

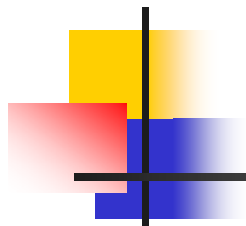


## 第四讲 汉语的句法结构分析（I）

---

詹卫东

<http://ccl.pku.edu.cn/course/nlp>



# 提纲

---

1. 句法结构分析概述
2. 形式文法
3. 句法结构分析算法
  - 自底向上和自顶向下分析法
  - Earley算法
  - Tomita算法
4. 小结



# 1 句法结构分析概述

句法：关于“词是如何组成句子的”（的）知识

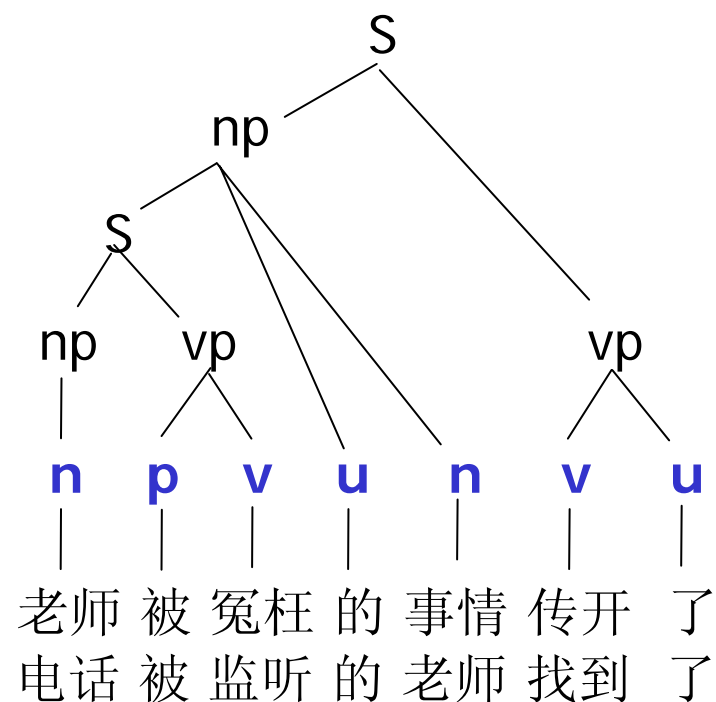
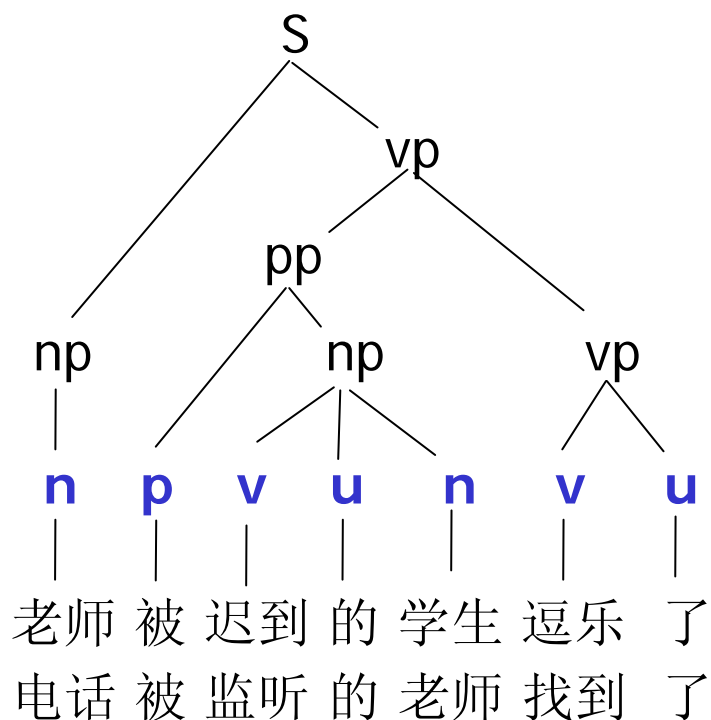
例：下面这些句子有何不同？

甲

乙

- |    |                    |   |                    |                      |
|----|--------------------|---|--------------------|----------------------|
| 1. | a 老师 被 迟到 的 学生 逗乐了 | ✓ | b 迟到 的 学生 把 老师 逗乐了 |                      |
|    |                    |   | ✗                  | c 老师 被 迟到 的 学生 被 逗乐了 |
| 2. | b 老师 被 冤枉 的 事情 传开了 | ✗ | b 冤枉 的 事情 把 老师 传开了 |                      |
|    |                    |   | ✓                  | c 老师 被 冤枉 的 事情 被 传开了 |
| 3. | c 电话 被 监听 的 老师 找到了 | ✓ | b 监听 的 老师 把 电话 找到了 |                      |
|    |                    |   | ✓                  | c 电话 被 监听 的 老师 被 找到了 |

# 句子内部结构的树图表示





# 自然语言的层次结构特性示例

---

1. 听说服装设计很吃香      ——      听说那套服装设计得很有品位
2. 听说孩子丢了                      ——      听说孩子丢了一只鞋
3. 听说北京队大败                      ——      听说北京队大败上海队

同一个线性字符串，根据所处上下文环境的不同而解释为不同的树结构！



# 如何进行句法结构分析

- 句法结构分析： 从“线性串”到“树结构”的映射。  
如何在—对多中求解最优的映射？
- 需要做两件事：
  1. 语言模型  
语法体系形式化描述的任务：
    - 语言成分有多少类（范畴）？
    - 成分间组合模式有多少种？
    - 成分组合的约束条件是什么？
  2. 搜索算法  
计算技术的任务：
    - 如何快速找到正确的结构树



## 2 形式文法 无穷字符序列的有穷表示法

---

一个形式文法 $G$ 由四个部分组成，可记作  $G = \{V_N, V_T, S, P\}$ ，其中：

$V_N$ ：称为文法 $G$ 的非终端符号字母表， $V_N$ 不出现在  
     $G$ 所表示的语言集合的句子中；

$V_T$ ：称为文法 $G$ 的终端符号字母表， $G$ 所表示的语  
    言的句子由  $V_T$  中的元素组成， $V_N \cap V_T = \phi$ ；

$S$ ：代表开始符号， $S \in V_N$ 。

$P$ ：代表一组变换式组成的集合， $P$  中的式子具有如下形式：

$$\alpha \rightarrow \beta$$



## 形式文法（续）

---

$$\alpha \rightarrow \beta$$

称为产生式规则（production rule）或 重写规则（rewriting rule）

产生式规则需要满足下面的条件：

- 1)  $\alpha$ 可以是  $V_N$  和  $V_T$  上的任意字符串，不能是空字符；
- 2)  $\beta$ 可以是  $V_N$  和  $V_T$  上的任意字符串，可以是空字符；
- 3)  $P$  中至少有一个产生式中的  $\alpha$  得由  $S$  来充当；





# 形式文法的Chomsky Hierarchy

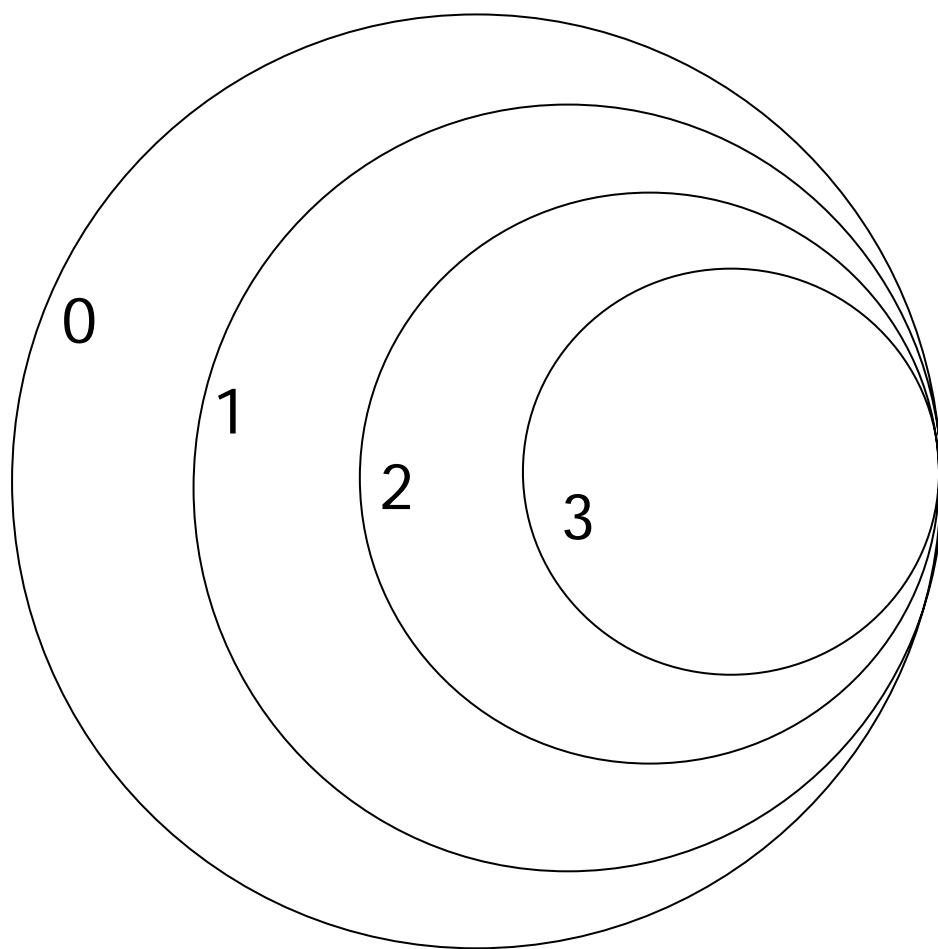
| 分级 | 名称  | 产生式规则的形式限制  |
|----|-----|---|
| 0  | PSG | $\alpha \rightarrow \beta$ with $\alpha \in (V_T \cup V_N)^+$ and $\beta \in (V_T \cup V_N)^*$  |
| 1  | CSG | $\alpha_1 A \alpha_2 \rightarrow \alpha_1 \beta \alpha_2$ with $A \in V_N$ and $\alpha_1, \alpha_2 \in (V_T \cup V_N)^*$ and $\beta \in (V_T \cup V_N)^+$ |
| 2  | CFG | $A \rightarrow \beta$ with $A \in V_N$ and $\beta \in (V_T \cup V_N)^*$   |
| 3  | RG  | $A \rightarrow \beta B$ or $A \rightarrow \beta$ with $A, B \in V_N$ and $\beta \in V_T^*$  |

Noam Chomsky, 1956, Three Models For The Description Of Language, *IRE Transactions on Information Theory*, 2 (1956), pp.113-124.

Noam Chomsky, 1959, On Certain Formal Properties Of Grammars, *Information and Control*, Vol. 2 (1959), pp.137-167.



# 形式文法的Chomsky Hierarchy



$G_0$ : 无限制重写文法 PSG

$G_1$ : 上下文相关文法 CSG

$G_2$ : 上下文无关文法 CFG

$G_3$ : 正则文法 RG

$L_0$ : 递归可枚举语言

$L_1$ : 上下文相关语言

$L_2$ : 上下文无关语言

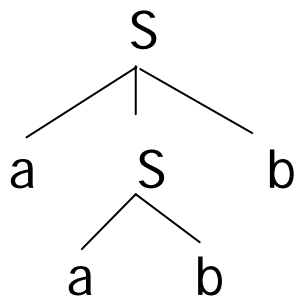
$L_3$ : 正则语言



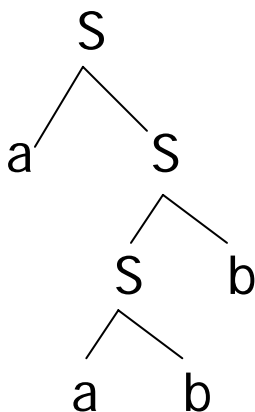
## 用CFG来描述语言

对于语言  $L = \{ab, aabb, aaabbb, \dots, a^n b^n, \dots\}$   $n$  是自然数。  
请写出  $L$  的上下文无关文法。

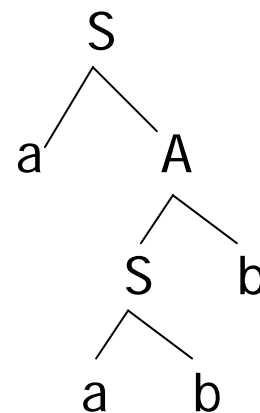
$G_1$  (1)  $S \rightarrow a S b$   
(2)  $S \rightarrow a b$



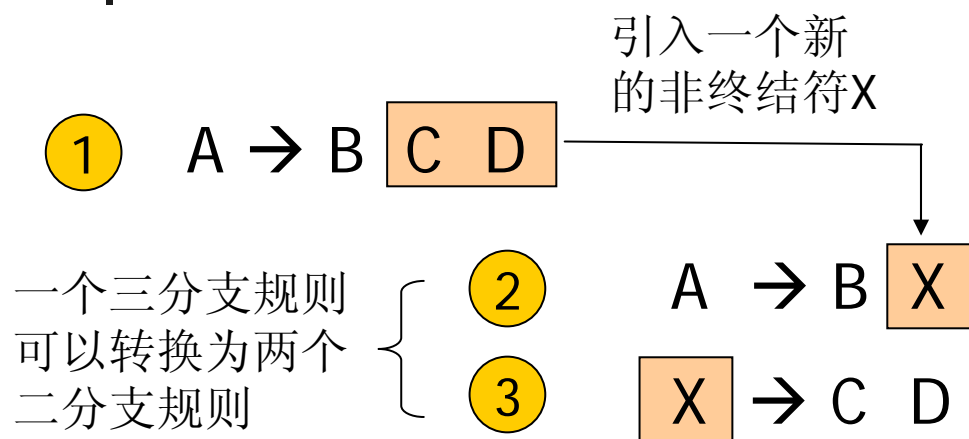
$G_2$  (1)  $S \rightarrow a S$   
(2)  $S \rightarrow S b$   
(3)  $S \rightarrow a b$



$G_3$  (1)  $S \rightarrow a A$   
(2)  $A \rightarrow S b$   
(3)  $S \rightarrow a b$

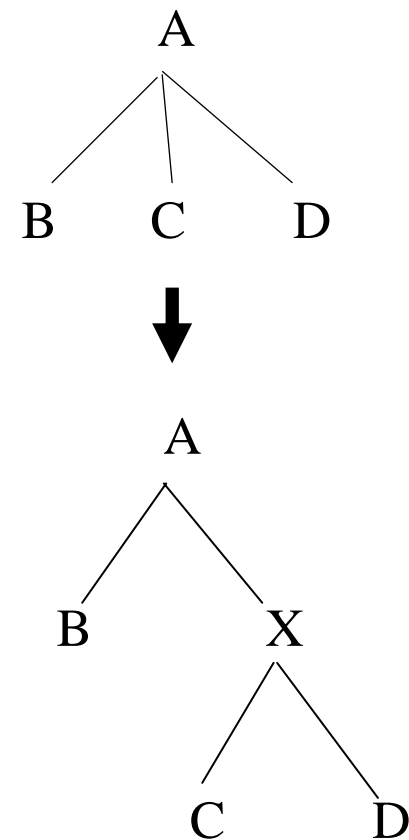


# 乔姆斯基范式 (Chomsky Normal Form, CNF)



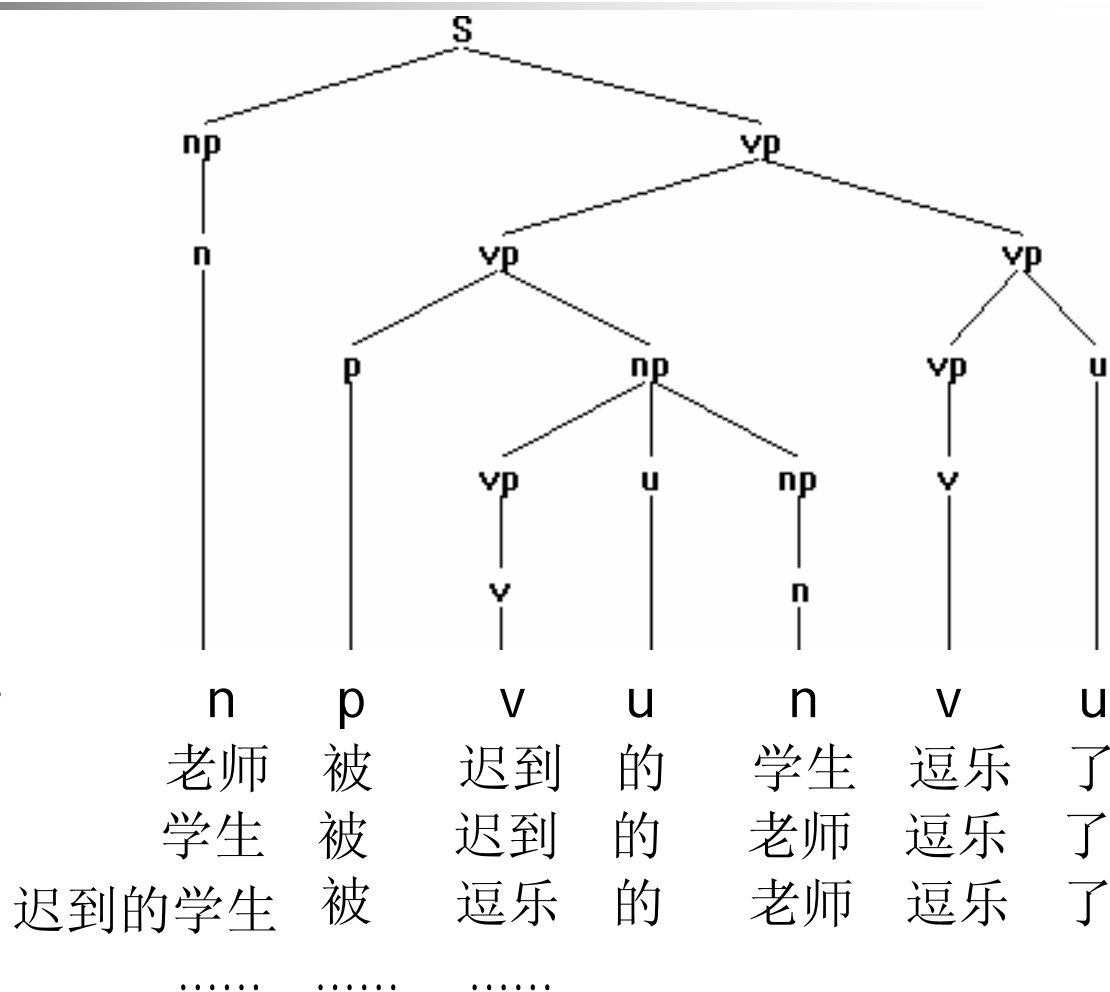
CNF 的规则形式

- $A \rightarrow B C$
- $A \rightarrow a$



# 用CFG描述自然语言

1.  $S \rightarrow np \ vp$
2.  $np \rightarrow vp \ u \ np$
3.  $vp \rightarrow pp \ vp$
4.  $vp \rightarrow vp \ u$
5.  $pp \rightarrow p \ np$
6.  $np \rightarrow n$
7.  $vp \rightarrow v$
8.  $n \rightarrow \text{老师} \mid \text{学生} \dots$
9.  $v \rightarrow \text{迟到} \mid \text{逗乐} \dots$
10.  $p \rightarrow \text{被} \dots$
11.  $u \rightarrow \text{的} \mid \text{了} \dots$





## 文法的三个作用

---

- 生成：产生语言L中所有的句子；
- 判定：一个字符串是否属于语言L；
- 分析：得到L中句子的结构树；



# 课后练习

---

- 1 写出可以产生汉语自然数表达式的CFG
- 2 用你写的CFG，画出下列数字的分析树：

一亿零三百万

三万六千五百八十一



## 3 句法结构分析算法

---

自底向上 (bottom-up) 基于规约的方法。

从待分析字符串开始，用待分析字符串去匹配CFG规则箭头的右部字符，匹配成功后替换为左部字符，直到S。

自顶向下 (top-down) 基于预测的方法。

从CFG规则中的S规则开始，将CFG规则箭头左部的符号展开，直到形成以终结符开始的序列，用该序列去匹配待分析字符串，直到完全匹配上。



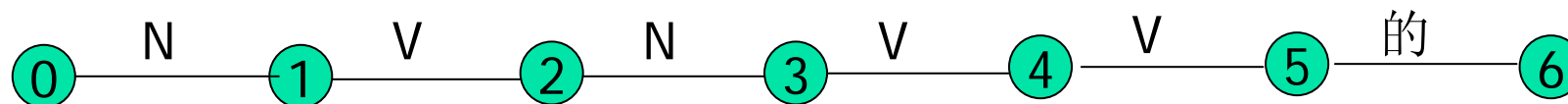


# Earley算法

---

Top-down（为主）与 bottom-up（为辅）相结合  
预测能力 + 数据驱动

张三是县长派来的 — N V N V V 的





## 基本概念：状态(state)

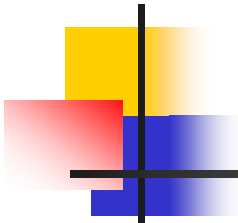
一个状态由3部分组成：

- (1) 上下文无关文法规则
- (2) 圆点  $\cdot$  （圆点左边的部分是已分析的，右边是待分析的）
- (3) 状态的起止位置：
  - 整数  $i$ ：状态起点（已分析子串的起点）
  - 整数  $j$ ：状态终点（已分析子串的终点）  $i \leq j$

比如：  $\langle S \rightarrow NP \cdot VP \ [0, 4] \rangle$

点规则

点在最右端，为 [完成状态]  
否则，为 [未完成状态]



## 基本操作/算子 (operator)

1. **预测 (Predictor) :** 如果圆点右方是一个非终结符, 那么以该非终结符为左部的规则都有匹配的希望, 也就是说分析器可以预测这些规则都可以建立相应的项目。
2. **扫描 (Scanner) :** 如果圆点右方是一个终结符, 就将圆点向右方扫描一个字符间隔, 把匹配完的字符“让”到左方。
3. **归约 (Completer) :** 如果圆点右方没有符号 (即圆点已经在状态的结束位置), 那么表示当前状态所做的预测已经实现, 因而可以将当前状态 ( $S_i$ ) 与已有的包含当前状态的状态 ( $S_j$ ) 进行归约 (合并), 从而扩大  $S_j$  覆盖的子串范围



## 算子的形式定义

**Predicator:** 对于状态  $Z \rightarrow \alpha \cdot X \beta \ [j, k]$  其中 $X$ 是非终结符  
对于语法中每条形如  $X \rightarrow \gamma$  的规则, 都可以形成一个新状态:  $X \rightarrow \cdot \gamma \ [k, k]$

**Scanner:** 对于状态  $Z \rightarrow \alpha \cdot X \beta \ [j, k]$  其中 $X$ 是终结符  
如果 $X$ 与输入字符串中第 $k$ 个字符匹配, 就形成一个新状态:  $Z \rightarrow \alpha X \cdot \beta \ [j, k+1]$

**Completer :** 对于一个已经“完成”的状态  $Z \rightarrow \gamma \cdot \ [j, k]$   
如果已有状态集合中有形如  $X \rightarrow \alpha \cdot Z \beta \ [i, j]$  这样的状态, 就形成一个新状态:  $X \rightarrow \alpha Z \cdot \beta \ [i, k]$

说明: 以上 $\alpha, \beta, \gamma$ 是终结符或非终结符串, 其中 $\alpha, \beta$ 均可为空字符  $i \leq j \leq k$



## Earley算法：算法描述

设输入字符串长度为 $n$ , 字符间隔可记做 $0, 1, 2, \dots, n$

- (1) 将文法规则中形如  $S \rightarrow \alpha$  的规则形成为状态：  
 $\langle S \rightarrow \cdot \alpha \ [0, 0] \rangle$  加入到状态集合中（种子状态/seed state）
- (2) 对当前分析句子的每个词，依次进行循环：  
对状态集中的每个状态，依次进行循环：
  - i) 如果当前状态是[未完成状态]，且点后不是终结符，则  
执行Predicator;
  - ii) 如果当前状态是[未完成状态]，且点后是终结符，则  
执行Scanner;
  - iii) 如果当前状态是[完成状态]，则  
执行Completer;
- (3) 如果最后得到形如  $\langle S \rightarrow \alpha \cdot \ [0, n] \rangle$  这样的状态，那么  
输入字符串被接受为合法的句子，否则分析失败

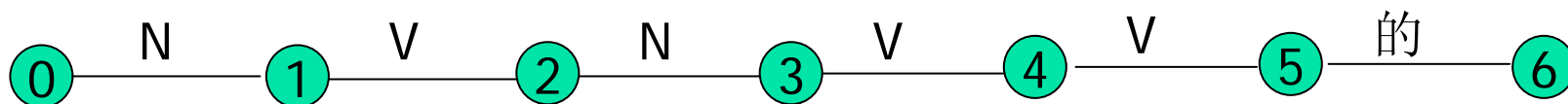
## Earley算法过程示例

张三是县长派来的  
老虎是瞎子打死的  
主意是董永想出来的

.....

N V N V V 的

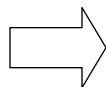
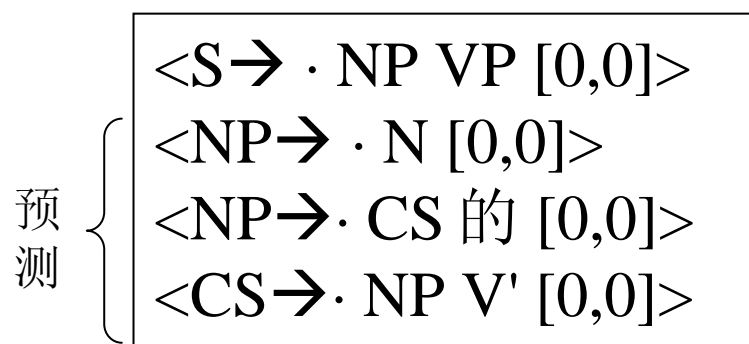
- (1)  $S \rightarrow NP VP$
- (2)  $NP \rightarrow N$
- (3)  $NP \rightarrow CS$  的
- (4)  $CS \rightarrow NP V'$
- (5)  $VP \rightarrow V NP$
- (6)  $V' \rightarrow V V$



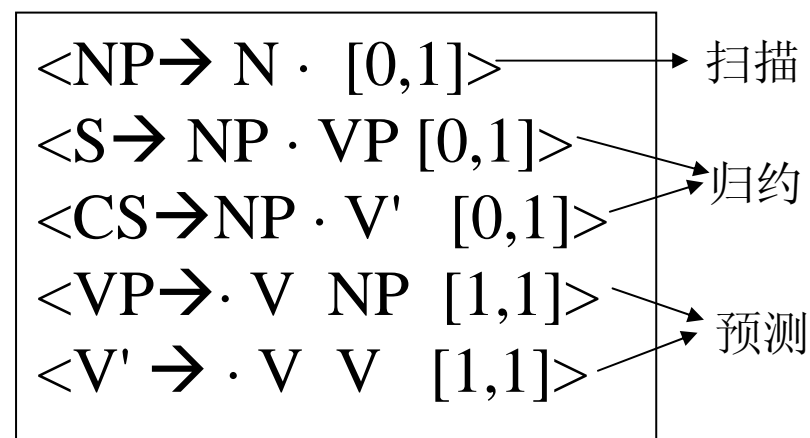
# Earley算法过程示例-1

- (1)  $S \rightarrow NP \ VP$
- (2)  $NP \rightarrow N$
- (3)  $NP \rightarrow CS \text{ 的}$
- (4)  $CS \rightarrow NP \ V'$
- (5)  $VP \rightarrow V \ NP$
- (6)  $V' \rightarrow V \ V$

当前间隔 ①



当前间隔 ②



预测 → 扫描 → 归约 → 预测 → 扫描 → 归约 → 预测 .....

## Earley算法过程示例-2

- (1)  $S \rightarrow NP \ VP$
- (2)  $NP \rightarrow N$
- (3)  $NP \rightarrow CS \text{ 的}$
- (4)  $CS \rightarrow NP \ V'$
- (5)  $VP \rightarrow V \ NP$
- (6)  $V' \rightarrow V \ V$

当前间隔 ①

$\langle NP \rightarrow N \cdot [0,1] \rangle$   
 $\langle S \rightarrow NP \cdot VP [0,1] \rangle$   
 $\langle CS \rightarrow NP \cdot V' [0,1] \rangle$   
 $\langle VP \rightarrow \cdot V \ NP [1,1] \rangle$   
 $\langle V' \rightarrow \cdot V \ V [1,1] \rangle$

当前间隔 ②

$\langle VP \rightarrow V \cdot NP [1,2] \rangle$   
 $\langle V' \rightarrow V \cdot V [1,2] \rangle$   
 $\langle NP \rightarrow \cdot N [2,2] \rangle$   
 $\langle NP \rightarrow \cdot CS \text{ 的} [2,2] \rangle$   
 $\langle CS \rightarrow \cdot NP \ V' [2,2] \rangle$   

---

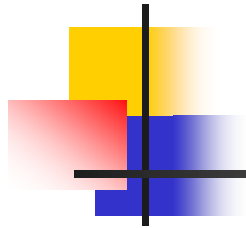
 $\langle S \rightarrow NP \cdot VP [0,1] \rangle$   
 $\langle CS \rightarrow NP \cdot V' [0,1] \rangle$

扫描

预测

-----> 保留状态





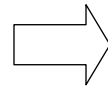
## Earley算法过程示例-3

- (1)  $S \rightarrow NP \ VP$
- (2)  $NP \rightarrow N$
- (3)  $NP \rightarrow CS \text{ 的}$
- (4)  $CS \rightarrow NP \ V'$
- (5)  $VP \rightarrow V \ NP$
- (6)  $V' \rightarrow V \ V$

当前间隔 ②

$\langle VP \rightarrow V \cdot NP \ [1,2] \rangle$   
 $\langle V' \rightarrow V \cdot V \ [1,2] \rangle$   
 $\langle NP \rightarrow \cdot N \ [2,2] \rangle$   
 $\langle NP \rightarrow \cdot CS \text{ 的} \ [2,2] \rangle$   
 $\langle CS \rightarrow \cdot NP \ V' \ [2,2] \rangle$

$\langle S \rightarrow NP \cdot VP \ [0,1] \rangle$   
 $\langle CS \rightarrow NP \cdot V' \ [0,1] \rangle$



当前间隔 ③

$\langle NP \rightarrow N \cdot \ [2,3] \rangle$  ———→ 扫描  
 $\langle VP \rightarrow V \ NP \cdot \ [1,3] \rangle$   
 $\langle S \rightarrow NP \ VP \cdot \ [0,3] \rangle$  ———→ 归约  
 $\langle CS \rightarrow NP \cdot V' \ [2,3] \rangle$  ———→ 归约  
 $\langle V' \rightarrow \cdot V \ V \ [3,3] \rangle$  ———→ 预测

$\langle S \rightarrow NP \cdot VP \ [0,1] \rangle$   
 $\langle CS \rightarrow NP \cdot V' \ [0,1] \rangle$   
 $\langle NP \rightarrow \cdot CS \text{ 的} \ [2,2] \rangle$   
 $\langle VP \rightarrow V \cdot NP \ [1,2] \rangle$

## Earley算法过程示例-4

- (1)  $S \rightarrow NP \ VP$
- (2)  $NP \rightarrow N$
- (3)  $NP \rightarrow CS$  的
- (4)  $CS \rightarrow NP \ V'$
- (5)  $VP \rightarrow V \ NP$
- (6)  $V' \rightarrow V \ V$

当前间隔 ③

|  |  |
|--|--|
| $\langle NP \rightarrow N \cdot [2,3] \rangle$             |  |
| $\langle VP \rightarrow V \ NP \cdot [1,3] \rangle$        |  |
| $\langle S \rightarrow NP \ VP \cdot [0,3] \rangle$ -----  |  |
| $\langle CS \rightarrow NP \cdot V' [2,3] \rangle$ -----   |  |
| $\langle V' \rightarrow \cdot V \ V \ [3,3] \rangle$       |  |
| -----  |  |
| $\langle S \rightarrow NP \cdot VP \ [0,1] \rangle$        |  |
| $\langle CS \rightarrow NP \cdot V' [0,1] \rangle$         |  |
| $\langle NP \rightarrow \cdot CS \text{ 的 } [2,2] \rangle$ |  |
| $\langle VP \rightarrow V \cdot NP [1,2] \rangle$          |  |

当前间隔 ④

$\langle V' \rightarrow V \cdot V [3,4] \rangle$  → 扫描

|  |
|--|
| $\langle S \rightarrow NP \cdot VP [0,1] \rangle$          |
| $\langle CS \rightarrow NP \cdot V' [0,1] \rangle$         |
| $\langle NP \rightarrow \cdot CS \text{ 的 } [2,2] \rangle$ |
| $\langle VP \rightarrow V \cdot NP [1,2] \rangle$          |
| $\langle CS \rightarrow NP \cdot V' [2,3] \rangle$         |
| $\langle S \rightarrow NP \ VP \cdot [0,3] \rangle$        |

## Earley算法过程示例-5

- (1)  $S \rightarrow NP \ VP$
- (2)  $NP \rightarrow N$
- (3)  $NP \rightarrow CS \text{ 的}$
- (4)  $CS \rightarrow NP \ V'$
- (5)  $VP \rightarrow V \ NP$
- (6)  $V' \rightarrow V \ V$

当前间隔 ④

|   |
|---|
| $\langle V' \rightarrow V \cdot V \ [3,4] \rangle$          |
| $\langle S \rightarrow NP \cdot VP \ [0,1] \rangle$         |
| $\langle CS \rightarrow NP \cdot V' \ [0,1] \rangle$        |
| $\langle NP \rightarrow \cdot CS \text{ 的} \ [2,2] \rangle$ |
| $\langle VP \rightarrow V \cdot NP \ [1,2] \rangle$         |
| $\langle CS \rightarrow NP \cdot V' \ [2,3] \rangle$        |
| $\langle S \rightarrow NP \ VP \cdot \ [0,3] \rangle$       |

当前间隔 ⑤

|  |    |
|--|----|
| $\langle V' \rightarrow V \ V \cdot \ [3,5] \rangle$       | 扫描 |
| $\langle CS \rightarrow NP \ V' \cdot \ [2,5] \rangle$     | 归约 |
| $\langle NP \rightarrow CS \cdot \text{的} \ [2,5] \rangle$ | 归约 |
| <hr/>  |    |
| $\langle S \rightarrow NP \cdot VP \ [0,1] \rangle$        |    |
| $\langle CS \rightarrow NP \cdot V' \ [0,1] \rangle$       |    |
| $\langle VP \rightarrow V \cdot NP \ [1,2] \rangle$        |    |
| $\langle S \rightarrow NP \ VP \cdot \ [0,3] \rangle$      |    |

## Earley算法过程示例-6

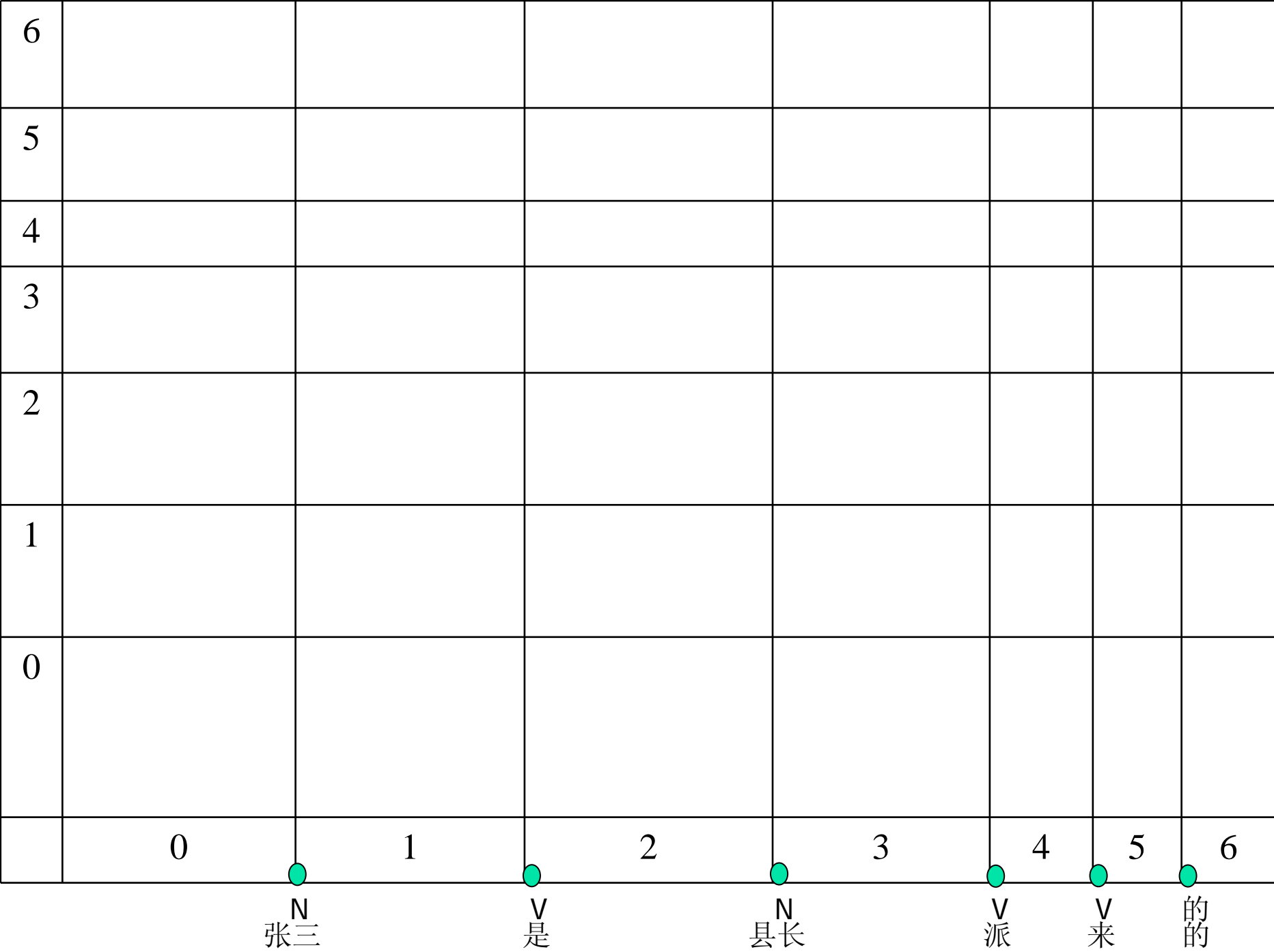
- (1)  $S \rightarrow NP \ VP$
- (2)  $NP \rightarrow N$
- (3)  $NP \rightarrow CS \text{ 的}$
- (4)  $CS \rightarrow NP \ V'$
- (5)  $VP \rightarrow V \ NP$
- (6)  $V' \rightarrow V \ V$

当前间隔 ⑤

|  |  |
|--|--|
| $\langle V' \rightarrow V \ V \cdot [3,5] \rangle$       |  |
| $\langle CS \rightarrow NP \ V' \cdot [2,5] \rangle$     |  |
| $\langle NP \rightarrow CS \cdot \text{的} [2,5] \rangle$ |  |
| <hr/>  |  |
| $\langle S \rightarrow NP \cdot VP [0,1] \rangle$        |  |
| $\langle CS \rightarrow NP \cdot V' [0,1] \rangle$       |  |
| $\langle VP \rightarrow V \cdot NP [1,2] \rangle$        |  |
| $\langle S \rightarrow NP \ VP \cdot [0,3] \rangle$      |  |







当前间隔 ⑥

|   |    |
|---|----|
| $\langle NP \rightarrow CS \text{ 的} \cdot [2,6] \rangle$ | 扫描 |
| $\langle VP \rightarrow V \ NP \cdot [1,6] \rangle$       | 归约 |
| $\langle S \rightarrow NP \ VP \cdot [0,6] \rangle$       | 归约 |
| <hr/>   |    |
| $\langle CS \rightarrow NP \cdot V' [0,1] \rangle$        |    |
| $\langle S \rightarrow NP \ VP \cdot [0,3] \rangle$       |    |









|   |             |        |         |        |        |        |   |    |
|---|-------------|--------|---------|--------|--------|--------|---|----|
| 6 |             |        |         |        |        |        |   |    |
| 5 |             |        |         |        |        |        |   |    |
| 4 |             |        |         |        |        |        |   |    |
| 3 |             |        |         |        |        |        |   |    |
| 2 |             |        |         |        |        |        |   |    |
| 1 |             |        |         |        |        |        |   |    |
| 0 | S → · NP VP |        |         |        |        |        |   | 种子 |
|   | 0           | 1      | 2       | 3      | 4      | 5      | 6 |    |
|   | N<br>张三     | V<br>是 | N<br>县长 | V<br>派 | V<br>来 | 的<br>的 |   |    |

|   |   |   |                    |   |                     |   |                    |                    |                      |
|---|---|---|--------------------|---|---------------------|---|--------------------|--------------------|----------------------|
| 6 |   |   |                    |   |                     |   |                    |                    |                      |
| 5 |   |   |                    |   |                     |   |                    |                    |                      |
| 4 |   |   |                    |   |                     |   |                    |                    |                      |
| 3 |   |   |                    |   |                     |   |                    |                    |                      |
| 2 |   |   |                    |   |                     |   |                    |                    |                      |
| 1 |   |   |                    |   |                     |   |                    |                    |                      |
| 0 | CS → · NP V '<br>NP → · N<br>NP → · CS 的<br>S → · NP VP |   |                    |   |                     |   |                    |                    | 预测<br>预测<br>预测<br>种子 |
|   | 0   | 1 | 2                  | 3 | 4                   | 5 | 6                  |                    |                      |
|   | <div>N<br/>张三</div>                                     |   | <div>V<br/>是</div> |   | <div>N<br/>县长</div> |   | <div>V<br/>派</div> | <div>V<br/>来</div> | <div>的<br/>的</div>   |

|   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|
| 6 |   |   |   |   |   |   |   |   |   |
| 5 |   |   |   |   |   |   |   |   |   |
| 4 |   |   |   |   |   |   |   |   |   |
| 3 |   |   |   |   |   |   |   |   |   |
| 2 |   |   |   |   |   |   |   |   |   |
| 1 | NP → N .  |   |   |   |   |   |   |   | 扫描  |
| 0 | CS → . NP V '<br>NP → . N<br>NP → . CS 的<br>S → . NP VP                             |   |   |   |   |   |   |   | 预测<br>预测<br>预测<br>种子  |
|   | 0   | 1 | 2   | 3 | 4   | 5 | 6   |   |   |
|   |  |   |  |   |  |   |  |  |  |
|   | N<br>张三   |   | V<br>是  |   | N<br>县长   |   | V<br>派  | V<br>来  | 的<br>的  |



|   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|
| 6 |   |   |   |   |   |   |   |   |   |
| 5 |   |   |   |   |   |   |   |   |   |
| 4 |   |   |   |   |   |   |   |   |   |
| 3 |   |   |   |   |   |   |   |   |   |
| 2 |   |   |   |   |   |   |   |   |   |
| 1 | S → NP · VP<br>CS → NP · V '<br>NP → N ·  |   |   |   |   |   |   |   | 归约<br>扫描  |
| 0 | CS → · NP V '<br>NP → · N<br>NP → · CS 的<br>S → · NP VP                             |   |   |   |   |   |   |   | 预测<br>预测<br>预测<br>种子  |
|   | 0   | 1 | 2   | 3 | 4   | 5 | 6   |   |   |
|   |  |   |  |   |  |   |  |  |  |
|   | N<br>张三   |   | V<br>是  |   | N<br>县长   |   | V<br>派  | V<br>来  | 的<br>的  |

|   |  |   |   |   |   |   |                      |
|---|--|---|---|---|---|---|----------------------|
| 6 |  |   |   |   |   |   |                      |
| 5 |  |   |   |   |   |   |                      |
| 4 |  |   |   |   |   |   |                      |
| 3 |  |   |   |   |   |   |                      |
| 2 |  |   |   |   |   |   |                      |
| 1 | $S \rightarrow NP \cdot VP$<br>$CS \rightarrow NP \cdot V'$<br>$NP \rightarrow N \cdot$                                | $VP \rightarrow \cdot V NP$<br>$V' \rightarrow \cdot V V$ |   |   |   |   | 预测<br>归约<br>扫描       |
| 0 | $CS \rightarrow \cdot NP V'$<br>$NP \rightarrow \cdot N$<br>$NP \rightarrow \cdot CS$ 的<br>$S \rightarrow \cdot NP VP$ |   |   |   |   |   | 预测<br>预测<br>预测<br>种子 |
|   | 0  | 1   | 2 | 3 | 4 | 5 | 6                    |

N  
张三







V  
是

N  
县长

V  
派

V  
来

的  
的

|   |  |   |   |   |   |   |   |                      |
|---|--|---|---|---|---|---|---|----------------------|
| 6 |  |   |   |   |   |   |   |                      |
| 5 |  |   |   |   |   |   |   |                      |
| 4 |  |   |   |   |   |   |   |                      |
| 3 |  |   |   |   |   |   |   |                      |
| 2 |  | VP → V · NP<br>V' → V · V   |   |   |   |   |   | 扫描                   |
| 1 | S → NP · VP<br>CS → NP · V'<br>NP → N ·                | VP → · V NP<br>V' → · V V   |   |   |   |   |   | 预测<br>归约<br>扫描       |
| 0 | CS → · NP V'<br>NP → · N<br>NP → · CS 的<br>S → · NP VP |   |   |   |   |   |   | 预测<br>预测<br>预测<br>种子 |
|   | 0  | 1   | 2   | 3   | 4   | 5   | 6   |                      |
|   |  |  |  |  |  |  |  |                      |
|   | N<br>张三  |   | V<br>是  |   | N<br>县长   | V<br>派  | V<br>来  | 的<br>的               |

|   |  |                           |                         |   |   |   |                      |
|---|--|---------------------------|-------------------------|---|---|---|----------------------|
| 6 |  |                           |                         |   |   |   |                      |
| 5 |  |                           |                         |   |   |   |                      |
| 4 |  |                           |                         |   |   |   |                      |
| 3 |  |                           |                         |   |   |   |                      |
| 2 |  | VP → V · NP<br>V' → V · V | NP → · N<br>NP → · CS 的 |   |   |   | 预测扫描                 |
| 1 | S → NP · VP<br>CS → NP · V'<br>NP → N ·                | VP → · V NP<br>V' → · V V |                         |   |   |   | 预测归约扫描               |
| 0 | CS → · NP V'<br>NP → · N<br>NP → · CS 的<br>S → · NP VP |                           |                         |   |   |   | 预测<br>预测<br>预测<br>种子 |
|   | 0  | 1                         | 2                       | 3 | 4 | 5 | 6                    |

N  
张三

V  
是

N  
县长







V  
派

V  
来

的  
的

|   |  |                           |   |        |        |   |                      |
|---|--|---------------------------|---|--------|--------|---|----------------------|
| 6 |  |                           |   |        |        |   |                      |
| 5 |  |                           |   |        |        |   |                      |
| 4 |  |                           |   |        |        |   |                      |
| 3 |  |                           |   |        |        |   |                      |
| 2 |  | VP → V · NP<br>V' → V · V | CS → · NP V'<br>NP → · N<br>NP → · CS 的 |        |        |   | 预测<br>预测<br>扫描       |
| 1 | S → NP · VP<br>CS → NP · V'<br>NP → N ·                | VP → · V NP<br>V' → · V V |   |        |        |   | 预测<br>归约<br>扫描       |
| 0 | CS → · NP V'<br>NP → · N<br>NP → · CS 的<br>S → · NP VP |                           |   |        |        |   | 预测<br>预测<br>预测<br>种子 |
|   | 0  | 1                         | 2                                       | 3      | 4      | 5 | 6                    |
|   | N<br>张三  | V<br>是                    | N<br>县长                                 | V<br>派 | V<br>来 | 的 |                      |

|   |  |                           |   |         |        |        |                      |
|---|--|---------------------------|---|---------|--------|--------|----------------------|
| 6 |  |                           |   |         |        |        |                      |
| 5 |  |                           |   |         |        |        |                      |
| 4 |  |                           |   |         |        |        |                      |
| 3 |  |                           | NP → N .                                |         |        |        | 扫描                   |
| 2 |  | VP → V . NP<br>V' → V . V | CS → . NP V'<br>NP → . N<br>NP → . CS 的 |         |        |        | 预测<br>预测<br>扫描       |
| 1 | S → NP . VP<br>CS → NP . V'<br>NP → N .                | VP → . V NP<br>V' → . V V |   |         |        |        | 预测<br>归约<br>扫描       |
| 0 | CS → . NP V'<br>NP → . N<br>NP → . CS 的<br>S → . NP VP |                           |   |         |        |        | 预测<br>预测<br>预测<br>种子 |
|   | 0  | 1                         | 2                                       | 3       | 4      | 5      | 6                    |
|   |  | N<br>张三                   | V<br>是                                  | N<br>县长 | V<br>派 | V<br>来 | 的<br>的               |

|   |   |                           |   |   |   |   |   |
|---|---|---------------------------|---|---|---|---|---|
| 6 |   |                           |   |   |   |   |   |
| 5 |   |                           |   |   |   |   |   |
| 4 |   |                           |   |   |   |   |   |
| 3 |   |                           | CS → NP · V '<br>NP → N ·   |   |   |   | 归约<br>扫描  |
| 2 |   | VP → V · NP<br>V' → V · V | CS → · NP V '<br>NP → · N<br>NP → · CS 的  |   |   |   | 预测<br>预测<br>扫描  |
| 1 | S → NP · VP<br>CS → NP · V '<br>NP → N ·  | VP → · V NP<br>V' → · V V |   |   |   |   | 预测<br>归约<br>扫描  |
| 0 | CS → · NP V '<br>NP → · N<br>NP → · CS 的<br>S → · NP VP                             |                           |   |   |   |   | 预测<br>预测<br>预测<br>种子  |
|   | 0   | 1                         | 2   | 3   | 4   | 5   | 6   |
|   |  |                           |  |  |  |  |  |
|   | N<br>张三   |                           | V<br>是  | N<br>县长   | V<br>派  | V<br>来  | 的<br>的  |

|   |   |                           |  |         |        |        |                      |
|---|---|---------------------------|--|---------|--------|--------|----------------------|
| 6 |   |                           |  |         |        |        |                      |
| 5 |   |                           |  |         |        |        |                      |
| 4 |   |                           |  |         |        |        |                      |
| 3 |   | VP → V NP .               | CS → NP . V '<br>NP → N .                |         |        |        | 归约<br>扫描             |
| 2 |   | VP → V . NP<br>V' → V . V | CS → . NP V '<br>NP → . N<br>NP → . CS 的 |         |        |        | 预测<br>预测<br>扫描       |
| 1 | S → NP . VP<br>CS → NP . V '<br>NP → N .                | VP → . V NP<br>V' → . V V |  |         |        |        | 预测<br>归约<br>扫描       |
| 0 | CS → . NP V '<br>NP → . N<br>NP → . CS 的<br>S → . NP VP |                           |  |         |        |        | 预测<br>预测<br>预测<br>种子 |
|   | 0   | 1                         | 2  | 3       | 4      | 5      | 6                    |
|   |   | N<br>张三                   | V<br>是                                   | N<br>县长 | V<br>派 | V<br>来 | 的<br>的               |



|   |  |   |   |   |   |   |                      |
|---|--|---|---|---|---|---|----------------------|
| 6 |  |   |   |   |   |   |                      |
| 5 |  |   |   |   |   |   |                      |
| 4 |  |   |   |   |   |   |                      |
| 3 | $S \rightarrow NP VP \cdot$  | $VP \rightarrow V NP \cdot$                               | $CS \rightarrow NP \cdot V'$<br>$NP \rightarrow N \cdot$  |   |   |   | 归约扫描                 |
| 2 |  | $VP \rightarrow V \cdot NP$<br>$V' \rightarrow V \cdot V$ | $CS \rightarrow \cdot NP V'$<br>$NP \rightarrow \cdot N$<br>$NP \rightarrow \cdot CS \text{ 的}$ |   |   |   | 预测<br>预测<br>扫描       |
| 1 | $S \rightarrow NP \cdot VP$<br>$CS \rightarrow NP \cdot V'$<br>$NP \rightarrow N \cdot$  | $VP \rightarrow \cdot V NP$<br>$V' \rightarrow \cdot V V$ |   |   |   |   | 预测<br>归约<br>扫描       |
| 0 | $CS \rightarrow \cdot NP V'$<br>$NP \rightarrow \cdot N$<br>$NP \rightarrow \cdot CS \text{ 的}$<br>$S \rightarrow \cdot NP VP$ |   |   |   |   |   | 预测<br>预测<br>预测<br>种子 |
|   | 0  | 1   | 2   | 3 | 4 | 5 | 6                    |

N  
张三

V  
是

N  
县长

V  
派

V  
来

的  
的

|   |  |   |   |                            |   |   |                      |
|---|--|---|---|----------------------------|---|---|----------------------|
| 6 |  |   |   |                            |   |   |                      |
| 5 |  |   |   |                            |   |   |                      |
| 4 |  |   |   |                            |   |   |                      |
| 3 | $S \rightarrow NP VP \cdot$  | $VP \rightarrow V NP \cdot$                               | $CS \rightarrow NP \cdot V'$<br>$NP \rightarrow N \cdot$                                | $V' \rightarrow \cdot V V$ |   |   | 预测<br>归约<br>扫描       |
| 2 |  | $VP \rightarrow V \cdot NP$<br>$V' \rightarrow V \cdot V$ | $CS \rightarrow \cdot NP V'$<br>$NP \rightarrow \cdot N$<br>$NP \rightarrow \cdot CS$ 的 |                            |   |   | 预测<br>预测<br>扫描       |
| 1 | $S \rightarrow NP \cdot VP$<br>$CS \rightarrow NP \cdot V'$<br>$NP \rightarrow N \cdot$                                | $VP \rightarrow \cdot V NP$<br>$V' \rightarrow \cdot V V$ |   |                            |   |   | 预测<br>归约<br>扫描       |
| 0 | $CS \rightarrow \cdot NP V'$<br>$NP \rightarrow \cdot N$<br>$NP \rightarrow \cdot CS$ 的<br>$S \rightarrow \cdot NP VP$ |   |   |                            |   |   | 预测<br>预测<br>预测<br>种子 |
|   | 0  | 1   | 2   | 3                          | 4 | 5 | 6                    |

N  
张三

V  
是

N  
县长

V  
派

V  
来

的  
的

|   |  |   |   |                            |   |   |                      |
|---|--|---|---|----------------------------|---|---|----------------------|
| 6 |  |   |   |                            |   |   |                      |
| 5 |  |   |   |                            |   |   |                      |
| 4 |  |   |   | $V' \rightarrow V \cdot V$ |   |   | 扫描                   |
| 3 | $S \rightarrow NP VP \cdot$  | $VP \rightarrow V NP \cdot$                               | $CS \rightarrow NP \cdot V'$<br>$NP \rightarrow N \cdot$                                | $V' \rightarrow \cdot V V$ |   |   | 预测<br>归约<br>扫描       |
| 2 |  | $VP \rightarrow V \cdot NP$<br>$V' \rightarrow V \cdot V$ | $CS \rightarrow \cdot NP V'$<br>$NP \rightarrow \cdot N$<br>$NP \rightarrow \cdot CS$ 的 |                            |   |   | 预测<br>预测<br>扫描       |
| 1 | $S \rightarrow NP \cdot VP$<br>$CS \rightarrow NP \cdot V'$<br>$NP \rightarrow N \cdot$                                | $VP \rightarrow \cdot V NP$<br>$V' \rightarrow \cdot V V$ |   |                            |   |   | 预测<br>归约<br>扫描       |
| 0 | $CS \rightarrow \cdot NP V'$<br>$NP \rightarrow \cdot N$<br>$NP \rightarrow \cdot CS$ 的<br>$S \rightarrow \cdot NP VP$ |   |   |                            |   |   | 预测<br>预测<br>预测<br>种子 |
|   | 0  | 1   | 2   | 3                          | 4 | 5 | 6                    |

N  
张三







V  
是

N  
县长

V  
派

V  
来

的  
的

|   |  |   |   |   |   |   |   |
|---|--|---|---|---|---|---|---|
| 6 |  |   |   |   |   |   |   |
| 5 |  |   |   | $V' \rightarrow V V \cdot$  |   |   | 扫描  |
| 4 |  |   |   | $V' \rightarrow V \cdot V$  |   |   | 扫描  |
| 3 | $S \rightarrow NP VP \cdot$  | $VP \rightarrow V NP \cdot$                               | $CS \rightarrow NP \cdot V'$<br>$NP \rightarrow N \cdot$                                | $V' \rightarrow \cdot V V$  |   |   | 预测<br>归约<br>扫描  |
| 2 |  | $VP \rightarrow V \cdot NP$<br>$V' \rightarrow V \cdot V$ | $CS \rightarrow \cdot NP V'$<br>$NP \rightarrow \cdot N$<br>$NP \rightarrow \cdot CS$ 的 |   |   |   | 预测<br>预测<br>扫描  |
| 1 | $S \rightarrow NP \cdot VP$<br>$CS \rightarrow NP \cdot V'$<br>$NP \rightarrow N \cdot$                                | $VP \rightarrow \cdot V NP$<br>$V' \rightarrow \cdot V V$ |   |   |   |   | 预测<br>归约<br>扫描  |
| 0 | $CS \rightarrow \cdot NP V'$<br>$NP \rightarrow \cdot N$<br>$NP \rightarrow \cdot CS$ 的<br>$S \rightarrow \cdot NP VP$ |   |   |   |   |   | 预测<br>预测<br>预测<br>种子  |
|   | 0  | 1   | 2   | 3   | 4   | 5   | 6   |
|   |                                     |   |      |  |  |  |  |
|   | N<br>张三  |   | V<br>是  | N<br>县长   | V<br>派  | V<br>来  | 的<br>的  |

|   |   |                            |  |             |        |        |                      |
|---|---|----------------------------|--|-------------|--------|--------|----------------------|
| 6 |   |                            |  |             |        |        |                      |
| 5 |   |                            | NP → CS . 的<br>CS → NP V ' .             | V ' → V V . |        |        | 归约<br>扫描             |
| 4 |   |                            |  | V ' → V . V |        |        | 扫描                   |
| 3 | S → NP VP .   | VP → V NP .                | CS → NP . V '<br>NP → N .                | V ' → . V V |        |        | 预测<br>归约<br>扫描       |
| 2 |   | VP → V . NP<br>V ' → V . V | CS → . NP V '<br>NP → . N<br>NP → . CS 的 |             |        |        | 预测<br>预测<br>扫描       |
| 1 | S → NP . VP<br>CS → NP . V '<br>NP → N .                | VP → . V NP<br>V ' → . V V |  |             |        |        | 预测<br>归约<br>扫描       |
| 0 | CS → . NP V '<br>NP → . N<br>NP → . CS 的<br>S → . NP VP |                            |  |             |        |        | 预测<br>预测<br>预测<br>种子 |
|   | 0   | 1                          | 2  | 3           | 4      | 5      | 6                    |
|   |   | N<br>张三                    | V<br>是                                   | N<br>县长     | V<br>派 | V<br>来 | 的<br>的               |

|   |  |   |   |                            |   |   |                      |
|---|--|---|---|----------------------------|---|---|----------------------|
| 6 |  |   | $NP \rightarrow CS \text{ 的} \cdot$   |                            |   |   | 扫描                   |
| 5 |  |   | $NP \rightarrow CS \cdot \text{ 的}$<br>$CS \rightarrow NP V' \cdot$                             | $V' \rightarrow V V \cdot$ |   |   | 归约<br>扫描             |
| 4 |  |   |   | $V' \rightarrow V \cdot V$ |   |   | 扫描                   |
| 3 | $S \rightarrow NP VP \cdot$  | $VP \rightarrow V NP \cdot$                               | $CS \rightarrow NP \cdot V'$<br>$NP \rightarrow N \cdot$  | $V' \rightarrow \cdot V V$ |   |   | 预测<br>归约<br>扫描       |
| 2 |  | $VP \rightarrow V \cdot NP$<br>$V' \rightarrow V \cdot V$ | $CS \rightarrow \cdot NP V'$<br>$NP \rightarrow \cdot N$<br>$NP \rightarrow \cdot CS \text{ 的}$ |                            |   |   | 预测<br>预测<br>扫描       |
| 1 | $S \rightarrow NP \cdot VP$<br>$CS \rightarrow NP \cdot V'$<br>$NP \rightarrow N \cdot$  | $VP \rightarrow \cdot V NP$<br>$V' \rightarrow \cdot V V$ |   |                            |   |   | 预测<br>归约<br>扫描       |
| 0 | $CS \rightarrow \cdot NP V'$<br>$NP \rightarrow \cdot N$<br>$NP \rightarrow \cdot CS \text{ 的}$<br>$S \rightarrow \cdot NP VP$ |   |   |                            |   |   | 预测<br>预测<br>预测<br>种子 |
|   | 0  | 1   | 2   | 3                          | 4 | 5 | 6                    |

N  
张三

V  
是

N  
县长

V  
派

V  
来

的  
的

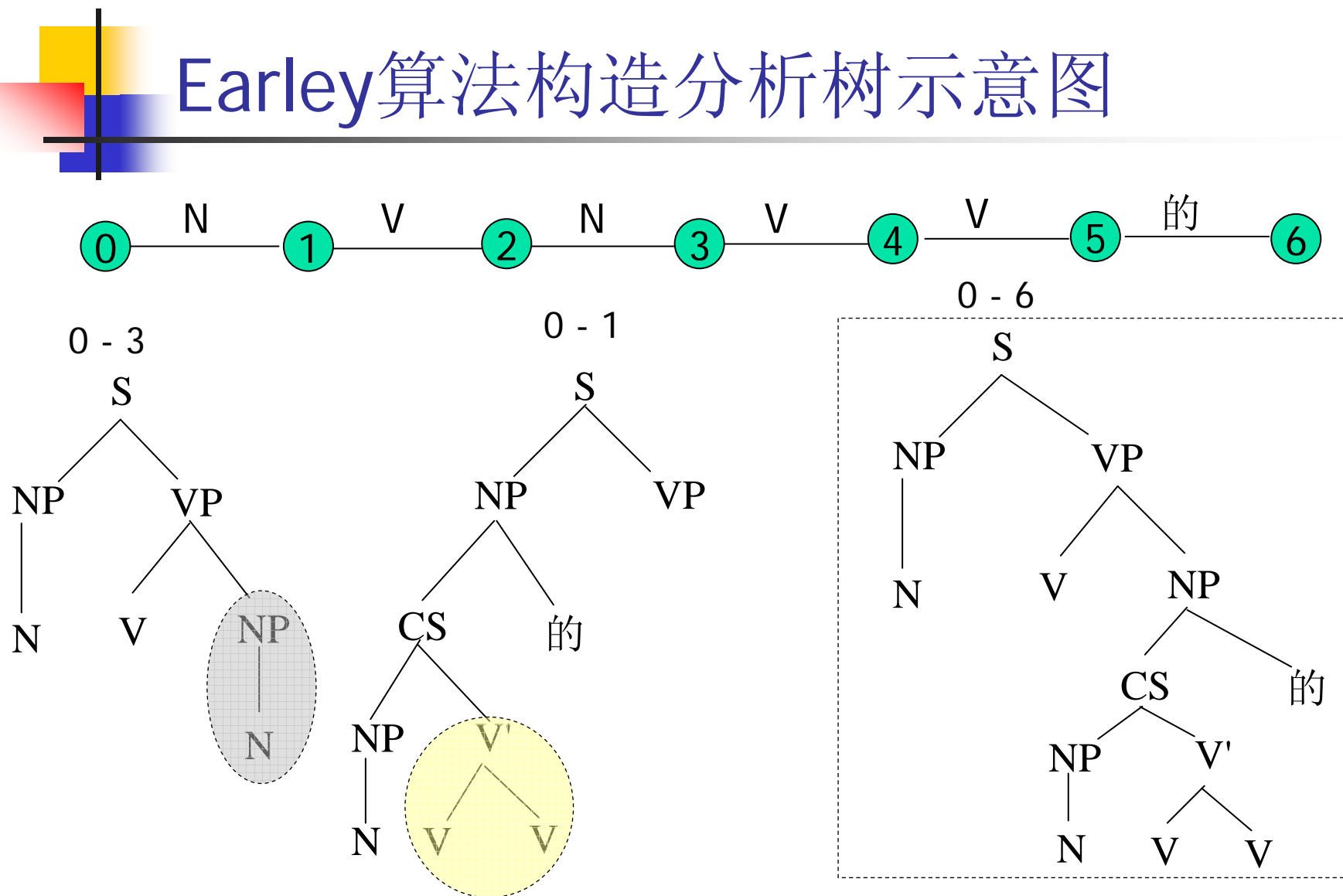
|   |  |   |   |                              |        |        |                      |
|---|--|---|---|------------------------------|--------|--------|----------------------|
| 6 |  | $VP \rightarrow V \ NP \ .$                                   | $NP \rightarrow CS \ 的 \ .$   |                              |        |        | 归约<br>扫描             |
| 5 |  |   | $NP \rightarrow CS \ . \ 的$<br>$CS \rightarrow NP \ V' \ .$                                 | $V' \rightarrow V \ V \ .$   |        |        | 归约<br>扫描             |
| 4 |  |   |   | $V' \rightarrow V \ . \ V$   |        |        | 扫描                   |
| 3 | $S \rightarrow NP \ VP \ .$  | $VP \rightarrow V \ NP \ .$                                   | $CS \rightarrow NP \ . \ V'$<br>$NP \rightarrow N \ .$                                      | $V' \rightarrow \cdot V \ V$ |        |        | 预测<br>归约<br>扫描       |
| 2 |  | $VP \rightarrow V \ . \ NP$<br>$V' \rightarrow V \ . \ V$     | $CS \rightarrow \cdot NP \ V'$<br>$NP \rightarrow \cdot N$<br>$NP \rightarrow \cdot CS \ 的$ |                              |        |        | 预测<br>预测<br>扫描       |
| 1 | $S \rightarrow NP \ . \ VP$<br>$CS \rightarrow NP \ . \ V'$<br>$NP \rightarrow N \ .$  | $VP \rightarrow \cdot V \ NP$<br>$V' \rightarrow \cdot V \ V$ |   |                              |        |        | 预测<br>归约<br>扫描       |
| 0 | $CS \rightarrow \cdot NP \ V'$<br>$NP \rightarrow \cdot N$<br>$NP \rightarrow \cdot CS \ 的$<br>$S \rightarrow \cdot NP \ VP$ |   |   |                              |        |        | 预测<br>预测<br>预测<br>种子 |
|   | 0  | 1   | 2   | 3                            | 4      | 5      | 6                    |
|   |  | N<br>张三   | V<br>是  | N<br>县长                      | V<br>派 | V<br>来 | 的<br>的               |

|   |  |   |   |                            |        |        |                      |
|---|--|---|---|----------------------------|--------|--------|----------------------|
| 6 | $S \rightarrow NP VP \cdot$  | $VP \rightarrow V NP \cdot$                               | $NP \rightarrow CS \text{ 的} \cdot$   |                            |        |        | 归约<br>扫描             |
| 5 |  |   | $NP \rightarrow CS \cdot \text{ 的}$<br>$CS \rightarrow NP V' \cdot$                             | $V' \rightarrow V V \cdot$ |        |        | 归约<br>扫描             |
| 4 |  |   |   | $V' \rightarrow V \cdot V$ |        |        | 扫描                   |
| 3 | $S \rightarrow NP VP \cdot$  | $VP \rightarrow V NP \cdot$                               | $CS \rightarrow NP \cdot V'$<br>$NP \rightarrow N \cdot$  | $V' \rightarrow \cdot V V$ |        |        | 预测<br>归约<br>扫描       |
| 2 |  | $VP \rightarrow V \cdot NP$<br>$V' \rightarrow V \cdot V$ | $CS \rightarrow \cdot NP V'$<br>$NP \rightarrow \cdot N$<br>$NP \rightarrow \cdot CS \text{ 的}$ |                            |        |        | 预测<br>预测<br>扫描       |
| 1 | $S \rightarrow NP \cdot VP$<br>$CS \rightarrow NP \cdot V'$<br>$NP \rightarrow N \cdot$  | $VP \rightarrow \cdot V NP$<br>$V' \rightarrow \cdot V V$ |   |                            |        |        | 预测<br>归约<br>扫描       |
| 0 | $CS \rightarrow \cdot NP V'$<br>$NP \rightarrow \cdot N$<br>$NP \rightarrow \cdot CS \text{ 的}$<br>$S \rightarrow \cdot NP VP$ |   |   |                            |        |        | 预测<br>预测<br>预测<br>种子 |
|   | 0  | 1   | 2   | 3                          | 4      | 5      | 6                    |
|   |  | N<br>张三   | V<br>是  | N<br>县长                    | V<br>派 | V<br>来 | 的<br>的               |



|   |  |   |   |                            |        |        |                      |
|---|--|---|---|----------------------------|--------|--------|----------------------|
| 6 | $S \rightarrow NP VP \cdot$  | $VP \rightarrow V NP \cdot$                               | $NP \rightarrow CS \text{ 的} \cdot$   |                            |        |        | 归约<br>扫描             |
| 5 |  |   | $NP \rightarrow CS \cdot \text{ 的}$<br>$CS \rightarrow NP V' \cdot$                             | $V' \rightarrow V V \cdot$ |        |        | 归约<br>扫描             |
| 4 |  |   |   | $V' \rightarrow V \cdot V$ |        |        | 扫描                   |
| 3 | $S \rightarrow NP VP \cdot$  | $VP \rightarrow V NP \cdot$                               | $CS \rightarrow NP \cdot V'$<br>$NP \rightarrow N \cdot$  | $V' \rightarrow \cdot V V$ |        |        | 预测<br>归约<br>扫描       |
| 2 |  | $VP \rightarrow V \cdot NP$<br>$V' \rightarrow V \cdot V$ | $CS \rightarrow \cdot NP V'$<br>$NP \rightarrow \cdot N$<br>$NP \rightarrow \cdot CS \text{ 的}$ |                            |        |        | 预测<br>预测<br>扫描       |
| 1 | $S \rightarrow NP \cdot VP$<br>$CS \rightarrow NP \cdot V'$<br>$NP \rightarrow N \cdot$  | $VP \rightarrow \cdot V NP$<br>$V' \rightarrow \cdot V V$ |   |                            |        |        | 预测<br>归约<br>扫描       |
| 0 | $CS \rightarrow \cdot NP V'$<br>$NP \rightarrow \cdot N$<br>$NP \rightarrow \cdot CS \text{ 的}$<br>$S \rightarrow \cdot NP VP$ |   |   |                            |        |        | 预测<br>预测<br>预测<br>种子 |
|   | 0  | 1   | 2   | 3                          | 4      | 5      | 6                    |
|   |  | N<br>张三   | V<br>是  | N<br>县长                    | V<br>派 | V<br>来 | 的                    |

# Earley算法构造分析树示意图





## 4 小结

- 语法模型：保证句法结构分析的准确    把事情做对
- 分析算法：保证句法结构分析的效率    把事情做好

TAG（树邻接语法）

语法模型

HPSG（中心驱动的短语结构语法）

FUG（功能合一语法）

LFG（词汇功能语法）

PCFG（概率上下文无关文法）

Link Grammar（链语法）

Dependency Grammar（依存语法）

.....

CYK算法

分析算法

ATN 算法

Probabilistic Earley算法

Chart-based Parsing算法

依存句法分析算法

链语法分析算法

.....



## 进一步阅读文献

---

- 冯志伟等译（2005）《自然语言处理综论》第1章，第10.3.2，第13章。
- 陆汝钤（2000）《人工智能》（上册），科学出版社
- 陆汝钤（1993）《数学·计算·逻辑》，湖南教育出版社
- Earley, J. (1970) An Efficient Context-Free-Parsing Algorithm, Communication of ACM, Vol. 6, No. 8, pp94-102.
- Tomita, Masaru, 1987, An Efficient Augmented Context Free Parsing Algorithm, Computational Linguistics Vol.13, Issue 1-2, pp.31-46.
- Dick Grune & Criel Jacobs, 1990, Parsing Techniques: A Practical Guide, First Published in 1990 by ELLIS HORWOOD LIMITED, Reprinted in 1997, 1998
- 刘挺、马金山，2009，汉语自动句法分析的理论与方法，《当代语言学》2009年第2期。