

并非侠义地指字符 (character)

▶2.1基本概念

▶串:有穷符号 (symbol) 序列

单词 → 字符

句子 → 单词

- ▶2.1基本概念
 - ▶串:有穷符号 (symbol) 序列
 - ▶ 串 s 的 长 度: |s|
 - **▶空串:** ε
 - ▶串上的运算
 - ▶连接 xy
 - >幂 sⁿ

▶2.1基本概念

▶串:有穷符号 (symbol) 序列

>字母表:有穷符号集合

>字母表上的运算 运算结果是串的集合

> 乘积

$$\sum_{1}\sum_{2}$$

长度为2的串构成的集合

〉幂

$$\sum n$$

长度为n的串构成的集合

$$\triangleright \sum \theta = \{\varepsilon\} \neq \Phi$$

$$ightharpoonup$$
 正闭包 $\Sigma^+=\Sigma$ U Σ^2 U Σ^3 U ... 长度为正数的串构成的集合

 \triangleright 克林闭包 $\Sigma^* = \Sigma^0 \cup \Sigma^+$

任意长度的串 (包括空串) 构成的集合

- ▶2.1基本概念
 - ▶串:有穷符号 (symbol) 序列
 - >字母表:有穷符号集合
- ▶语言: {句子} 语言是字符及其组合规则的统一体
 - ▶句子: 词串
 - ▶词:字符串
- > 句子本身并不是由随机抽取的一些单词组成的序列 (串)
 - 〉例
 - > I'm I do, sorry that afraid Dave I'm can't.
 - > I'm sorry Dave, I'm afraid I can't do that.

- ▶2.1基本概念
- ▶2.2文法的定义

自然语言的例子——句子的构成规则

- ▶ <句子> → <名词短语> <动词短语>
- ▶ <名词短语> → <形容词> <名词短语>
- ▶ <名词短语> → <名词>
- ▶ <动词短语> → <动词> <名词短语>
- ▶ <形 容词> → little
- < 名词> → boy

▶ <名词> → apple

> <动词> → eat

未用尖括号括起来部分表示语言的基本符号

尖括号括起来部分称为语法成分

- ▶2.1基本概念
- ▶2.2文法的定义

$$G = (V_T, V_N, P, S)$$

- $ightarrow V_T$: 终结符 (语言的基本符号) 集合
- 字母表 \rightarrow 如果G表示句法 (用来描述句子构成规则的文法) ,则每个终结符是一个单词 (token)
 - ▶如果G表示词法 (用来描述单词构成规则的文法) ,则每个终结符是一个字符

- ▶2.1基本概念
- ▶2.2文法的定义

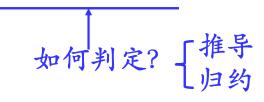
$$G = (V_T, V_N, P, S)$$

- $\triangleright V_T$: 终结符 (语言的基本符号token) 集合
- > 符号的约定 句

终结符
$$a,b,c$$
 终结符号串 u,v,\ldots,z 非终结符 A,B,C 文法符号 X,Y,Z 文法符号串 α,β,γ

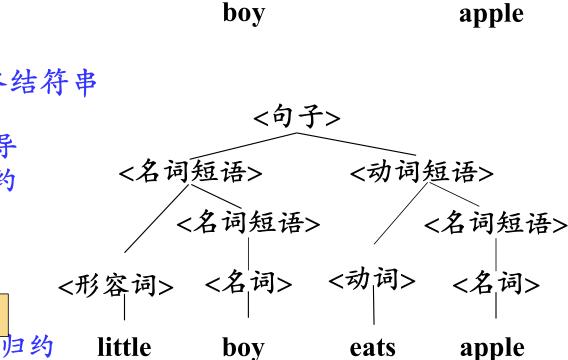
第2讲(语言及其文法)要点 推导 <向子>

- ▶2.1基本概念
- ▶2.2文法的定义
- ▶2.3语言的定义
 - ▶语言: {句子}
 - >句子: 满足文法规则的终结符串



例: little boy eats apple

I'm I do, sorry that afraid Dave I'm can't.



<形容词><名词短语> <动词> <名词短语>

<名词短语>

little

<动词短语>

<名词> eats <名词>

- ▶2.1基本概念
- ▶2.2文法的定义
- ▶2.3语言的定义
 - ▶语言: {句子}

$$L(G) = \{ w \mid S \Rightarrow^* w, w \in V_T^* \}$$

> 句子: 满足文法规则的终结符串

规则 终结符串

- ▶2.1基本概念
- ▶2.2文法的定义
- ▶2.3语言的定义
 - ▶语言: {句子}
 - ▶语言上的运算
 - \triangleright 并 $L_1 \cup L_2$
 - \rightarrow 连接 L_1L_2
 - ▶幂 Lⁿ

$$\triangleright L^{\theta} = \{\varepsilon\} \neq \Phi$$

- ightharpoonup 正闭包 $L^+=L\cup L^2\cup L^3\cup ...$
- ▶ 克林闭包 $L^* = L^0 \cup L^+$

- ▶2.1基本概念
- ▶2.2文法的定义
- ▶2.3语言的定义
- ▶2.4文法的分类
- ▶2.5CFG的分析树
 - ▶直接短语:高度为2的子树的边缘
 - \triangleright 直接短语一定是某产生式的右部 $E \rightarrow E + E$
 - ▶但产生式的右部不一定是给定句型的直接短语

