第2章 文法和语言

学习目的：

掌握文法和语言的相关概念，为以后的词法分析、语法分析、语义分析奠定基础。

第2章 文法和语言

2.1 文法的直观概念

语言：是由句子组成的集合，是一组记号所构成的集合。

- **汉语** —— 所有符合汉语语法的句子的全体
- **英语** —— 所有符合英语语法的句子的全体
- **程序设计语言** —— 所有该语言的程序的全体

语言研究

研究语言 ：

- **每个句子构成的规律**
- **每个句子的含义**
- **每个句子和使用者的关系**

研究程序设计语言:

- **每个程序构成的规律**
- **每个程序的含义**
- **每个程序和使用者的关系**

→ 语言研究的三个方面：

- 语法 Syntax：表示构成语言句子的各个记号之间的组合规律。
- 语义 Semantics：表示按照各种表示方法所表示的各个记号的特定含义。
- 语用 Pragmatics：表示在各个记号所出现的行为中，它们的来源、使用和影响。

4

形式语言与文法

- **形式语言**：只从**语法**这一侧面来看语言，这种意义下的语言称作“形式语言”。

  “**形式**”是指：语言的所有规则**只以什么符号串能出现的方式**来陈述。

  A = B + 3 * C　✓

  A = B + * C　✗

- **形式语言理论**：是对符号串集合的表示法、结构及其特性的研究。

- **文法**：描述词法、语法规则的工具。用一组**规则**严格定义句子的结构，即对含有“无穷句子”的语言进行“有穷的表示”。
  - 〈赋值语句〉 ::= 〈id〉=〈表达式〉
  - 〈表达式〉 ::= 〈项〉 + 〈项〉
  - 〈表达式〉 ::= 〈项〉 − 〈项〉
  - ……

5

以自然语言为例，用 EBNF 描述语言的规则

文法(EBNF)

〈句子〉::= 〈主语〉〈谓语〉
〈主语〉::= 〈代词〉|〈名词〉
〈代词〉::= 你 | 我 | 他
〈名词〉::= 王明 | 大学生 | 工人 | 英语
〈谓语〉::= 〈动词〉〈直接宾语〉
〈动词〉::= 是 | 学习
〈直接宾语〉::= 〈代词〉|〈名词〉

判断下列句子是否是该语言的句子？（用规则去推导句子）

1. *我是大学生*
2. *我大学生是*
3. *他学习英语*
4. *英语学习他*

<句子> => <主语><谓语>
=> <代词><谓语>
=> 我<谓语>
=> 我<动词><直接宾语>
=> 我是<直接宾语>
=> 我是<名词>
=> 我是大学生

第2章 文法和语言

2.1　文法的直观概念

2.2　符号和符号串

2.3　文法和语言的形式定义

2.4　文法的类型

2.5　上下文无关文法及其语法树

2.6　句型的分析

2.7　有关文法实际应用的一些说明

7

2.2 符号和符号串

字母表 Σ：元素的非空有穷集合。（又称为符号集）

符号：字母表中的元素。

例如：

1) **汉语的字母表：包括汉字、数字及标点符号等。**
2) **英文的字母表：{a, b, … z , A, B, … , Z}**
3) **二进制的字母表：{0, 1}**
4) **标识符的字母表： {a … z , A … Z , 0 … 9 , _ }**

2.2 符号和符号串

符号串：由字母表Σ中的符号组成的任何有穷序列。

- 空符号串 ε（没有符号的符号串）是Σ上的符号串。
- 符号串不仅表示符号组成，还表示符号的顺序。

例如： $\Sigma=\{0,1\}$

ε, 0, 1, 00, 01, 10, 11, …, 1011, … 都是Σ上的符号串

$01 \neq 10$

2. 2 符号和符号串

(1) 符号串的长度：符号串x中符号的个数，用|x|表示

例如：|aabc|=4

(2) 空符号串：ε 则 |ε|=0

(3) 头、尾、固有头、固有尾

z=xy

- x是z的头；y是z的尾
- 若y非空，x是z的固有头；若x非空，y是z的固有尾。

例如：符号串 abc

头：ε, a, ab, abc 尾：ε, c, bc, abc

固有头：ε, a, ab 固有尾：ε, c, bc

(4) 符号串的连接：x, y的连接即xy（把y的符号写在x符号后面）

例如：x=01 y=abc 则 xy=01abc yx=abc01

εx=xε=x

2.2 符号和符号串

(5) 符号串的方幂：对符号串x，把它自身连接n次得到符号串z，$z=xxx\ldots x$，记作：$z=x^n$ ， $x^0=\varepsilon$ ， $x^1=x$ ， $x^2=xx$ ， ……

例如：$x=01$ ，则$x^0=\varepsilon$ $x^1=01$ $x^2=0101$ $x^3=010101$

(6) 符号串集合：集合A中的一切元素都是某字母表上的符号串，则称A为该字母表上的符号串集合。

例如：$\Sigma=\{0, 1\}$ $A=\{0, 1, 00, 01, 10\cdots, 10001, \cdots\cdots\}$ 是

$B=\{10, 11, 101\}$ 是

$C=\{1a, 11011, b11\}$ 不是

(7) 符号串集合的乘积：$AB=\{xy \mid x\in A且y\in B\}$

例如：$A=\{01, 10\}$ $B=\{ab, cd\}$

$AB=\{01ab, 10ab, 01cd, 10cd\}$ 注意：“ab01”不在 AB 中

$\{\varepsilon\}A = A\{\varepsilon\} = A$

2.2 符号和符号串

(8) 集合的闭包：指定字母表V之后，可用V^*表示V上所有有穷长度的串的集合。 V^+ 为V的正闭包。

$$V^* = V^0 \cup V^1 \cup \ldots \cup V^n \ldots$$

$$V^+ = V^1 \cup \ldots \cup V^n \ldots$$

则：

$$V^* = V^0 \cup V^+$$

$$V^+ = V V^* = V^* V$$

$$V^+ = V^* - \{\varepsilon\}$$

例如：设 V={0，1} 则

$V^*=\{\varepsilon, 0, 1, 00, 01, 10, 11, 000, 001, \cdots\}$

$V^+=\{0, 1, 00, 01, 10, 11, 000, 001, \cdots\}$

第2章 文法和语言

2.3 文法和语言的形式定义

规则（重写规则、产生式、生成式）

是形如 $\alpha \rightarrow \beta$ 或 $\alpha ::= \beta$ 的（α，β）有序对。

左部　右部

其中 $\alpha \in V^+$，　$\beta \in V^*$

举例：

〈程序〉 → 〈分程序〉 .

〈条件语句〉 → IF〈条件〉THEN〈语句〉

14

文法的定义

- 文法G 定义为四元组（V_N，V_T，P，S）
  - V_N ： 非终结符集
  - V_T ： 终结符集
  - P ： 产生式集合（规则集合）
  - S ： 开始符号（识别符号）

其中，

- V_N、V_T 和 P 是非空有穷集
- $S \in V_N$ ，并且 S 至少在一条规则中作为左部出现
- $V_N \cap V_T = \phi$
- $V = V_N \cup V_T$ ， V 称为文法G的字母表

15

例1 文法G =（V_N, V_T, P, S）

V_N = { S }

V_T ={ 0, 1 }

P={ S→0S1, S→01 }

S为开始符号

例2 文法G =（V_N, V_T, P, S）

V_N = {标识符, 字母, 数字}

V_T = {a, b, c, …x, y, z, 0, 1, …, 9}

P = {<标识符>→<字母>

<标识符>→<标识符><字母>

<标识符>→<标识符><数字>

<字母>→a, …, <字母>→z

<数字>→0, …, <数字>→9 }

S = <标识符>

例1 文法G=（V_N，V_T，P，S）

V_N = { S }

V_T ={ 0, 1 }

P={ S→0S1, S→01 }

S为开始符号

例1的简写形式：

G：S→0S1

S→01

或

G[S]：S→0S1

S→01

例2 文法G=（V_N，V_T，P，S）

V_N = {标识符，字母，数字}

V_T = {a, b, c, …x, y, z, 0, 1, …, 9}

P = {<标识符>→<字母>

<标识符>→<标识符><字母>

<标识符>→<标识符><数字>

<字母>→a, …, <字母>→z

<数字>→0, …, <数字>→9 }

S = <标识符>

文法的简写形式：

- 只写出产生式
- G写成G[S]，S是开始符号
  或 第一条产生式左部是开始符号
- 非终结符用尖括号括起或 大写字母
  终结符不用尖括号括起或 小写字母

推导的定义

- **直接推导 “⇒”**

  $\alpha \rightarrow \beta$ 是文法G的产生式， $\gamma, \delta \in V^*$，

  若有v, w满足：$v=\gamma\alpha\delta$ ， $w=\gamma\beta\delta$，

  则说：v（应用规则 $\alpha \rightarrow \beta$）直接产生w

  或说：w是v的直接推导

  或说：w直接归约到v

  记作：$v \Rightarrow w$

推导的定义

- **直接推导 “$\Rightarrow$”**

$\alpha\rightarrow\beta$ 是文法G的产生式， γ，$\delta\in V^*$，

若有v, w满足：$v=\gamma\alpha\delta$ ， $w=\gamma\beta\delta$，

则说：v（应用规则 $\alpha\rightarrow\beta$ ）直接产生w

或说：w是v的直接推导

或说：w直接归约到v

记作：$v\Rightarrow w$

例3 G: $S\rightarrow 0S1$

$S\rightarrow 01$

直接推导：

$S\Rightarrow 0S1$

$0S1\Rightarrow 0011$

$0S1\Rightarrow 00S11$

$\overset{+}{\Rightarrow}$ 和 $\overset{*}{\Rightarrow}$

若存在$v=w_0 \Rightarrow w_1 \Rightarrow \ldots \Rightarrow w_n=w$ ，(n>0)

则称 v 推导出 w（推导长度为n），

或称 v 产生 w

或称 w 归约到 v

记作 $v \overset{+}{\Rightarrow} w$

若有$v \overset{+}{\Rightarrow} w$，或$v=w$，则记为$v \overset{*}{\Rightarrow} w$

例4 G:S→0S1

S→01

$0S1 \Rightarrow 00S11 \Rightarrow 000S111 \Rightarrow 00001111$

即 $0S1 \overset{+}{\Rightarrow} 00001111$

$0S1 \overset{*}{\Rightarrow} 00001111$

句型、句子

- 句型

  设G[S]是一文法，如果符号串x是从开始符号推导出来的，即$S \stackrel{*}{\Rightarrow} x$，则称x是文法G[S]的句型。

句型、句子

- 句型

  设G[S]是一文法，如果符号串x是从开始符号推导出来的，即$S \overset{*}{\Rightarrow} x$，则称x是文法G[S]的句型。

- 句子

  x仅由V_T组成（即$S \overset{*}{\Rightarrow} x$，且$x \in V_T^*$），则称x是G[S]的句子。

例5 G[S]： S→0S1， S→01

由于存在$S \Rightarrow 0S1 \Rightarrow 00S11 \Rightarrow 000111$

句型：S 0S1 00S11 000111 00001111

句子：01 0011 000111 00001111

语言

由文法G生成的语言记为L(G)

它是文法G的一切句子的集合：

$L(G)=\{x \mid S \overset{*}{\Rightarrow} x$，其中S为文法的开始符号，且$x \in V_T^*\}$

重点掌握：①根据文法，写出对应的语言

②构造出一种语言的文法

例6　G：S→0S1，　S→01

则 $L(G)=\{0^n 1^n \mid n \geq 1\}$

练习

例7 文法G[S]：

$S \rightarrow AB$

$A \rightarrow aAb$

$A \rightarrow ab$

$B \rightarrow Bc$

$B \rightarrow \varepsilon$

求对应的语言。

练习

例7 文法G[S]： $S \rightarrow AB$ 求对应的语言。

$A \rightarrow aAb$

$A \rightarrow ab$

$B \rightarrow Bc$

$B \rightarrow \varepsilon$

答案：$L[G]=\{ a^m b^m c^n \mid m>0,\ n >=0 \}$

练习

例8 $L[G]=\{ a^m b^n \mid m, n > 0 \}$求对应文法

例9 $L[G]=\{ a^m b^n \mid m, n \geqslant 0 \}$求对应文法

练习

例8 $L[G]=\{ a^m b^n \mid m, n > 0 \}$求对应文法

答案：$S \to AB$
$A \to aA \mid a$
$B \to bB \mid b$

或

$S \to aS$
$S \to aB$
$B \to bB \mid b$

例9 $L[G]=\{ a^m b^n \mid m, n \geq 0 \}$求对应文法

练习

独自增长型！

例8 $L[G]=\{ a^m b^n \mid m, n > 0 \}$求对应文法

答案：$S \rightarrow AB$

$A \rightarrow aA \mid a$

$B \rightarrow bB \mid b$

或

$S \rightarrow aS$

$S \rightarrow aB$

$B \rightarrow bB \mid b$

例9 $L[G]=\{ a^m b^n \mid m, n \geq 0 \}$求对应文法

答案：$S \rightarrow AB$

$A \rightarrow aA \mid \varepsilon$

$B \rightarrow bB \mid \varepsilon$

或

$S \rightarrow aS$

$S \rightarrow B$

$B \rightarrow bB \mid \varepsilon$

练习

例10 $L[G]=\{ a^n b^n \mid n > 0 \}$求对应文法

例11 $L[G]=\{ a^n b^n \mid n \geq 0 \}$求对应文法

练习： $L[G]=\{ a^n b^{2n} \mid n > 0 \}$ 求对应文法。

练习

例10　$L[G]=\{ a^n b^n \mid n > 0 \}$求对应文法

答案：$S \to aSb$　或　$A \to aB \mid ab$

$S \to ab$　　　$B \to Ab$

例11　$L[G]=\{ a^n b^n \mid n \geq 0 \}$求对应文法

练习：$L[G]=\{ a^n b^{2n} \mid n > 0 \}$ 求对应文法。

练习

卷心菜型！

例10　$L[G]=\{ a^n b^n \mid n > 0 \}$求对应文法

答案：$S \to aSb$　或　$A \to aB \mid ab$
$S \to ab$　　$B \to Ab$

例11　$L[G]=\{ a^n b^n \mid n \geq 0 \}$求对应文法

答案：$S \to aSb$　或　$A \to aB \mid \varepsilon$
$S \to \varepsilon$　　$B \to Ab$

练习：$L[G]=\{ a^n b^{2n} \mid n > 0 \}$ 求对应文法。

练习

卷心菜型！

例10 $L[G]=\{ a^n b^n \mid n > 0 \}$求对应文法

答案：$S \rightarrow aSb$ 或 $A \rightarrow aB \mid ab$

$S \rightarrow ab$ $B \rightarrow Ab$

例11 $L[G]=\{ a^n b^n \mid n \geqslant 0 \}$求对应文法

答案：$S \rightarrow aSb$ 或 $A \rightarrow aB \mid \varepsilon$

$S \rightarrow \varepsilon$ $B \rightarrow Ab$

练习：$L[G]=\{ a^n b^{2n} \mid n > 0 \}$ 求对应文法。

$S \rightarrow aSbb$ 或 $A \rightarrow aB \mid \varepsilon$

$S \rightarrow \varepsilon$ $B \rightarrow Abb$

练习

卷心菜型！

例12　$L[G]=\{\ a^n\ b\ b^n\ |\ n>0\ \}$求对应文法（独心卷心菜）

例13　$L[G]=\{\ a^m\ b^n\ |\ n>=m>=1\ \}$求对应文法(混合卷心菜)

练习

卷心菜型！

例12　$L[G]=\{ a^n\ b\ b^n \mid n>0 \}$求对应文法（独心卷心菜）

答案：$S \rightarrow aAb$

$A \rightarrow aAb$

$A \rightarrow b$

例13　$L[G]=\{ a^m b^n \mid n>=m>=1 \}$求对应文法(混合卷心菜)

练习

卷心菜型！

例12 L[G]={ a^n b b^n | n >0 }求对应文法（独心卷心菜）

答案：S→aAb
A→aAb
A→b

例13 L[G]={ a^m b^n | n>=m>=1 }求对应文法(混合卷心菜)

答案：S→aAb
A→aAb
A→Ab
A→ ε

文法的等价

- 若 $L(G_1) = L(G_2)$，则称文法G_1和G_2是等价的。
  即：两个不同的文法 能够产生 相同的语言。

例如 文法$G_1[A]$：$A \rightarrow 0R$ 与 $G_2[S]$：$S \rightarrow 0S1$ 等价

$A \rightarrow 01$　　　　$S \rightarrow 01$

$R \rightarrow A1$

36

第2章 文法和语言

2.4 文法的类型

- 1956年，Chomsky建立形式语言的描述
- 通过对产生式施加不同的限制，Chomsky将文法分为四种类型：

0型文法(PSG)：对任一产生式 $\alpha \rightarrow \beta$

$\alpha \in (V_N \cup V_T)^*$，且至少含一个$V_N$

$\beta \in (V_N \cup V_T)^*$

即：对产生式没有任何限制

例如：A0→1A0 ， A1→B

2.4 文法的类型

1型文法(CSG)：

对任一产生式 $\alpha \rightarrow \beta$

$|\beta| >= |\alpha|$，　仅仅 $S \rightarrow \varepsilon$ 除外

产生式的形式描述：$\alpha_1 A \alpha_2 \rightarrow \alpha_1 \beta \alpha_2$

其中，α_1、α_2、$\beta \in (V_N \cup V_T)^*$，$\beta \neq \varepsilon$，$A \in V_N$

即：A只有出现在 $\alpha_1 \alpha_2$ 的上下文中，才允许用 β 替换。

产生的语言称“上下文有关语言”。

例如：0A0→011000　　1A1→101011

2.4 文法的类型

2型文法(CFG)：

对任一产生式 $\alpha \rightarrow \beta$

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

通常产生式的形式描述：$A \rightarrow \beta$ $(A \in V_N)$

即：β取代A时，与A所处的上下文无关。

产生的语言称“上下文无关语言”。

例如：G[S]：S→01

S→0S1

2.4 文法的类型

3型文法(RG)：

每个产生式均为 $A \to aB$ 或 $A \to a$ —— 右线性

$A \to Ba$ 或 $A \to a$ —— 左线性

其中，A、$B \in V_N$，$a \in V_T^*$

产生的语言称“正规语言”、“正则语言”。

例如：G[S]：

$S \to 0A \mid 0$

$A \to 1B \mid B$

$B \to 1 \mid 0$

文法举例

- 例14：2型文法（上下文无关文法）

文法G[S]：S→ 0 A

S→ 1 B

S→ 0

A→ 0 A

A→ 1 B

B→ 1 B

B→ 1

B→ 0

第2章 文法和语言

2.1 文法的直观概念

2.2 符号和符号串

2.3 文法和语言的形式定义

2.4 文法的类型

2.5 上下文无关文法及其语法树

2.6 句型的分析

2.7 有关文法实际应用的一些说明

43

2.5 上下文无关文法及其语法树

- 上下文无关文法 有足够的能力描述现今程序设计语言的语法结构。
  - 算术表达式
  - 语句
    - 赋值语句
    - 条件语句
    - 读语句
    - ……

算术表达式上下文无关文法表示

文法G[E]:

E → E+E

E → E*E

E → (E)

E → i

条件语句上下文无关文法表示

〈条件语句〉→if 〈条件〉 then 〈语句〉

〈条件语句〉→if 〈条件〉 then 〈语句〉 else 〈语句〉

45

上下文无关文法的语法树

- 用于描述上下文无关文法的句型推导的直观方法

例：G[S]：

$S \rightarrow aAS$

$A \rightarrow SbA$

$A \rightarrow SS$

$S \rightarrow a$

$A \rightarrow ba$

46

上下文无关文法的语法树

- 用于描述上下文无关文法的句型推导的直观方法

例：G[S]：

S→aAS

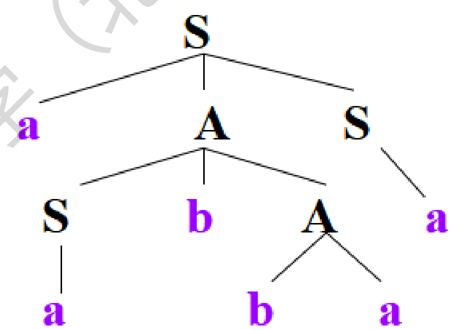
A→SbA

A→SS

S→a

A→ba

句型aabbaa的语法树（推导树）



上下文无关文法的语法树

- 用于描述上下文无关文法的句型推导的直观方法

例：G[S]：

S→aAS

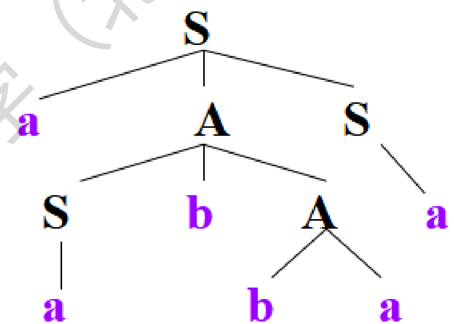
A→SbA

A→SS

S→a

A→ba

句型aabbaa的语法树（推导树）



叶子结点：树中没有子孙的结点。

从左到右读出推导树的叶子标记，所得的句型为推导树的结果。也把该推导树称为该句型的语法树。

推导过程中施用产生式的顺序

例：G[S]：

S→aAS

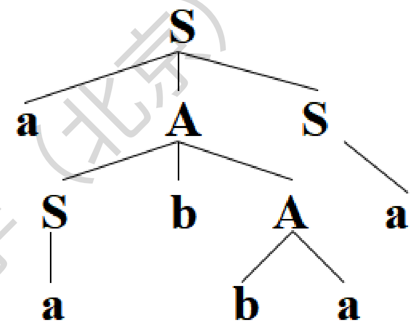
A→SbA

A→SS

S→a

A→ba

句型aabbaa的语法树（推导树）



49

推导过程中施用产生式的顺序

例：G[S]：

S→aAS

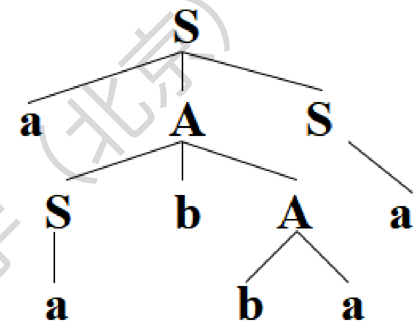
A→SbA

A→SS

S→a

A→ba

句型aabbaa的语法树（推导树）



S⇒aAS⇒aAa⇒aSbAa⇒aSbbaa⇒aabbaa 最右推导（规范推导）

S⇒aAS⇒aSbAS⇒aabAS⇒aabbaS⇒aabbaa 最左推导

S⇒aAS⇒aSbAS⇒aSbAa⇒aabAa⇒aabbaa

50

问题：一个句型是否对应唯一的一棵语法树？

例：G[E]：

E → i

E → E+E

E → E*E

E → (E)

问题：一个句型是否对应唯一的一棵语法树？

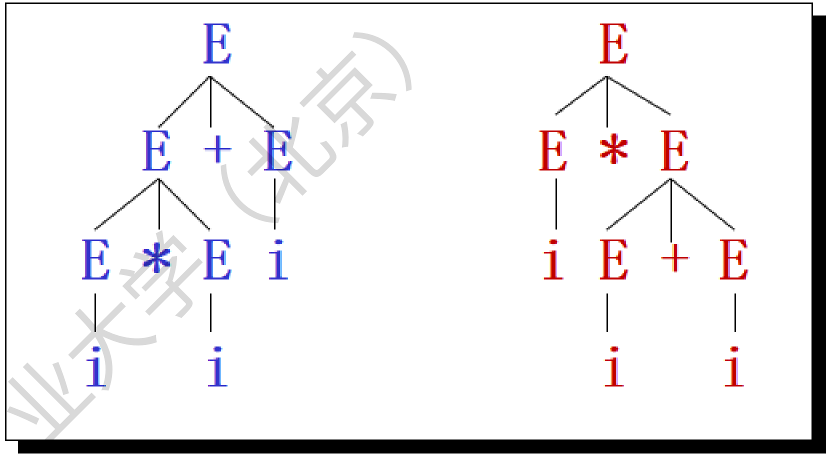
例：G[E]:

E → i

E → E+E

E → E*E

E → (E)



句型 i*i+i 的两个不同的最左推导：

推导1：E ⇒ E+E ⇒ E*E+E ⇒ i*E+E ⇒ i*i+E ⇒ i*i+i

推导2：E ⇒ E*E ⇒ i*E ⇒ i*E+E ⇒ i*i+E ⇒i*i+i

问题：一个句型是否对应唯一的一棵语法树？

例：G[E]:

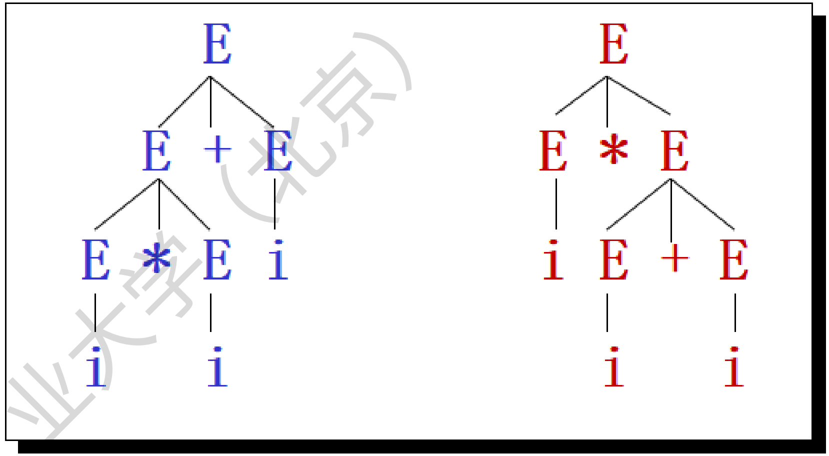
二义文法

E → i

E → E+E

E → E*E

E → (E)



句型 i*i+i 的两个不同的最左推导：

推导1：E ⇒ E+E ⇒ E*E+E ⇒ i*E+E ⇒ i*i+E ⇒ i*i+i

推导2：E ⇒ E*E ⇒ i*E ⇒ i*E+E ⇒ i*i+E ⇒i*i+i

53

二义文法

- 若一个文法存在某个句子，该句子对应两棵不同的语法树，则称这个文法是二义的。
  - 程序设计语言要求：文法不能是二义的！

- 产生某上下文无关语言的每一个文法都是二义的，则称此语言是先天二义的。

54

二义文法

- 如何证明文法是二义的？

找出一个句子，对应的语法树（或最左推导过程）不唯一。

P. 34　9. 文法 $S \rightarrow S(S)S \mid \varepsilon$

(1) 生成的语言是什么

(2) 该文法是二义的吗？请说明理由。

二义文法

- 如何证明文法是二义的？

找出一个句子，对应的语法树（或最左推导过程）不唯一。

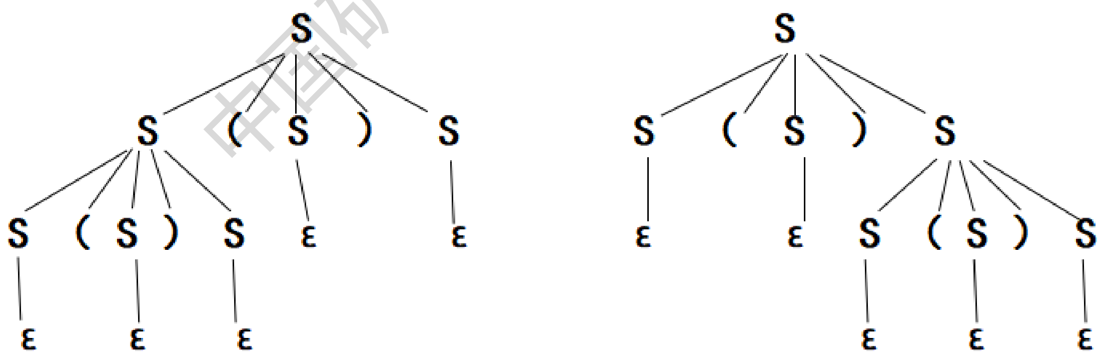
P. 34　9. 文法 $S \rightarrow S(S)S \mid \varepsilon$

(1) 生成的语言是什么

(2) 该文法是二义的吗？请说明理由。

答：(1) 可以任意嵌套、任意连接的( )串

(2) 对于句子( )( )可以构造两棵语法树，该文法二义。



56

改造二义文法

- 例：将 二义文法 改造为 无二义文法

G[E]：E → i
E → E+E
E → E*E
E → (E)

G[E]：E → T|E+T
T → F|T*F
F → （E）|i

或规定优先顺序和结合律

第2章 文法和语言

2.1 文法的直观概念

2.2 符号和符号串

2.3 文法和语言的形式定义

2.4 文法的类型

2.5 上下文无关文法及其语法树

2.6 句型的分析

2.7 有关文法实际应用的一些说明

2.6 句型的分析

- 句型分析
  - 就是识别一个符号串是否为某文法的句型，是某个推导的构造过程。
- 分析程序
  - 在语言的编译实现中，把完成句型分析的程序称为分析程序或识别程序。
  - 分析算法又称识别算法。
  - 从左到右的分析算法：

    总是从左到右地识别输入符号串。首先识别符号串中的最左符号，进而依次识别右边的一个符号。

分析算法分类

- 自上而下分析法
  - 从文法的开始符号出发，反复使用各种产生式，寻找与输入符号匹配的推导。
- 自下而上分析法
  - 从输入符号串开始，逐步进行归约，直至归约到文法的开始符号。

两种方法反映了两种不同的语法树的构造过程。

自上而下的分析

例　　文法G：S → cAd

A → ab

A → a

识别输入串 cabd 是否该文法的句子。

61

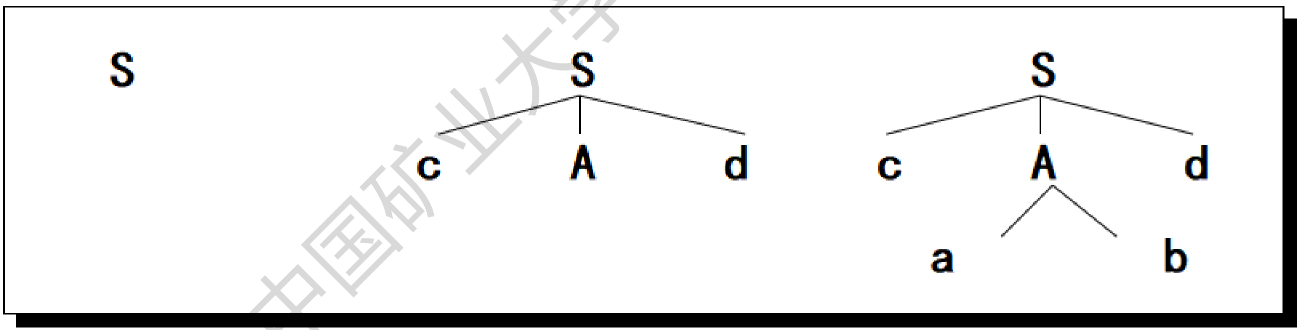
自上而下的分析

例　　文法G：$S \rightarrow cAd$

$A \rightarrow ab$

$A \rightarrow a$

识别输入串 cabd 是否该文法的句子。



推导过程：$S \Rightarrow cAd \Rightarrow cabd$

自下而上的分析

例　　文法G：$S \rightarrow cAd$

$A \rightarrow ab$

$A \rightarrow a$

识别输入串 cabd 是否该文法的句子

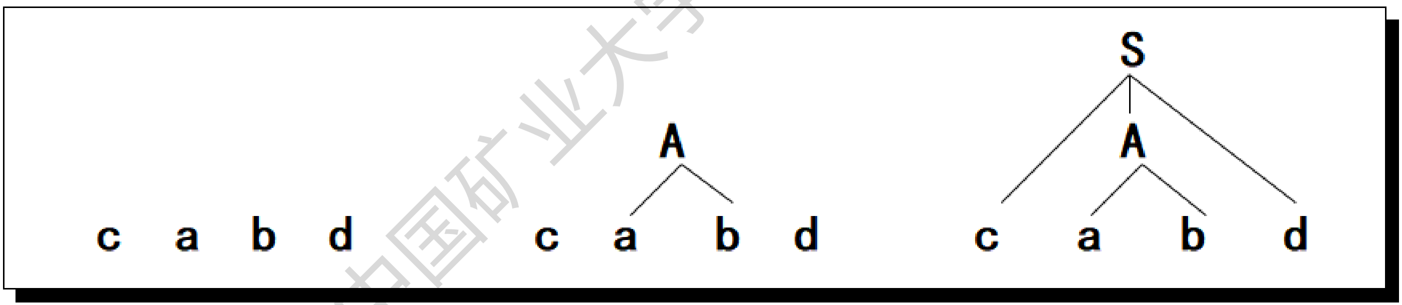
自下而上的分析

例　　文法G：$S \rightarrow cAd$

$A \rightarrow ab$

$A \rightarrow a$

识别输入串 <u>cabd</u> 是否该文法的句子



归约的过程：$S \Rightarrow cAd \Rightarrow cabd$

64

句型分析的有关问题

1）选择使用哪个产生式进行自顶向下推导？

2）如何确定“可归约串”进行自下而上归约？

在自下而上的分析方法中，在分析程序工作的每一步，都是从当前串中选择一个子串，将它归约到某个非终结符号，该子串称为“可归约串”。如何确定“可归约串”？

引出“句柄”的概念。

短语、直接短语、句柄

- 短语
  - 文法G[S]，　$\alpha\beta\delta$是G的一个句型，如果：

    $S \overset{*}{\Rightarrow} \alpha A \delta$　且　$A \overset{+}{\Rightarrow} \beta$

    则称β是句型$\alpha\beta\delta$相对于非终结符A的短语。
- 直接短语：若有$A \Rightarrow \beta$则称β是句型$\alpha\beta\delta$相对于规则$A \rightarrow \beta$的直接短语（或简单短语）。
- 句柄：一个句型的最左直接短语称为该句型的句柄。

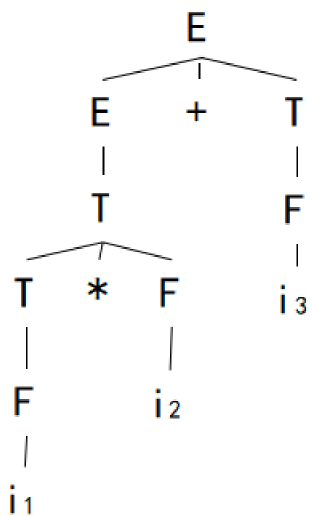
短语、直接短语、句柄

例. 求句型i*i+i的 短语、直接短语、句柄

G[E]：E→E+T|T

T→T*F|F

F→(E)|i



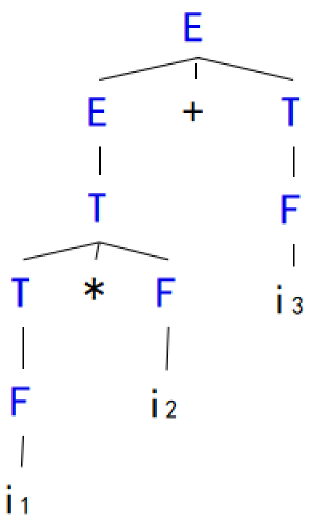
短语、直接短语、句柄

例. 求句型i*i+i的 短语、直接短语、句柄

G[E]: E→E+T|T

T→T*F|F

F→(E)|i



短语

i*i+i

i*i

i1

i2

i3

68

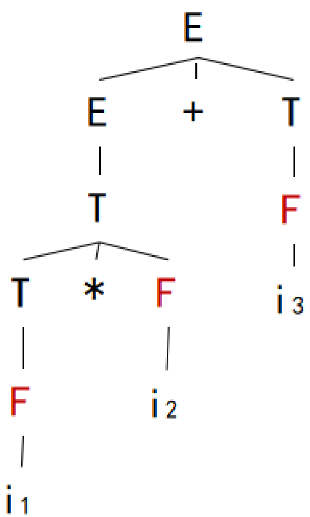
短语、直接短语、句柄

例. 求句型i*i+i的 短语、直接短语、句柄

G[E]：E→E+T|T

T→T*F|F

F→(E)|i



短语	直接短语	句柄（最左直接短语）
i*i+i		
i*i		
i_1	i_1	i_1
i_2	i_2	
i_3	i_3	

第2章 文法和语言

2.7 有关文法实际应用的一些说明

实际应用中，需要对文法提出一些限制条件。

但这些限制并不真正限制该文法描述的语言。

限制1：文法中不得含有“有害规则”

有害规则：引起文法二义性的规则。如A→A

限制2：文法中不得含有“多余规则”

多余规则：文法中任何句子的推导都用不到的规则。

(1) 规则中有不可到达的非终结符

(2) 规则中有不可终止的非终结符

为保证无多余规则，必须满足的两个条件：

1）A必须在某句型中出现。

2）必须能从A推出终结符号串t来。

2.7 有关文法实际应用的一些说明

限制3：不含“左递归”规则。 （如：A→A...）

限制4：有些文献规定：不含“空规则”（A→ε）；但有些文献允许“空规则”的出现。

2.7 有关文法实际应用的一些说明

限制1：文法中不得含有“有害规则”

限制2：文法中不得含有“多余规则”

限制3：不含“左递归”规则

限制4：有些文献规定：不含“空规则”

例 化简文法

G[S]：S→Be

B→Ce

B→Af

A→Ae

A→e

C→Cf

D→f

化简后

G[S]：S→Be

B→Af

A→Ae

A→e

73

第2章 文法和语言（小结）

1. 符号和符号串相关概念
   (1) 字母表
   (2) 符号串
   长度、头、尾、固有头、固有尾、连接、幂
   集合、集合的乘积、集合的闭包、正闭包

2. 文法和语言
   (1) 规则、文法
   (2) 推导、句子、句型
   (3) 语言：一切句子的集合
   (4) 文法→语言　　　语言→文法
   (5) 文法的等价性

第2章 文法和语言（小结）

3. 文法的类型

0型——短语文法
1型——上下文有关文法
2型——上下文无关文法
3型——正规文法、正则文法（左线性、右线性）

4. 上下文无关文法及其语法树

（1）语法树的构造
（2）推导过程：最左推导、最右推导
（3）文法二义性：一个句子对应两个或以上语法树

5. 句型的分析

（1）自上而下
（2）自下而上
（3）短语、直接短语、句柄

75

练习

1. “符号就是字符”，这种说法正确吗？

A. 正确　　　　B. 不正确

练习

1. “符号就是字符”，这种说法正确吗？

A. 正确　　　B. 不正确

B

2. 文法G[S]： $S \rightarrow A0$

$S \rightarrow B1$

$A \rightarrow S1$

$A \rightarrow 1$

$B \rightarrow S0$

$B \rightarrow 0$

该文法是Chomsky___(1)___型文法，该文法所描述的所有只含四个符号的句子是___(2)___。

2. 文法G[S]：S→A0

S→B1

A→S1

A→1

B→S0

B→0

该文法是Chomsky__(1)__型文法，该文法所描述的所有只含四个符号的句子是___(2)___。

(1) 3

(2) 1010，0110，1001，0101

79

3. 文法G[S]：

S→A

A→B|if A then A else A

B→C|B+C|+C

C→D|C*D|*D

D→x|(A)|-D

哪些是终结符，哪些是非终结符？

3. 文法G[S]:

S→A

A→B|if A then A else A

B→C|B+C|+C

C→D|C*D|*D

D→x|(A)|-D

哪些是终结符，哪些是非终结符？

答:

终结符: if、then、else、+、*、x、(、)、-

非终结符: S 、A 、B 、C 、D

4. 文法G[S]：

$S \rightarrow a \mid aB$

$B \rightarrow aS$

该文法所描述的语言是__________?

A. $L(G)=\{a^i \mid i \geq 0\}$ B. $L(G)=\{a^{2i} \mid i \geq 0\}$

C. $L(G)=\{a^{2i+1} \mid i \geq 0\}$ D. $L(G)=\{a^{2i+1} \mid i \geq 1\}$

4. 文法G[S]：

$S \rightarrow a \mid aB$

$B \rightarrow aS$

该文法所描述的语言是 ___C___ ？

A. $L(G)=\{a^i \mid i \geq 0\}$　　B. $L(G)=\{a^{2i} \mid i \geq 0\}$

C. $L(G)=\{a^{2i+1} \mid i \geq 0\}$　D. $L(G)=\{a^{2i+1} \mid i \geq 1\}$

83

5. 一个语言的文法是________?

A. 唯一的　B. 不唯一　C. 个数有限的

5. 一个语言的文法是_____B_____?

A. 唯一的　B. 不唯一　C. 个数有限的

6. 已知语言$L=\{a^{2n}bb^n \mid n>=1\}$，

写出对应文法。

86

6. 已知语言L=$\{a^{2n}bb^n | n>=1\}$，

写出对应文法。

答案：$S \rightarrow aaAb$

$A \rightarrow aaAb | b$

7. 文法G[E]： $E \rightarrow E+T \mid T$

$T \rightarrow T*F \mid F$

$F \rightarrow (E) \mid a$

写出该文法句型 E+F*(E+T)的短语、直接短语和句柄。

7. 文法G[E]： $E \rightarrow E+T \mid T$

$T \rightarrow T*F \mid F$

$F \rightarrow (E) \mid a$

写出该文法句型 E+F*(E+T) 的短语、直接短语和句柄。

短语：	E+F*(E+T)	F*(E+T)	F	(E+T)	E+T
直接短语：			F		E+T
句柄：			F		

8. ______正则文法能产生下面的语言：

$L=\{a^nb^n | n>=1\}$

A. 存在一个　　B. 不存在　　C. 无法判别

8. ______正则文法能产生下面的语言：

$L=\{a^nb^n \mid n>=1\}$

A. 存在一个　　B. 不存在　　C. 无法判别

答案：B

91

9. 文法G[S]：$S \rightarrow a \mid (T) \mid \varepsilon$

$T \rightarrow T,S \mid S$

请给出句子(a,(a,a))的最左、最右推导。

9. 文法G[S]：$S \to a \mid (T) \mid \varepsilon$

$T \to T, S \mid S$

请给出句子(a, (a, a))的最左、最右推导。

最左推导	最右推导
$S \Rightarrow (T)$	$S \Rightarrow (T)$
$\Rightarrow (T, S)$	$\Rightarrow (T, S)$
$\Rightarrow (S, S)$	$\Rightarrow (T, (T))$
$\Rightarrow (a, S)$	$\Rightarrow (T, (T, S))$
$\Rightarrow (a, (T))$	$\Rightarrow (T, (T, a))$
$\Rightarrow (a, (T, S))$	$\Rightarrow (T, (S, a))$
$\Rightarrow (a, (S, S))$	$\Rightarrow (T, (a, a))$
$\Rightarrow (a, (a, S))$	$\Rightarrow (S, (a, a))$
$\Rightarrow (a, (a, a))$	$\Rightarrow (a, (a, a))$

93

课后习题

8. 上下文无关文法

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

(1) 通过此文法如何生成串 aa+a*，并为该串构造推导树。
(2) 该文法生成的语言是什么?

课后习题

8. 上下文无关文法

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

(1) 通过此文法如何生成串 aa+a*，并为该串构造推导树。
(2) 该文法生成的语言是什么？

答:

（1） $S \Rightarrow SS*$
$\Rightarrow Sa*$
$\Rightarrow SS+a*$
$\Rightarrow Sa+a*$
$\Rightarrow aa+a*$

(2) 语言是
a，或者以*或+结尾的a、*、+组成的串。

课后作业

12（1-6）

13（1-4）

18

课后习题

12. 构造产生如下语言的上下文无关文法

(1) $\{a^n b^n \mid n \geq 0\}$

(2) $\{a^m b^n \mid m \geq n \geq 0\}$

(3) $\{\underline{\mu a \omega} b \mid \mu,\omega \in \{a,b\}^* \wedge |\mu|=|\omega|\}$

(4) $\{a^n b^m \mid n \geq 2m \geq 0\}$

(6) $\{\omega\omega^R \mid \omega \in \{a,b\}^*\}$ R表示反向串

97

课后习题

12. 构造产生如下语言的上下文无关文法

(1) $\{a^n b^n \mid n \geq 0\}$

G: $S \rightarrow aSb$

$S \rightarrow \varepsilon$

(2) $\{a^m b^n \mid m \geq n \geq 0\}$

G: $S \rightarrow aSb$

$S \rightarrow aS$

$S \rightarrow \varepsilon$

(3) $\{\underline{\mu a \omega} b \mid \mu, \omega \in \{a,b\}^* \wedge |\mu| = |\omega|\}$

G: $S \rightarrow Tb$

$T \rightarrow aTa$

$T \rightarrow bTb$

$T \rightarrow aTb$

$T \rightarrow bTa$

$T \rightarrow a$

98

课后习题

12. 构造产生如下语言的上下文无关文法

(4) $\{a^n b^m \mid n \geq 2m \geq 0\}$

G: $S \rightarrow aaSb$

$S \rightarrow aS$

$S \rightarrow \varepsilon$

(6) $\{\omega\omega^R \mid \omega \in \{a,b\}^*\}$　R表示反向串

G: $S \rightarrow aSa$

$S \rightarrow bSb$

$S \rightarrow \varepsilon$

课后习题

13. 构造产生如下语言的上下文无关文法

(1) $\{a^n b^m c^{2m} \mid n,m \ge 0\}$

(2) $\{\omega c \omega^R \mid \omega \in \{a,b\}^*\}$

(3) $\{a^m b^n c^k \mid m=n$ 或 $n=k, m,n,k \ge 0\}$

(4) $\{a^m b^n c^k \mid m=k$ 或 $n=k, m,n,k \ge 0\}$

课后习题

13. 构造产生如下语言的上下文无关文法

(1) $\{a^n b^m c^{2m} \mid n,m \geq 0\}$

G: $S \rightarrow AB$

$A \rightarrow aA$

$A \rightarrow \varepsilon$

$B \rightarrow bBcc$

$B \rightarrow \varepsilon$

(2) $\{\omega c \omega^R \mid \omega \in \{a,b\}^*\}$

G: $S \rightarrow aSa$

$S \rightarrow bSb$

$S \rightarrow c$

课后习题

13. 构造产生如下语言的上下文无关文法

(3) $\{a^m b^n c^k \mid m=n \text{ 或 } n=k, m,n,k \geq 0\}$

G: $S \to AC \mid BD$

$A \to aAb$

$A \to \varepsilon$

$C \to cC$

$C \to \varepsilon$

$B \to aB$

$B \to \varepsilon$

$D \to bDc$

$D \to \varepsilon$

解题思路：

$\underline{a^m b^m}\,\underline{c^k}$ $\underline{a^m}\,\underline{b^n c^n}$

A C B D

102

课后习题

13. 构造产生如下语言的上下文无关文法

(4) $\{a^m b^n c^k \mid m=k$ 或 $n=k, m,n,k \geq 0\}$

G: $S \rightarrow A|B$

$A \rightarrow aAc$

$A \rightarrow K$

$K \rightarrow bK$

$K \rightarrow \varepsilon$

$B \rightarrow aB$

$B \rightarrow C$

$C \rightarrow bCc$

$C \rightarrow \varepsilon$

解题思路:

$\underbrace{a^m b^n c^m}_{A}$ $\underbrace{a^m b^n c^n}_{B}$

103

课后习题

18. 构造生成下述语言的三型文法

(1) $\{a^n \mid n \geq 0\}$

(2) $\{a^n b^m \mid n,m \geq 1\}$

(3) $\{a^n b^m c^k \mid n,m,k \geq 0\}$

课后习题

18. 构造生成下述语言的三型文法

(1) $\{a^n \mid n \geq 0\}$

G: $S \rightarrow aS$

$S \rightarrow \varepsilon$

(2) $\{a^n b^m \mid n,m \geq 1\}$

G: $A \rightarrow aA$

$A \rightarrow aB$

$B \rightarrow bB$

$B \rightarrow b$

(3) $\{a^n b^m c^k \mid n,m,k \geq 0\}$

G: $A \rightarrow aA$

$A \rightarrow B$

$B \rightarrow bB$

$B \rightarrow C$

$C \rightarrow cC$

$C \rightarrow \varepsilon$

105