讲稿（教学内容、步骤）

第 2 章 文法和语言（1）

1. 语言：是由句子组成的集合，是一组记号所构成的集合。
   (1) 汉语 —— 所有符合汉语语法的句子的全体
   (2) 英语 —— 所有符合英语语法的句子的全体
   (3) 程序设计语言 —— 所有该语言的程序的全体

2. 语言研究的三个方面：
   (1) 每个句子构成的规律——语法 Syntax：表示构成语言句子的各个记号之间的组合规律。
   (2) 每个句子的含义——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。
   (3) 每个句子和使用者的关系——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。

   则程序设计语言的研究内容包括：
   (1) 语法：每个程序构成的规律
   (2) 语义：每个程序的含义
   (3) 语用：每个程序和使用者的关系

3. 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”。
   (1) “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。
   【举例】 A = B + 3 * C √
   A = B + * C ×
   (2) 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。

4. 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”。
   【举例】
   <赋值语句> ::= <id>=<表达式>
   <表达式> ::= <项> + <项>
   <表达式> ::= <项> - <项>
   ……
   【举例】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

5. 字母表：元素的非空有穷集合。（又称为符号集）

6. 符号：字母表中的元素。

【举例】

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 , _ }

7. 符号串：由字母表 中的符号组成的任何有穷序列。

(1) 空符号串ε(没有符号的符号串)是Σ上的符号串。

(2) 符号串不仅表示符号组成，还表示符号的顺序。

【举例】若 Σ={0,1}

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

注意：$01 \neq 10$

(3) 符号串的长度：符号串 x 中符号的个数，用|x|表示

【举例】 |aabc|=4

|ε|=0

(4) 头、尾、固有头、固有尾

【举例】若符号串 abc

头：ε, a, ab, abc　　尾：ε, c, bc, abc

固有头：ε, a, ab　　固有尾：ε, c, bc

(5) 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）

【举例】若符号串 x=01　符号串 y=abc

则 xy=01abc　　yx=abc01

注意：εx=xε=x

(6) 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...x，记作 $z=x^n$。

则 $x^0=\varepsilon$ ， $x^1=x$ ， $x^2=xx$ ， ......

【举例】若 x=01

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

(7) 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。

【举例】若字母表 Σ={0, 1}

A ={0,1,00,01,10...，10001,......}　是字母表Σ上的符号串集合

B ={10,11,101}　是

C ={1a,11011,b11}　不是

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

【举例】若 $A=\{01,10\}$　$B=\{ab, cd\}$

则 $AB=\{01ab,10ab,01cd,10cd\}$

注意：① “ab01”不在 AB 中

② $\{\varepsilon\}A = A\{\varepsilon\} = A$

(9) 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。

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

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

则：$V^* = V^0 \cup V^+$

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

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

例如：设 $V=\{0，1\}$ 则

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

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

8. 规则（重写规则、产生式、生成式）

是形如$\alpha\to\beta$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>

9. 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

【举例】

例1 文法G =（V_N，V_T，P，S）

V_N = { S }

V_T ={ 0，1 }

P={ $S\to 0S1$，　$S\to 01$ }

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>

文法的简写：

- 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

例1的简写形式：

G： S→0S1
　　S→01

或

G[S]： S→0S1
　　　S→01

10. 推导

(1) 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 => w

【举例】

例3 G：S→0S1
S→01

直接推导：

$$S \Rightarrow 0S1$$
$$0S1 \Rightarrow 0011$$
$$0S1 \Rightarrow 00S11$$

(2) $\overset{+}{\Rightarrow}$ **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** $v \overset{*}{\Rightarrow} w$

【举例】

例4 G: $S \to 0S1$

$S \to 01$

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

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

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

11. 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。

12. 句子

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

【举例】

例5 G[S]：$S \to 0S1$， $S \to 01$

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

句型：S 0S1 00S11 000111 00001111

句子：01 0011 000111 00001111

【讨论】句型和句子的异同?

13. 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

【举例】

例 6 G: $S \to 0S1$， $S \to 01$

则 $L(G)=\{0^n 1^n \mid n \geqslant 1\}$

例 7 文法 G[S]: $S \to AB$ 求对应的语言。

$A \to aAb$

$A \to ab$

$B \to Bc$

$B \to \varepsilon$

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

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

答案： $S \to AB$

$A \to aA \mid a$

B→bB | b

或：S→aS

S→aB

B→bB | b

例 9　L[G]={ $a^m b^n$ | m , n ≥ 0} 求对应文法　（独自增长型）

答案：S→AB

A→aA | ε

B→bB | ε

或：S→aS

S→B

B→bB | ε

【讨论】例 8 和例 9 的特点？

例 10　L[G]={ $a^n b^n$ | n > 0} 求对应文法　（卷心菜型）

答案：S→aSb

S→ab

或：A→aB | ab

B→Ab

例 11　L[G]={ $a^n b^n$ | n ≥ 0} 求对应文法　（卷心菜型）

答案：S→aSb

S→ ε

或：A→aB | ε

B→Ab

【讨论】例 11 和例 12 的特点？

【练习】 L[G]={ $a^n b^{2n}$ | n >0 }　求对应文法

答案：S→aSbb

S→ ε

或　A→aB | ε

B→Abb

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

答案：S→aAb

A→aAb

A→b

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

答案：S→aAb

A→aAb

A→Ab

A→ε

【讨论】例 12 和例 13 的特点？

14. 文法的等价

若 L(G1)= L(G2)，则称文法 G1 和 G2 是等价的。

即：两个不同的文法 能够产生 相同的语言。

【举例】

文法 $G_1[A]$： A→0R 与 $G_2[S]$： S→0S1 等价

A→01 S→01

R→A1

15. 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：<u>对产生式没有任何限制</u>

例如：A0→1A0 ， A1→B

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

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

产生的语言称“上下文有关语言”。

例如：0A0→011000 1A1→101011

2型文法(CFG)：

对任一产生式α→β

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

通常产生式的形式描述：A→β（$A \in V_N$）

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

3型文法(RG)：

每个产生式均为 A→aB 或 A→a —— 右线性

A→Ba 或 A→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→0A | 0

A→1B | B

B→1 | 0

【举例】

例 14：2 型文法（上下文无关文法）

G[S]: S→0A

S→1B

S→0

A→0A

A→1B

B→1B

B→1

B→0

【课后作业】12(1-6)　13(1-4)　18

12. 构造产生如下语言的上下文无关文法

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

G: S→aSb

S→ ε

(2) $\{a^mb^n \mid m \geqslant n \geqslant 0\}$

G: S→aSb

S→aS

S→ ε

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

G: S→Tb

T→aTa

T→bTb

T→aTb

T→bTa

T→a

另 **G:　S→Tb**

T→ATA

T→a

A→a | b

推导 **S=>Tb**

=>ATAb

=>AATAAb

=>AAATAAAb

(4) $\{a^nb^m \mid n \geqslant 2m \geqslant 0\}$

G: S→aaSb

S→aS

S→ ε

(5) $\{a^nb^m \mid n \geqslant 0,\ m \geqslant 0,$ 且 $3n \geqslant m \geqslant 2n\}$

G: S→aSbb

S→aSbbb

S→ ε

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

G: S→aSa

S→bSb

S→ ε

13. 构造产生如下语言的上下文无关文法

(1) $\{a^nb^mc^{2m} \mid n,m \geqslant 0\}$

G: S→AB

A→aA

A→ ε

B→bBcc

B→ ε

(2) { ω c ω ^{R} | ω ∈{a,b}* }

G: S→aSa

S→bSb

S→c

(3) {$a^m b^n c^k$ | m=n 或 n=k, m,n,k≥0}

G: S→AC | BD

A→aAb

A→ ε

C→cC

C→ ε

B→aB

B→ ε

D→bDc

D→ ε

(4) {$a^m b^n c^k$ | m=k 或 n=k, m, n, k≥0}

G: S→A|B

A→aAc

A→K

K→bK

K→ ε

B→aB

B→C

C→bCc

C→ ε

(5) {$a^n b^m c^k$ | n≤m+k , k≤n, 且 m,n,k≥1}

G: S→aAc

A→BD

B→aBb

B→ ε

D→bD

D→b

(6) {$a^n b^k c^m$ | n,k,m≥0 且 n≥m}

G: S→aSC

S→B

B→bB

B→ ε

C→c

C→ ε

(7) {$a^n b^n c^m d^m$ | n≥1, m≥1}∪{$a^n b^m c^m d^n$ | n≥1, m≥1}

G: S→AB
S→T
A→aAb
A→ab
B→cBd
B→cd
T→aTd
T→aDd
D→bDc
D→bc

(8) $\{\omega_1 c\omega_2 c \ldots c\omega_k cc\omega_j^R \mid k \geqslant 1 \wedge 1 \le j \le k$ 且对任何 $1 \le i \le k$，有 $\omega i \in \{a,b\}^+\}$　R 表示反向串

G: S→AB
A→CcA
A→ ε
C→aC
C→bC
C→a
C→b
B→aBa
B→bBb
B→cAc

18. 构造生成下述语言的三型文法

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

G: S→aS
S→ ε

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

G: A→aA
A→aB
B→bB
B→b

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

G: A→aA
A→B
B→bB
B→C
C→cC
C→ ε