



中山大學
SUN YAT-SEN UNIVERSITY



国家超级计算广州中心
NATIONAL SUPERCOMPUTER CENTER IN GUANGZHOU

DCS²⁹⁰

Compilation Principle

编译原理

第四章 语法分析 (2)

郑馥丹

zhengfd5@mail.sysu.edu.cn

CONTENTS

目录

01

自顶向下分析
Top-Down Parsing

02

LL(1)分析
LL(1) Parsing

03

算符优先分析
Operator
Precedence Parsing

04

LR分析
LR Parsing

4. 确定的自顶向下语法分析[predictive]

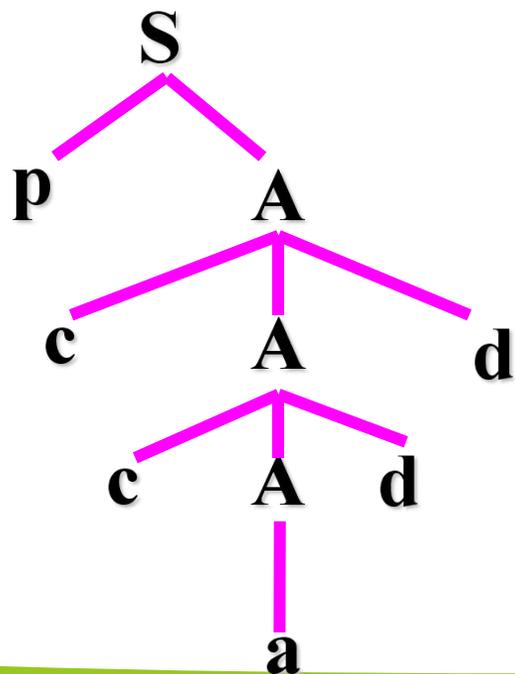
• 例：设有文法G[S]:

$$S \rightarrow pA|qB$$

$$A \rightarrow cAd|a$$

$$B \rightarrow dB|b$$

若输入串W=pccadd，自顶向下的推导过程为：



$$\underline{S} \Rightarrow p\underline{A} \Rightarrow pc\underline{A}d \Rightarrow pcc\underline{A}dd \Rightarrow pccadd$$

该推导过程是确定的！

原因：

(1) 每个产生式的右部由终结符开头

(2) 同一个非终结符的不同产生式的右部由不同的终结符开头。

因此，在推导过程中可以根据当前的输入符号**唯一确定选哪个产生式**往下推导，分析过程是确定的。

4. 确定的自顶向下语法分析[predictive]

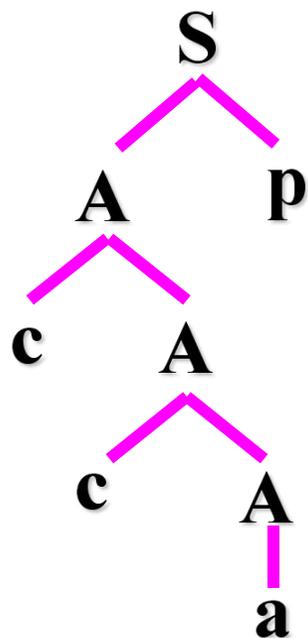
- 例：设有文法G[S]:

$$S \rightarrow Ap | Bq$$

$$A \rightarrow a | cA$$

$$B \rightarrow b | dB$$

若输入串W=ccap, 自顶向下的推导过程为:



$$\underline{S} \Rightarrow \underline{A}p \Rightarrow c\underline{A}p \Rightarrow cc\underline{A}p \Rightarrow cc\underline{a}p$$

该推导过程可能是确定的

当:

- (1) 产生式右部以终结符或非终结符开头 (**无空产生式**) ;
- (2) **同一非终结符的不同产生式的右部由不同的符号开头。**

对于这种文法, 在推导过程选用哪个产生式不直观, 关键是判断产生式右部推出的**开始符号集——FIRST集**, 分析过程可能是确定的。

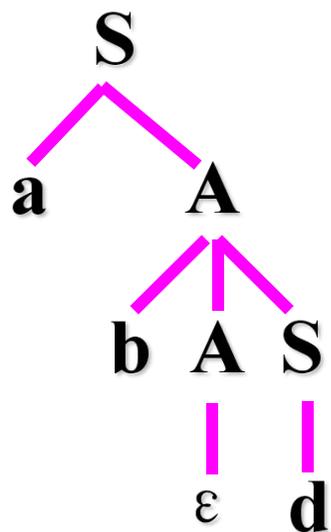
4. 确定的自顶向下语法分析[predictive]

- 例：设有文法G[S]

$$S \rightarrow aA | d$$

$$A \rightarrow bAS | \varepsilon$$

若输入串W=abd，自顶向下的推导过程为：



$$\underline{S} \Rightarrow \underline{a} \underline{A} \Rightarrow \underline{ab} \underline{A} \underline{S} \Rightarrow \underline{ab} \underline{S} \Rightarrow \underline{abd}$$

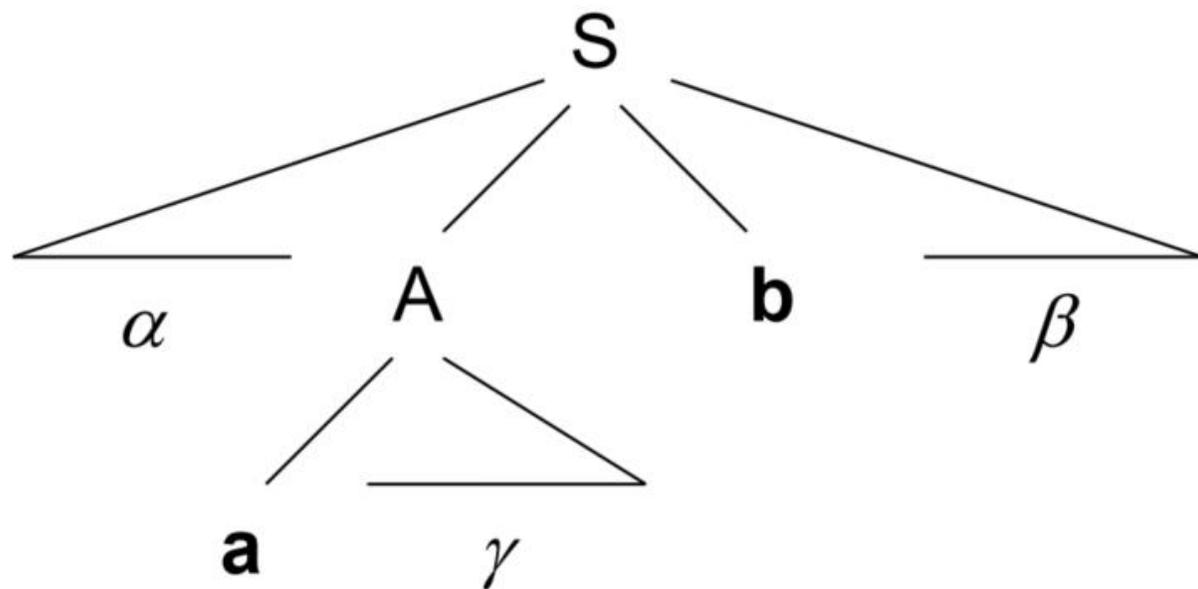
该推导过程**可能**是确定的

文法的特点是：**包含空产生式**。
对于空产生式左部的非终结符，
关键是判断该非终结符的**后跟
符号集——FOLLOW集**，分
析过程可能是确定的。

4. 确定的自顶向下语法分析[predictive]

• **FIRST集**和**FOLLOW集**的直观理解

$\mathbf{a} \in \text{FIRST}(A)$ and $\mathbf{b} \in \text{FOLLOW}(A)$

• 开始符号集——**FIRST集**

– 设 $G=(V_N, V_T, P, S)$ 是上下文无关文法, $\beta \in (V_N \cup V_T)^*$,

$\text{FIRST}(\beta) = \{ \mathbf{a} \in V_T \mid \beta \Rightarrow^* \mathbf{a} \dots \dots \}$

– 若 $\beta \Rightarrow^* \epsilon$ 则规定 $\epsilon \in \text{FIRST}(\beta)$

– 直观上说, 文法符号串 β 的**开始符号集是由 β 推导出的开头的终结符 (包括 ϵ) 组成的。**

4. 确定的自顶向下语法分析[predictive]

- 开始符号集——**FIRST集**

例：文法G[S]：

$S \rightarrow Ap$ $\text{FIRST}(Ap) = \{a, c\}$

$S \rightarrow Bq$ $\text{FIRST}(Bq) = \{b, d\}$

$A \rightarrow a$ $\text{FIRST}(a) = \{a\}$

$A \rightarrow cA$ $\text{FIRST}(cA) = \{c\}$

$B \rightarrow b$ $\text{FIRST}(b) = \{b\}$

$B \rightarrow dB$ $\text{FIRST}(dB) = \{d\}$

可进行确定的自顶向下分析

由于**同一非终结符的两个产生式**的右部推导出来的**开始符号集不相交**，因此可根据当前输入符属于哪个产生式右部的开始符号集而决定**唯一选哪个产生式进行推导**，可以进行确定的自顶向下分析。

4. 确定的自顶向下语法分析[predictive]

- 开始符号集——**FIRST集**

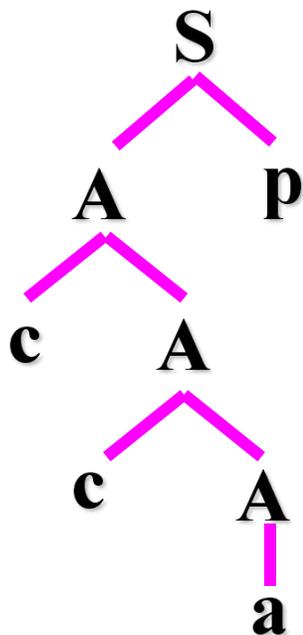
若输入串 $W=ccap$,自顶向下的推导过程为:

例: 文法 $G[S]$:

$S \rightarrow Ap|Bq$

$A \rightarrow a|cA$

$B \rightarrow b|dB$



<u>$S \rightarrow Ap$</u>	$FIRST(Ap) = \{a, c\}$
<u>$S \rightarrow Bq$</u>	$FIRST(Bq) = \{b, d\}$
<u>$A \rightarrow a$</u>	$FIRST(a) = \{a\}$
<u>$A \rightarrow cA$</u>	$FIRST(cA) = \{c\}$
$B \rightarrow b$	$FIRST(b) = \{b\}$
$B \rightarrow dB$	$FIRST(dB) = \{d\}$

$\underline{S} \Rightarrow \underline{A}p \Rightarrow c\underline{A}p \Rightarrow cc\underline{A}p \Rightarrow ccap$

4. 确定的自顶向下语法分析[predictive]

• 开始符号集——**FIRST集**

– 在求各个非终结符的FIRST集之前，先确定它们是否能 $\Rightarrow^* \epsilon$

✓ 第1次扫描——扫描文法中的产生式

- 对能直接推出 ϵ 的产生式左部的终结符标“是”。
- 删除所有右部含有终结符的产生式。若以某一非终结符为左部的所有产生式都被删除，则该非终结符不能推出 ϵ ，将其标为“否”。

例：G[S]

S \rightarrow AB	B \rightarrow ϵ
S \rightarrow bC	C \rightarrow AD
A \rightarrow b	C \rightarrow b
A \rightarrow ϵ	D \rightarrow aS
B \rightarrow aD	D \rightarrow c

非终结符	S	A	B	C	D
第 1 次扫描		是	是		否
第 2 次扫描					

4. 确定的自顶向下语法分析[predictive]

• 开始符号集——**FIRST集**– 先确定各个非终结符是否能 $\Rightarrow^* \epsilon$

✓ 第2次扫描——扫描产生式右部的符号

- 对每个产生式 $p: A \rightarrow X_1 \dots X_n$ ，如果 X_1, \dots, X_n 都被标为“是”（即 X_1, \dots, X_n 都能推出 ϵ ），则A也能推出 ϵ ，将其标为“是”。
- 如果 $A \rightarrow Y_1 \dots Y_n$ 中， $Y_1 \dots Y_n$ 中任一个已被标为“否”，则删掉该产生式，若这么做使得A的所有产生式都被删去，则A不能推出 ϵ ，将其标为“否”。

$S \rightarrow AB$	$B \rightarrow \epsilon$
$S \rightarrow bC$	$C \rightarrow AD$
$A \rightarrow b$	$C \rightarrow b$
$A \rightarrow \epsilon$	$D \rightarrow aS$
$B \rightarrow aD$	$D \rightarrow c$

非终结符	S	A	B	C	D
第 1 次扫描		是	是		否
第 2 次扫描	是			否	

4. 确定的自顶向下语法分析[predictive]

• 开始符号集——FIRST集

– 对**每一文法符号** $X(X \in V_T \cup V_N)$, 求**FIRST(X)**的算法:

✓ 对每个 $a \in V_T$: $\text{FIRST}(a) = \{a\}$

✓ 对每个 $A \in V_N$: 若 $A \Rightarrow^* \varepsilon$, 则 $\varepsilon \in \text{FIRST}(A)$

✓ 对每个 $A \in V_N$: 若 $A \rightarrow a \dots$, $a \in V_T$, 则 $a \in \text{FIRST}(A)$

✓ 若 X, Y_1, Y_2, \dots, Y_n 都 $\in V_N$, 有产生式 $X \rightarrow Y_1 Y_2 \dots Y_n$, 当 Y_1, Y_2, \dots, Y_{n-1} 都 $\Rightarrow^* \varepsilon$ 时, $\text{FIRST}(Y_1) - \{\varepsilon\}, \text{FIRST}(Y_2) - \{\varepsilon\}, \dots, \text{FIRST}(Y_{n-1}) - \{\varepsilon\}, \text{FIRST}(Y_n)$ 都包含在 $\text{FIRST}(X)$ 中

✓ 当所有 $Y_i \Rightarrow^* \varepsilon$, 则 $\text{FIRST}(X) = \text{FIRST}(Y_1) \cup \text{FIRST}(Y_2) \cup \dots \cup \text{FIRST}(Y_n) \cup \{\varepsilon\}$

4. 确定的自顶向下语法分析[predictive]

- 开始符号集——**FIRST集**

例G[S]:

$S \rightarrow AB$ $S \rightarrow bC$

$A \rightarrow b$ $A \rightarrow \epsilon$

$B \rightarrow aD$ $B \rightarrow \epsilon$

$C \rightarrow AD$ $C \rightarrow b$

$D \rightarrow aS$ $D \rightarrow c$

已求出能推出 ϵ 的非终结符集为{A,B,S}

	First集(0)	First集(1)	First集(2)	First集(3)
S	ϵ	ϵ b	ϵ b a	ϵ b a
A	ϵ	ϵ b	ϵ b	ϵ b
B	ϵ	ϵ a	ϵ a	ϵ a
C		b	b a c	b a c
D		a c	a c	a c
a	a	a	a	a
b	b	b	b	b

4. 确定的自顶向下语法分析[predictive]

• 开始符号集——FIRST集

- 利用求出每个文法符号的FIRST集求符号串的FIRST集, 设 $\alpha = X_1X_2 \dots X_n$:
 - ✓ 当 X_1 不能 $\Rightarrow^* \epsilon$, 则 $\text{FIRST}(\alpha) = \text{FIRST}(X_1)$
 - ✓ 若对任何 $j (1 \leq j < n)$ 都有 $\epsilon \in \text{FIRST}(X_j)$, 则 $\text{FIRST}(\alpha) = (\text{FIRST}(X_1) - \{\epsilon\}) \cup \dots \cup (\text{FIRST}(X_j) - \{\epsilon\}) \cup \text{FIRST}(X_{j+1})$
 - ✓ 若对所有 $i (1 \leq i \leq n)$, 都有 $\epsilon \in \text{FIRST}(X_i)$, 则 $\text{FIRST}(\alpha) = \text{FIRST}(X_1) \cup \dots \cup \text{FIRST}(X_n) \cup \{\epsilon\}$

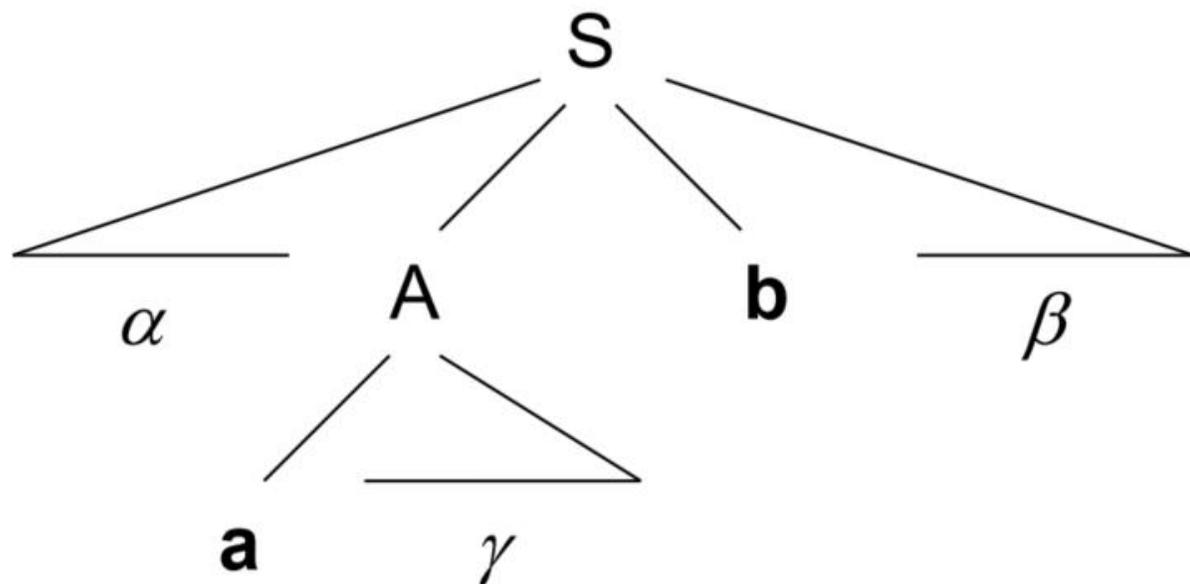
4. 确定的自顶向下语法分析[predictive]

• 开始符号集——**FIRST集**例: $G[S]$ $S \rightarrow AB|bC$ $A \rightarrow b|\epsilon$ $B \rightarrow aD|\epsilon$ $C \rightarrow AD|b$ $D \rightarrow aS|c$ 已求出非终结符的**First**集合如下: $\text{First}(S) = \{a, b, \epsilon\}$ $\text{First}(A) = \{b, \epsilon\}$ $\text{First}(B) = \{a, \epsilon\}$ $\text{First}(C) = \{a, b, c\}$ $\text{First}(D) = \{a, c\}$ 产生式右部符号串
的开始符集合为: $S \rightarrow AB$ $\text{FIRST}(AB) = \text{FIRST}(A) \cup \text{FIRST}(B) \cup \{\epsilon\} = \{a, b, \epsilon\}$ $S \rightarrow bC$ $\text{FIRST}(bC) = \{b\}$ $A \rightarrow \epsilon$ $\text{FIRST}(\epsilon) = \{\epsilon\}$ $A \rightarrow b$ $\text{FIRST}(b) = \{b\}$ $C \rightarrow AD$ $\text{FIRST}(AD) = (\text{FIRST}(A) - \{\epsilon\}) \cup \text{FIRST}(D) = \{b, a, c\}$ $C \rightarrow b$ $\text{FIRST}(b) = \{b\}$ $D \rightarrow aS$ $\text{FIRST}(aS) = \{a\}$ $D \rightarrow c$ $\text{FIRST}(c) = \{c\}$

4. 确定的自顶向下语法分析[predictive]

• **FIRST集**和**FOLLOW集**的直观理解

$\mathbf{a} \in \text{FIRST}(A)$ and $\mathbf{b} \in \text{FOLLOW}(A)$

• 后跟符号集——**FOLLOW集**

– 设 $G=(V_T, V_N, P, S)$ 是上下文无关文法,

$A \in V_N$, $\text{FOLLOW}(A)=\{b | S \Rightarrow^* \dots Ab \dots, b \in V_T\}$,

– 若有 $S \Rightarrow^* \dots A$, 则规定 $\$ \in \text{FOLLOW}(A)$

(注: ‘\$’ 做为输入串的结束符)

– 直观上说, 非终结符 A 的**后跟符号集**是由句型中紧跟 A 后的那些终结符 (包括 $\$$) 组成。

4. 确定的自顶向下语法分析[predictive]

• 后跟符号集——**FOLLOW**集

例：文法G[S]:

$$S \rightarrow aA|d$$

$$A \rightarrow bAS|\varepsilon$$

➤ 由 $S \Rightarrow^* aA$ 得 $\$ \in \text{FOLLOW}(A)$

由 $S \Rightarrow^* abAS \Rightarrow^* abAaA$ 得 $a \in \text{FOLLOW}(A)$

... $\Rightarrow^* abAd$ 得 $d \in \text{FOLLOW}(A)$

$\text{FOLLOW}(A) = \{\$, a, d\}$

➤ 由 $S \Rightarrow^* S$ 得 $\$ \in \text{FOLLOW}(S)$

由 $\underline{S} \Rightarrow a\underline{A} \Rightarrow ab\underline{AS} \Rightarrow abbASS \Rightarrow abbA\underline{S}aA$

... $\Rightarrow abbA\underline{S}d$

$\text{FOLLOW}(S) = \{\$, a, d\}$

4. 确定的自顶向下语法分析[predictive]

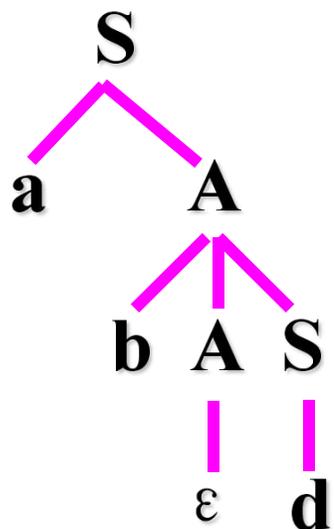
• 后跟符号集——**FOLLOW**集

例：文法G[S]:

$$S \rightarrow aA|d$$

$$A \rightarrow bAS|\varepsilon$$

若输入串W=abd，自顶向下的推导过程为：



$$\text{FIRST}(bAS) = \{b\}$$

$$\text{FOLLOW}(A) = \{\#, a, d\}$$

$$\underline{S} \Rightarrow \underline{a} \underline{A} \Rightarrow \underline{ab} \underline{A} \underline{S} \Rightarrow \underline{ab} \underline{S} \Rightarrow \underline{abd}$$

可进行确定的自顶向下分析

对于非终结符A的两个产生式
 $A \rightarrow bAS$ 和 $A \rightarrow \varepsilon$:

当输入符号 $\in \text{FIRST}(bAS) = \{b\}$ 时，
 选 $A \rightarrow bAS$ 推导；

当输入符号 $\in \text{FOLLOW}(A) = \{\#, a, d\}$
 时，选 $A \rightarrow \varepsilon$ 推导。

由于 $\text{FIRST}(bAS) \cap \text{FOLLOW}(A) = \emptyset$ ，
 所以可进行**确定的**自顶向下分析。

4. 确定的自顶向下语法分析[predictive]

• 后跟符号集——**FOLLOW集**– 计算**每个非终结符A的FOLLOW(A)集**

✓ (1) 对所有 $A \in V_N$, 令 $FOLLOW(A) = \{ \}$; 对开始符号 S , 令 $FOLLOW(S) = \{ \$ \}$

因为 $S \Rightarrow^* S$, 显然 $\$ \in FOLLOW(S)$

✓ (2) 对每条产生式 $A \rightarrow xBy$, 考察产生式右部的每一非终结符 B , $x, y \in V^*$:

• 如果 y 不能推出 ϵ : $FOLLOW(B) = FOLLOW(B) \cup First(y)$

• 否则, 若 $y \Rightarrow^* \epsilon$: $FOLLOW(B) = FOLLOW(B) \cup (First(y) - \{ \epsilon \}) \cup FOLLOW(A)$

若 $a \in FOLLOW(A)$, 则表明 $S \Rightarrow^* \dots Aa \dots$,

由于 $A \rightarrow xBy$, 且 $y \Rightarrow^* \epsilon$, 则有: $S \Rightarrow^* \dots Aa \dots \Rightarrow^* \dots xBya \Rightarrow^* \dots xBa \dots$,

即 $S \Rightarrow^* \dots xBa \dots$, 所以 $a \in FOLLOW(B)$

注意: ϵ 不存在于任何FOLLOW集中

✓ (3) 重复(2), 直至对所有 $A \in V_N$, $FOLLOW(A)$ **收敛为止**。

4. 确定的自顶向下语法分析[predictive]

• 后跟符号集——**FOLLOW集**

例G[S]:

[1]S→AB[2]S→bC

[3]A→b

[4]A→ε

[5]B→aD

[6]B→ε

[7]C→AD

[8]C→b

[9]D→aS

[10]D→c

已求出非终结符的First集合如下:

First(S)={a,b, ε} First(A)={b, ε}

First(B)={a, ε} First(C)={a,b,c}

First(D)={a,c}

已求出能推出ε的非终结符集为**{A,B,S}**

	Follow集(0)	Follow集(1)	Follow集(2)
S	\$	\$	\$
A		a \$ c	a \$ c
B		\$	\$
C		\$	\$
D		\$	\$

随堂练习 (5)

- 对文法 $G[E]$: $E \rightarrow TE'$, $E' \rightarrow +TE' | \varepsilon$, $T \rightarrow FT'$, $T' \rightarrow *FT' | \varepsilon$, $F \rightarrow (E) | n$, 求每个非终结符的FIRST集和FOLLOW集

(1) 能产生 ε 的非终结符: E' , T'

(2) $FIRST(E) = \{ (, n \}$

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

$FIRST(T) = \{ (, n \}$

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

$FIRST(F) = \{ (, n \}$

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

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

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

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

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

4. 确定的自顶向下语法分析[predictive]

• LL(1)文法

– 含义

- ✓ 第一个L表示**从左到右**扫描输入串
- ✓ 第二个L表示分析过程用**最左推导**
- ✓ 1表明只需**向前看一个符号**便可以决定选哪个产生式进行推导，类似地LL(k)文法需要向前看K个符号才可以确定选用哪个产生式。

– 定义

- ✓ 一个上下文无关文法是LL(1)文法的充分必要条件是，若存在产生式 $A \rightarrow \alpha | \beta$ ，则：
 - $\text{FIRST}(\alpha) \cap \text{FIRST}(\beta) = \Phi$
 - $\epsilon \in \text{FIRST}(\beta) \Rightarrow \text{FIRST}(\alpha) \cap \text{FOLLOW}(A) = \Phi$

两个条件同时满足!

LL(1)文法就能进行确定的自顶向下分析