$

$S \rightarrow Qc \mid c$

$Q \rightarrow Rb \mid b$

$R \rightarrow Sa \mid a$

$S \rightarrow Sabc \mid abc \mid bc \mid c$

$Q \rightarrow Sab \mid ab \mid b$

$R \rightarrow Sa \mid a$

■  $S$  变成

$S \rightarrow Sabc \mid abc \mid bc \mid c$

■ 消除  $S$  的直接左递归后:

$S \rightarrow abcS' \mid bcS' \mid cS'$

$S' \rightarrow abcS' \mid \varepsilon$

$Q \rightarrow Sab \mid ab \mid b$

$R \rightarrow Sa \mid a$

■ 例 考虑文法  $G(S)$

$$S \rightarrow Qc \mid c$$

$$Q \rightarrow Rb \mid b$$

$$R \rightarrow Sa \mid a$$

■ 消除  $S$  的直接左递归后：

$$S \rightarrow abcS' \mid bcS' \mid cS'$$

$$S' \rightarrow abcS' \mid \varepsilon$$

$$Q \rightarrow Sab \mid ab \mid b$$

$$R \rightarrow Sa \mid a$$

■ 关于  $Q$  和  $R$  的规则已是多余的，化简为：

$$S \rightarrow abcS' \mid bcS' \mid cS'$$

$$S' \rightarrow abcS' \mid \varepsilon \quad (4.4)$$

- 注意，由于对非终结符排序的不同，最后所得的文法在形式上可能不一样。但不难证明，它们都是等价的。
- 例如，若对文法 (4.3) 的非终结符排序选为  $S$ 、 $Q$ 、 $R$ ，那么，最后所得的无左递归文法是：

$$\begin{aligned} S &\rightarrow Qc \mid c \\ Q &\rightarrow Rb \mid b \end{aligned}$$

$$\begin{aligned} R &\rightarrow bcaR' \mid caR' \mid aR' \\ (4.5) \end{aligned}$$

$$R' \rightarrow bcaR' \mid \varepsilon$$

$$\begin{aligned} S &\rightarrow abcS' \mid bcS' \mid cS' \\ S' &\rightarrow abcS' \mid \varepsilon \end{aligned} \quad (4.4)$$

- 文法 (4.4) 和 (4.5) 的等价性是显然的。

# 小结

- 自上而下分析面临的问题
  - 文法的左递归性
  - 回溯
- 消除文法的左递归
  - 直接左递归的消除
  - 间接左递归的消除