



第八章 句法分析（一）

詹卫东

<http://ccl.pku.edu.cn/doubtfire>



提纲

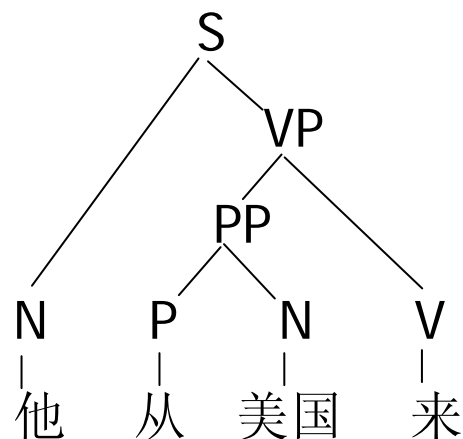
- 1 句法分析概说
- 2 自顶向下分析法
- 3 自底向上分析法
- 4 左角分析法（自底向上与自顶向下结合）
- 5 CYK分析法（自底向上分析法）
- 6 ATN分析法（自顶向下分析法）
- 7 Earley算法（自顶向下分析法）

1 句法分析概说

- 文法 G —— 语言 L
- 分析算法
 - (1) 给定一个字符串S, 判定S是否属于L
 - (2) 给定一个字符串S, 如果S属于L, 给出S对应的树结构

他从美国来

* 他从来美国



[S [N 他] [VP [PP [P 从] [N 美国]] [V 来]]] 3



句法分析的不同类型

- 自顶向下 (Top-down) 句法分析 predicative parser
- 自底向上 (Bottom-up) 句法分析 shift-reduce parser

- 确定性句法分析 deterministic parser
- 非确定性句法分析 non-deterministic parser

- 传统的非概率分析方法
- 概率方法 (PCFG)

- 完全句法分析
- 部分句法分析 (partial parsing / shallow parsing)

自然语言 —— 歧义无处不在

- 我是县长
我是县长派来的

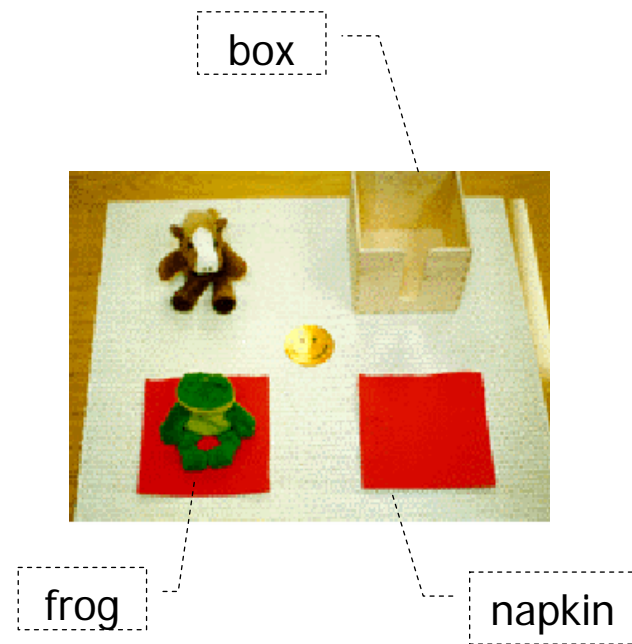
garden-path sentence/ 花园路径句

- Put the frog on the napkin in the box

- 麻子无头发黑脸大脚不大好看

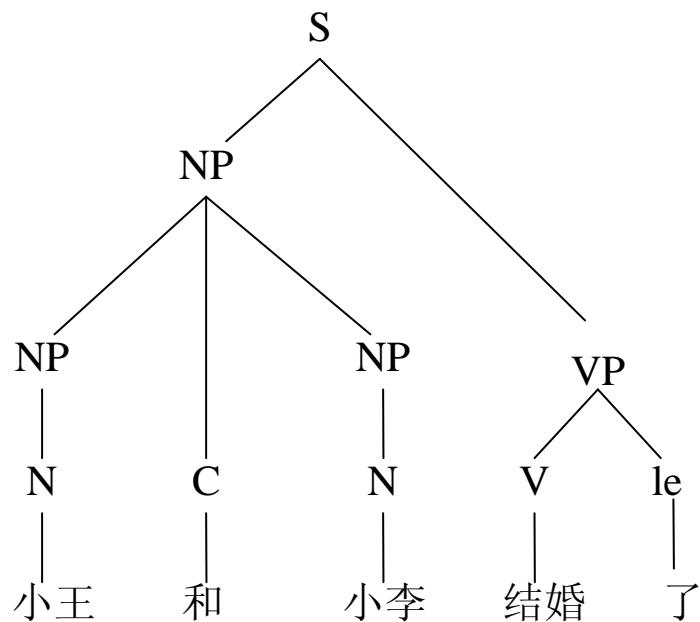
麻子无 头发黑 脸大脚不大 好看

麻子 无头发 黑脸 大脚 不大好看

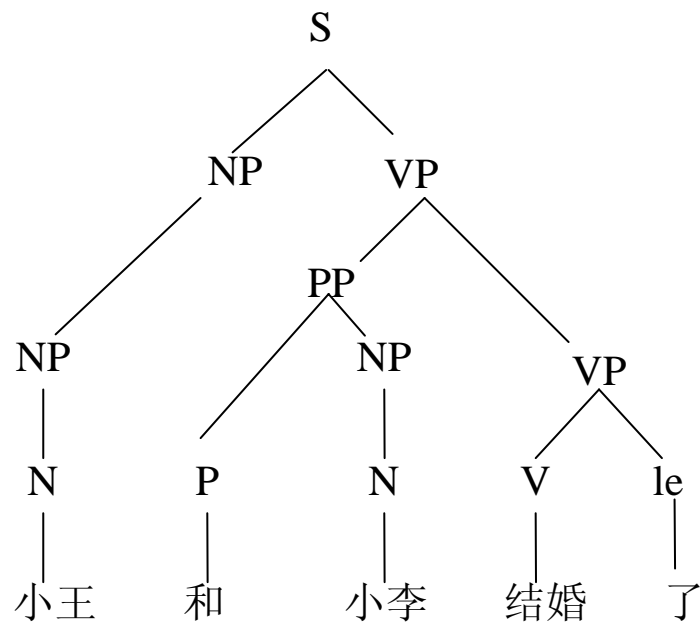


自然语言 —— 歧义无处不在（续）

小王和小李结婚了



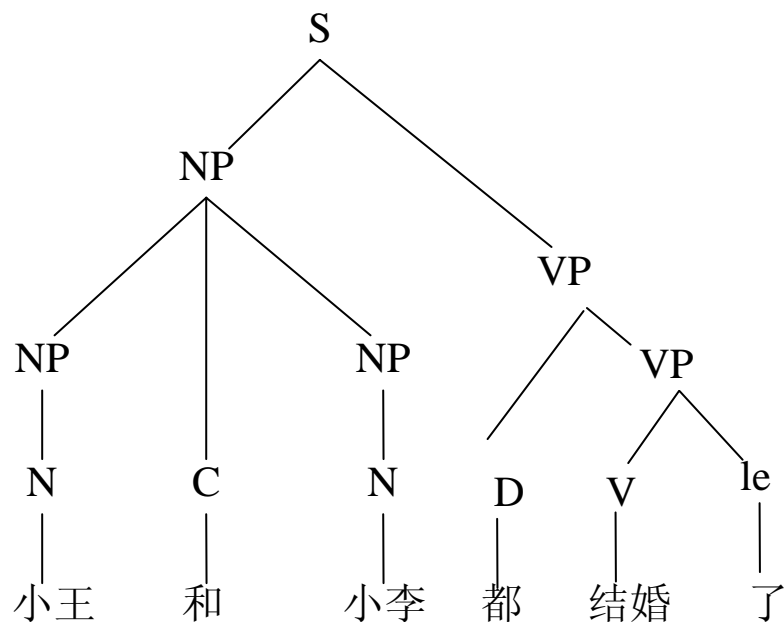
小王和小李都结婚了



自然语言 —— 歧义无处不在（续）

小王和小李结婚了

小王和小李都结婚了



自然语言文法 —— over generalization

实例	大眼睛姑娘 大钢铁公司
模式	A N N

实例	打伤了猎人的眼睛 打伤了眼睛的猎人
模式	V 了 N 的 N

实例	董永和七仙女的母亲 董永和七仙女的孩子
模式	N 和 N 的 N



2 自顶向下分析法

- 自顶向下的方法又称为基于**预测**的方法。

这种方法是先产生对后面将要出现的成分的预期，然后再通过逐步吃进待分析的字符串来验证预期。

- 如果预期得到了证明，就说明待分析的字符串可以被分析为所预期的句法结构。
- 如果某一个环节上预期出了差错，那就要用另外的预期来替换（即回溯）。
- 如果所有环节上所有可能的预期都被吃进的待分析字符串所“反驳”，那就说明待分析的字符串不可能是一个合法的句子，分析失败。



分析示例

张三是县长派来的
苍蝇是瞎子打死的
主意是董永想出来的

.....

N V N V V 的

词典:

张三: N

县长: N

是: V

派: V

来: V

的: de

规则:

(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) $V' \rightarrow V V$

自顶向下分析法示例-1

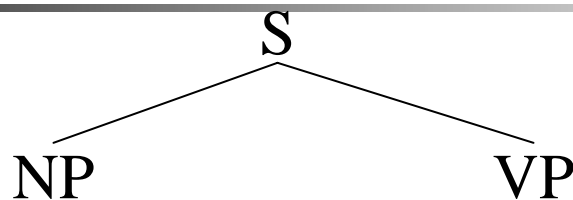
- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

查词典

N V N V V de
| | | | | |
张三 是 县长 派 来 的

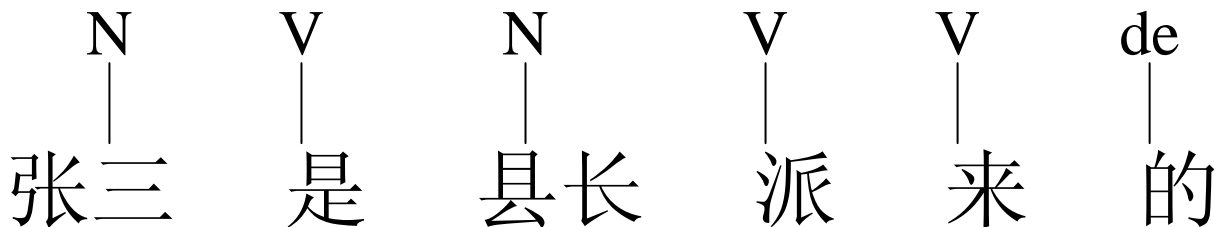
自顶向下分析法示例-2

- (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) $V' \rightarrow V V$



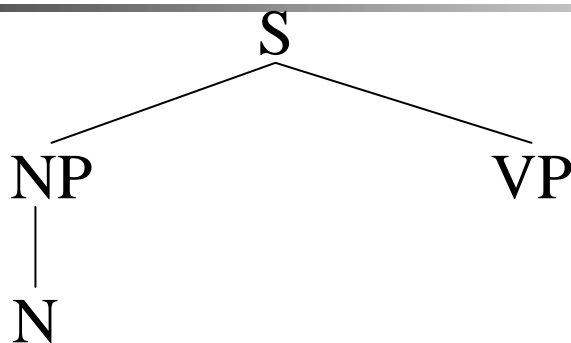
使用规则:

$S \rightarrow NP VP$



自顶向下分析法示例-3

- (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) $V' \rightarrow V V$



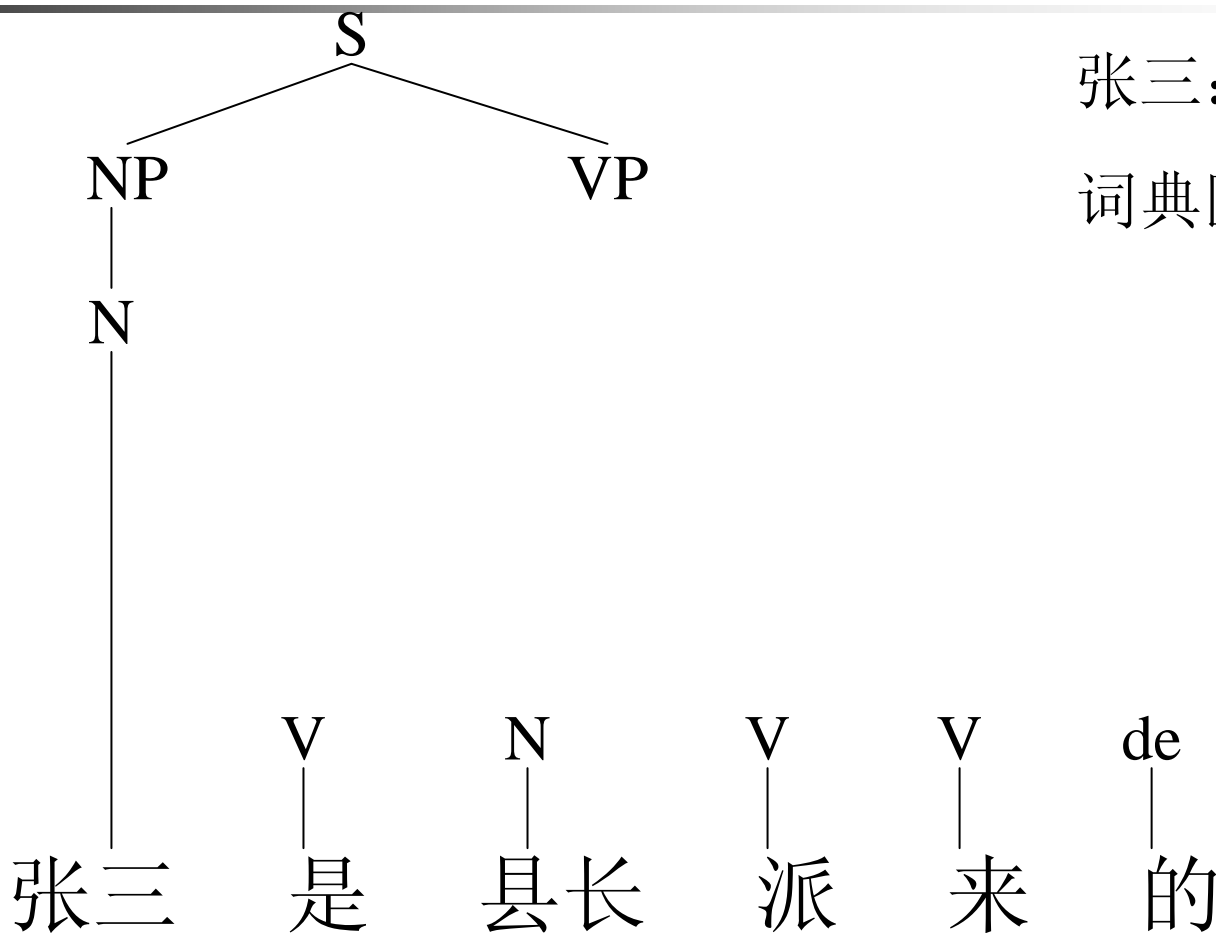
使用规则:

$NP \rightarrow N$



- (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) $V' \rightarrow V V$

自顶向下分析法示例-4

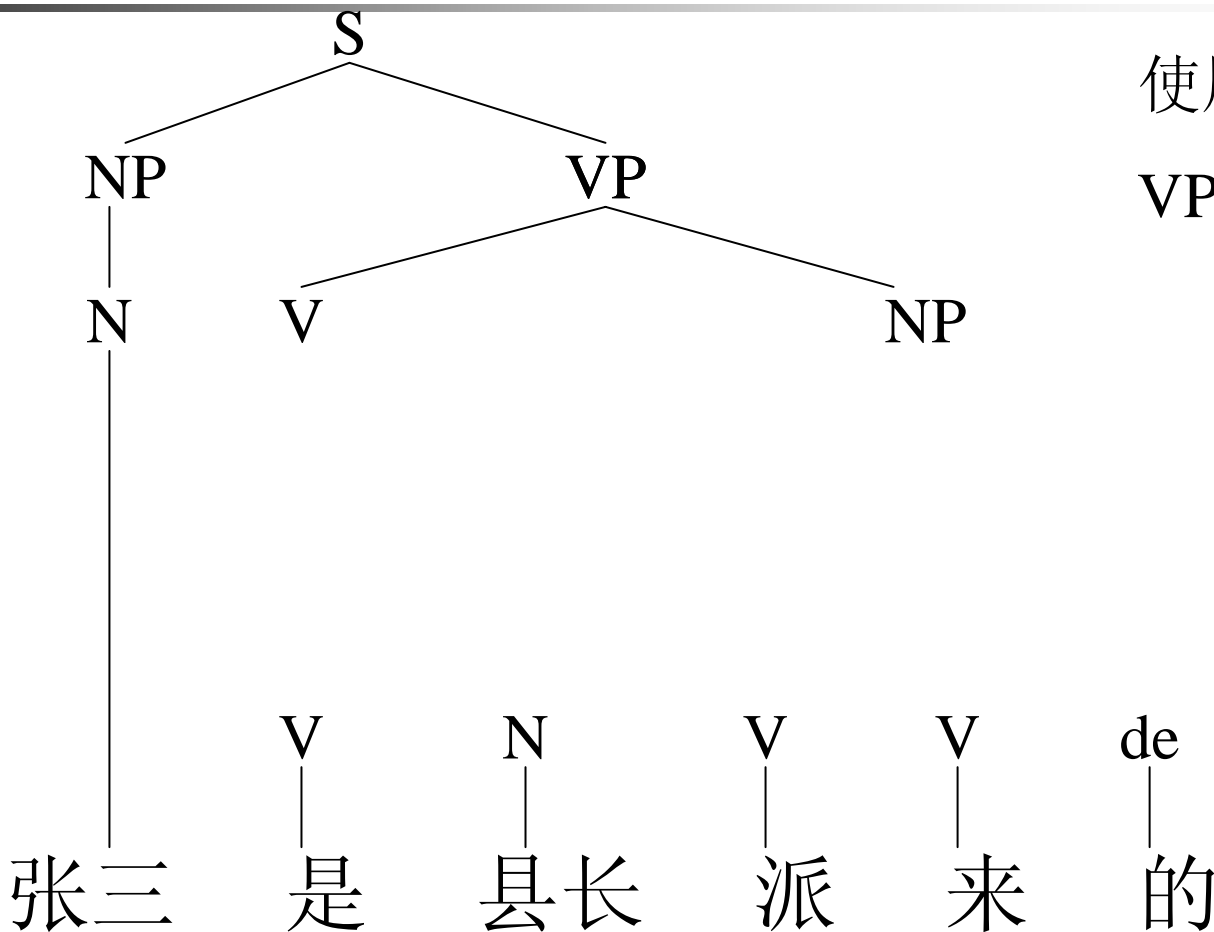


张三: N

词典匹配成功

- (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) $V' \rightarrow V V$

自顶向下分析法示例-5

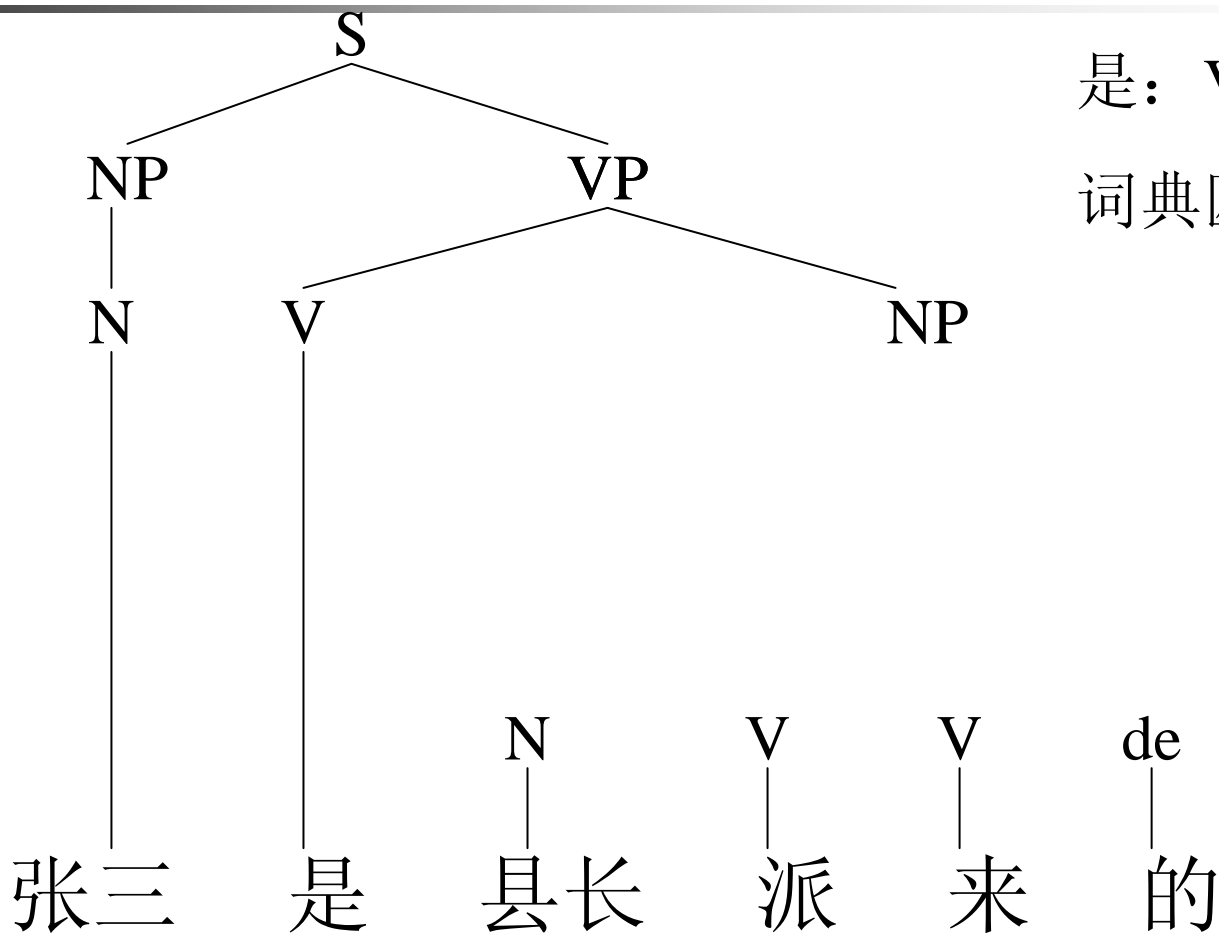


使用规则:

$VP \rightarrow V NP$

- (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) $V' \rightarrow V V$

自顶向下分析法示例-6

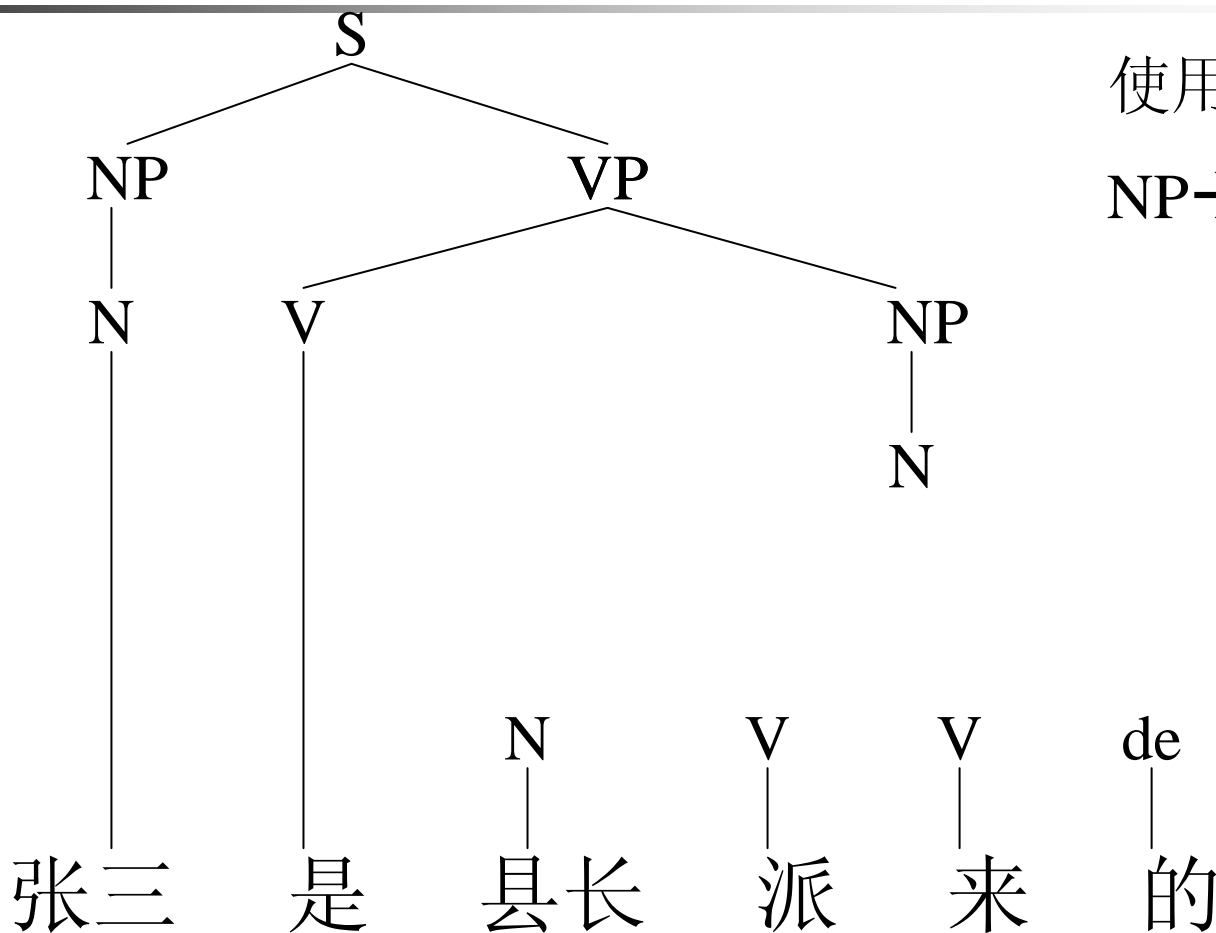


是: V

词典匹配成功

- (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) $V' \rightarrow V V$

自顶向下分析法示例-7

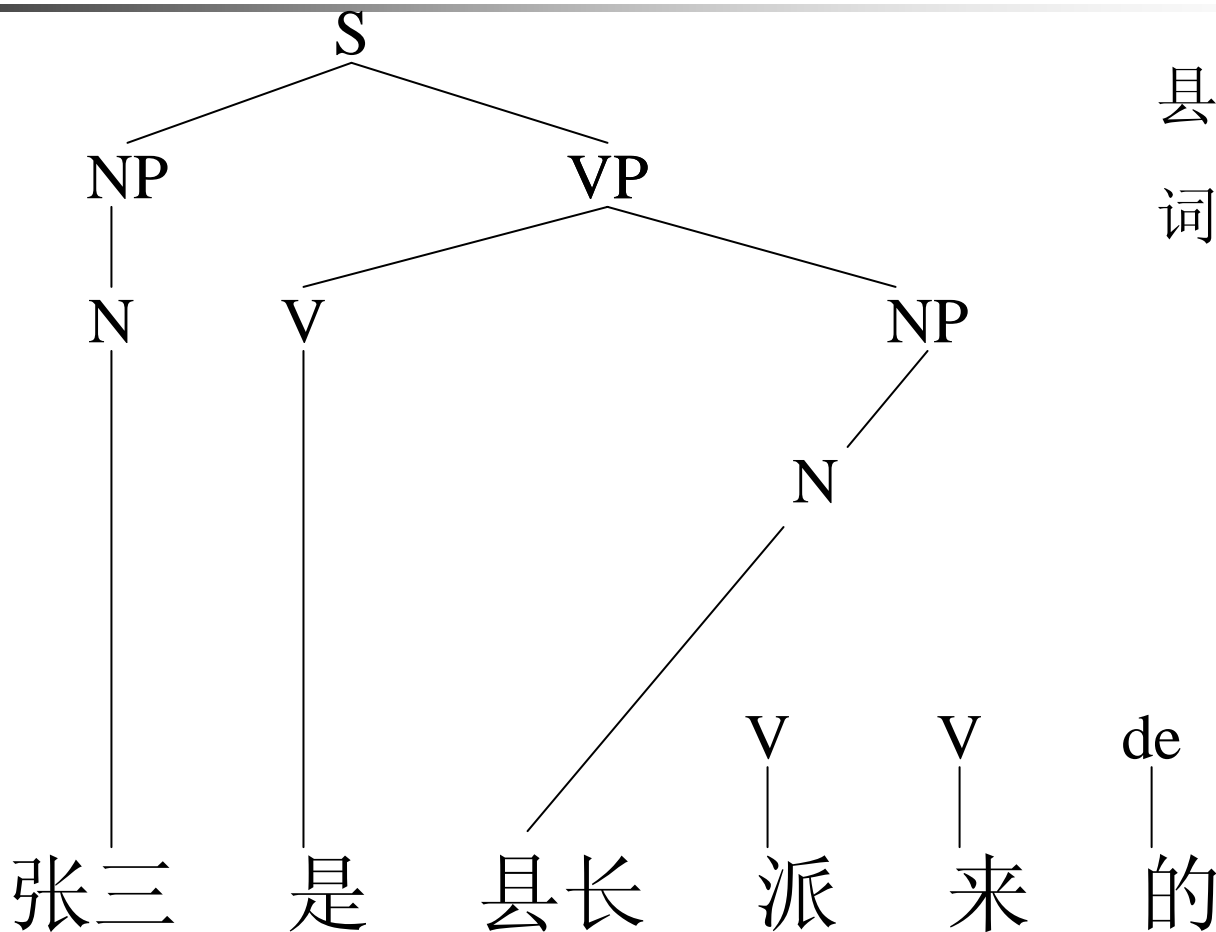


使用规则:

$NP \rightarrow N$

- (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) $V' \rightarrow V V$

自顶向下分析法示例-8

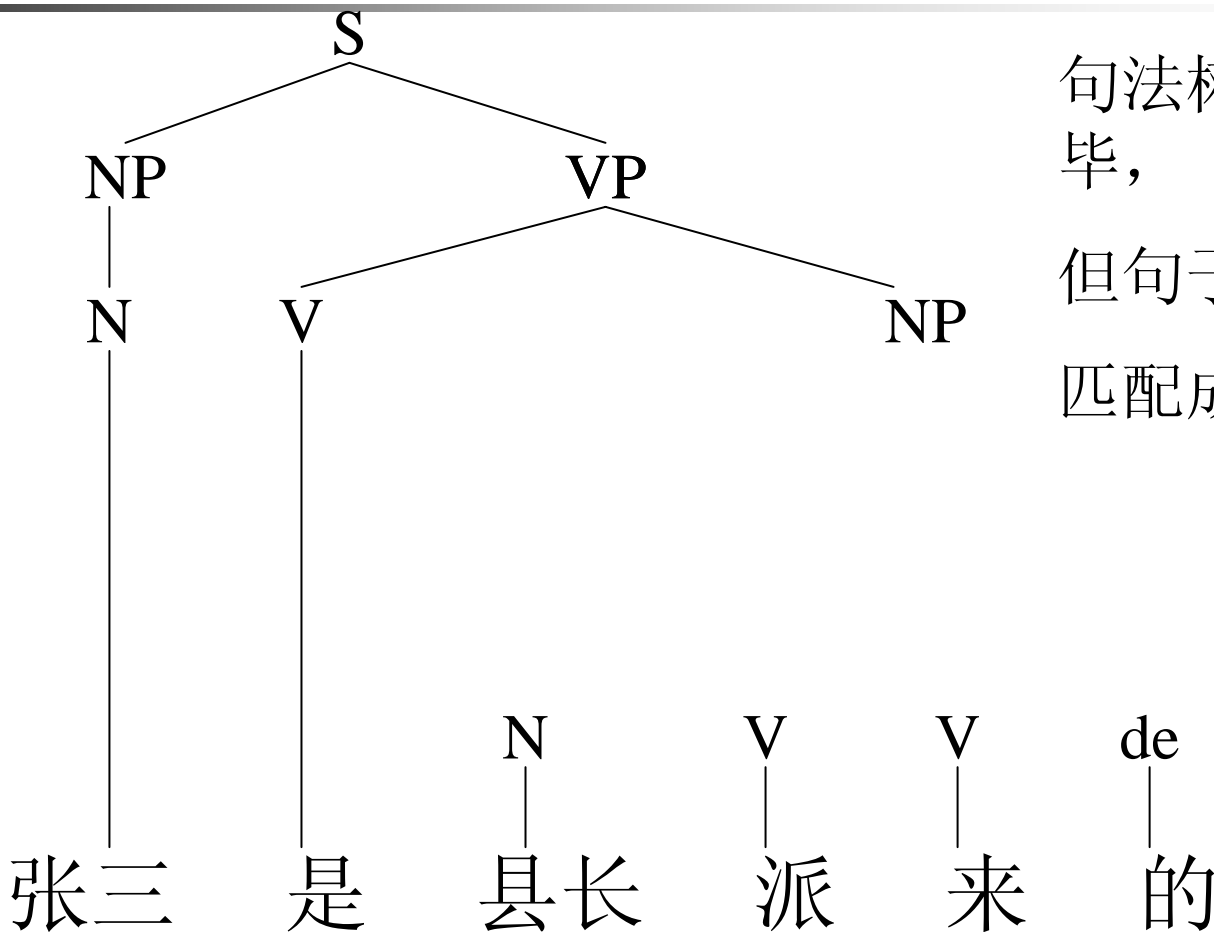


县长: N

词典匹配成功

- (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) $V' \rightarrow V V$

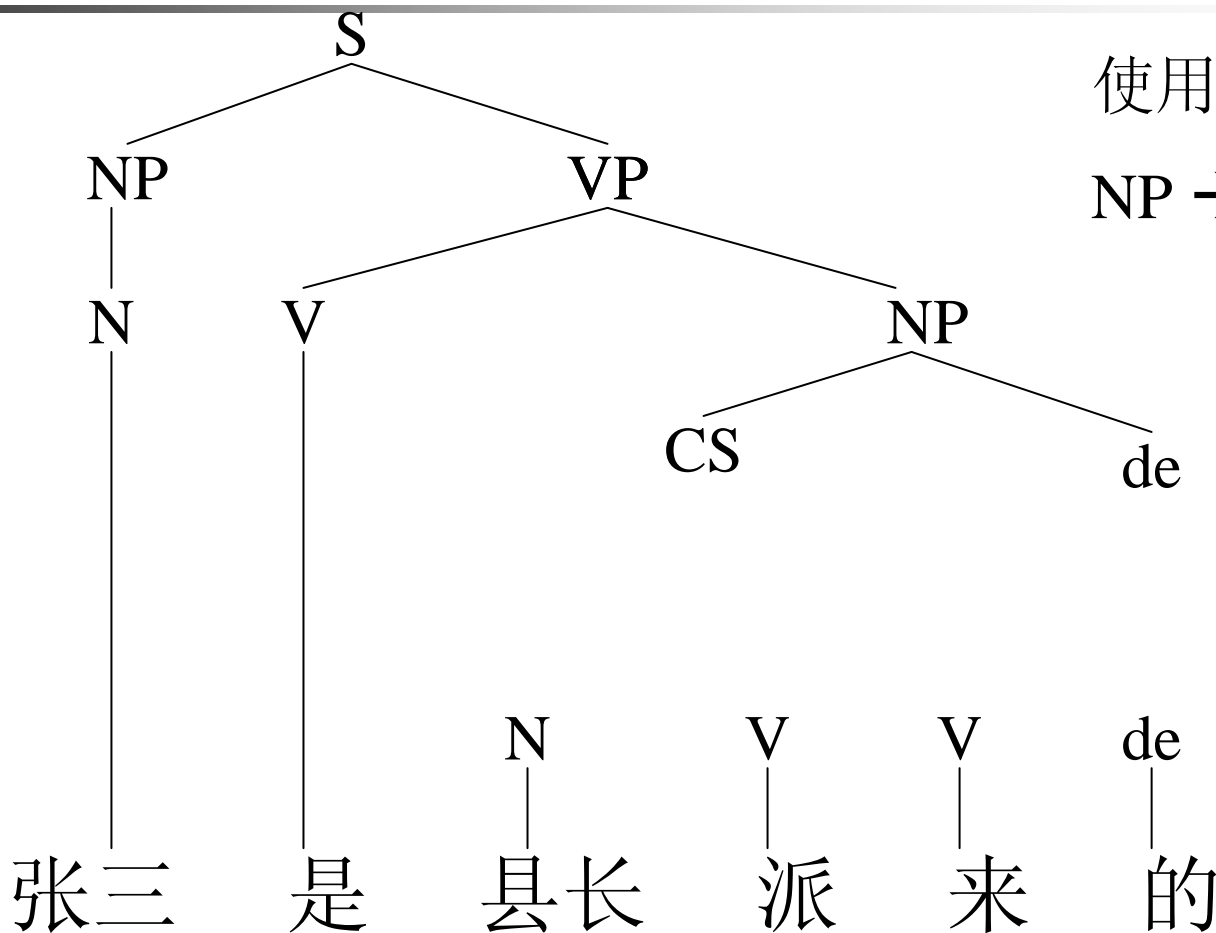
自顶向下分析法示例-9



句法树扩展完毕，
但句子没有完全匹配成功，回溯

- (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) $V' \rightarrow V V$

自顶向下分析法示例-10

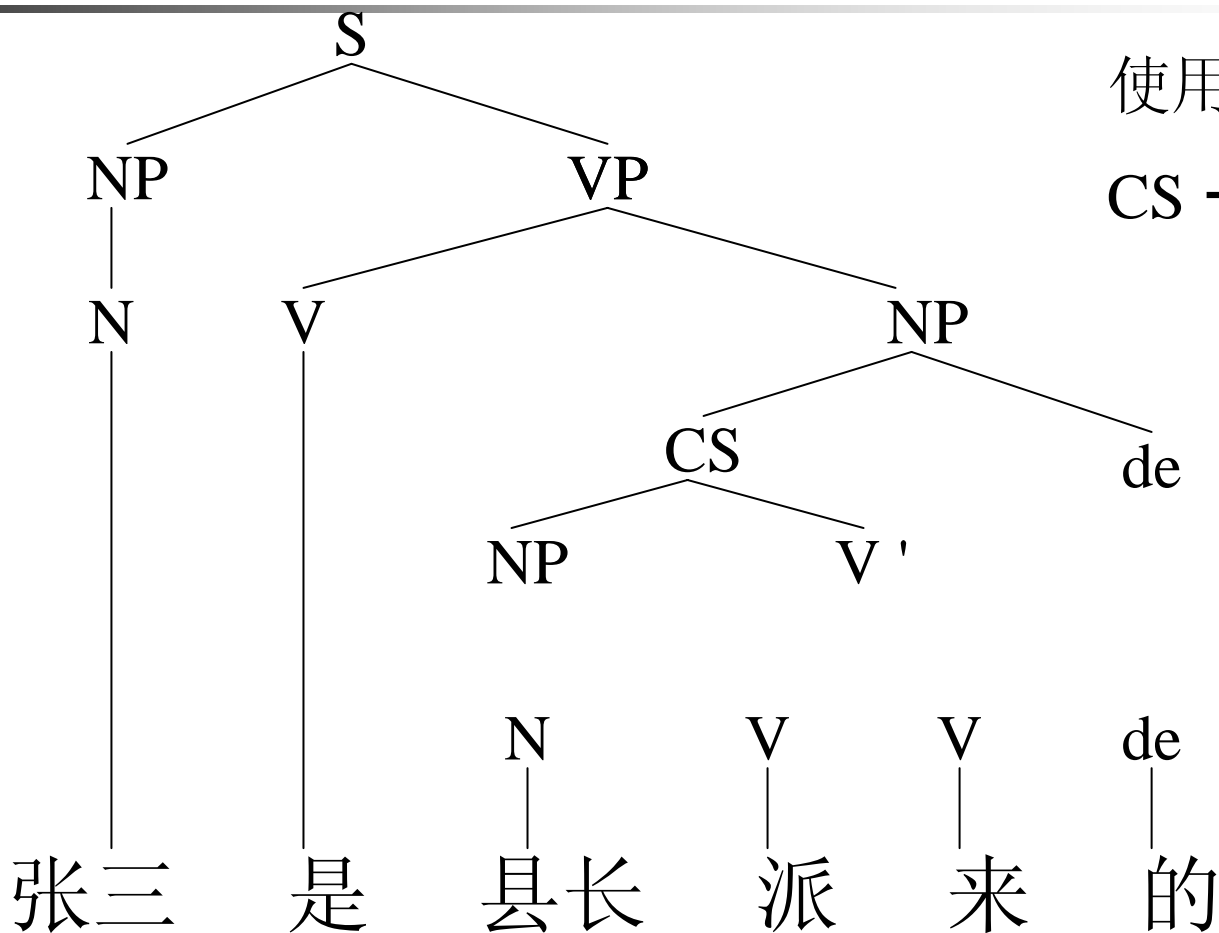


使用规则:

$NP \rightarrow CS$ de

- (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) $V' \rightarrow V V$

自顶向下分析法示例-11

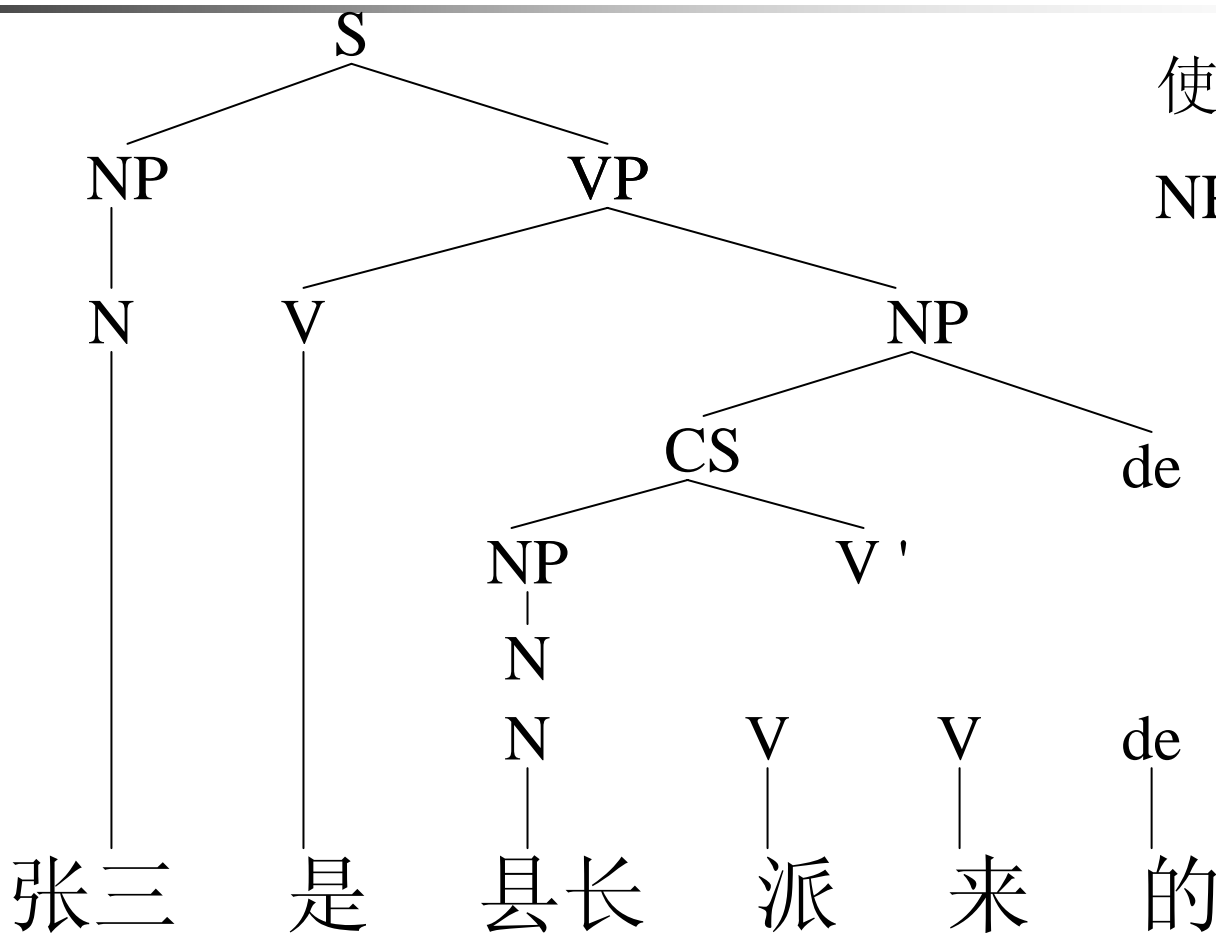


使用规则:

$CS \rightarrow NP V'$

- (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) $V' \rightarrow V V$

自顶向下分析法示例-12

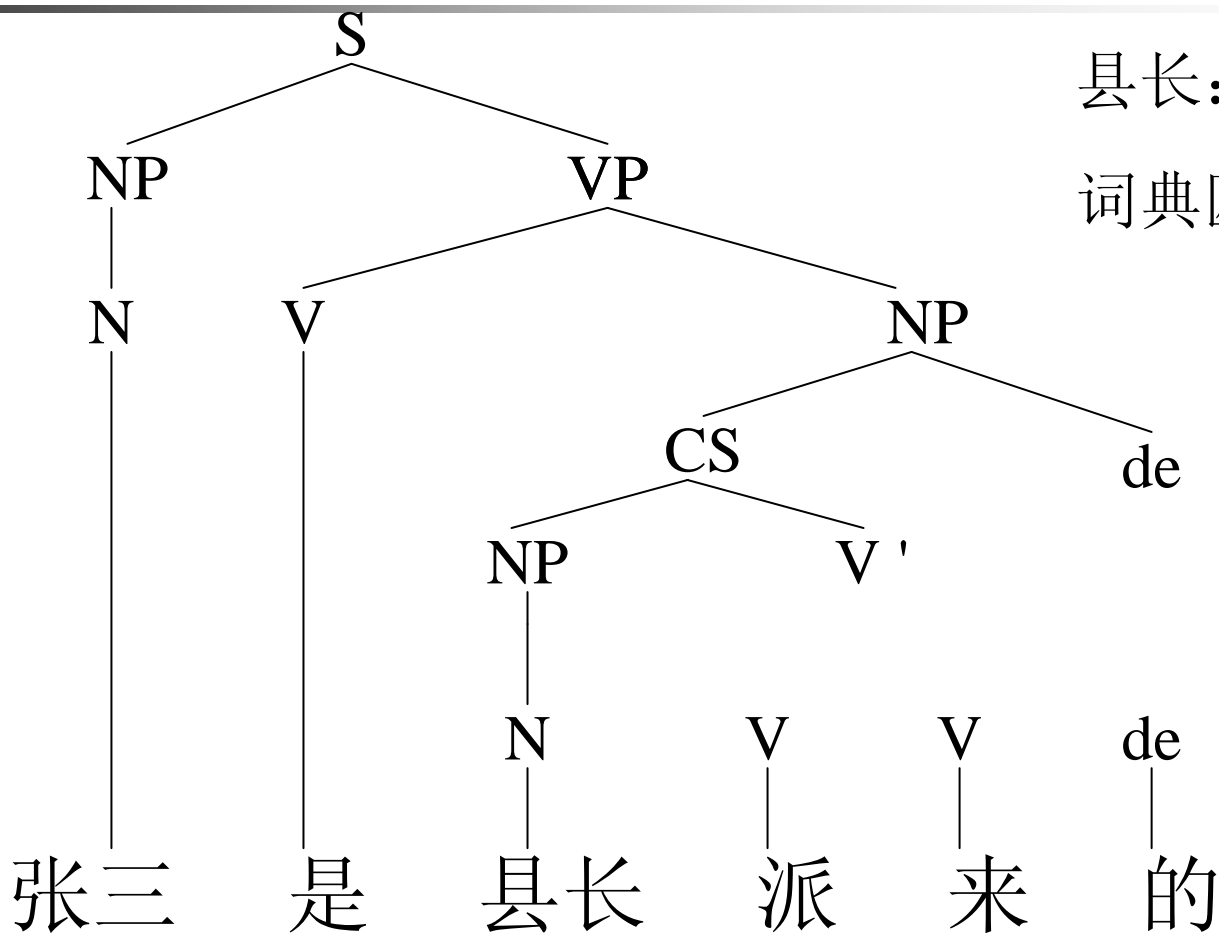


使用规则:

$NP \rightarrow N$

- (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) $V' \rightarrow V V$

自顶向下分析法示例-13

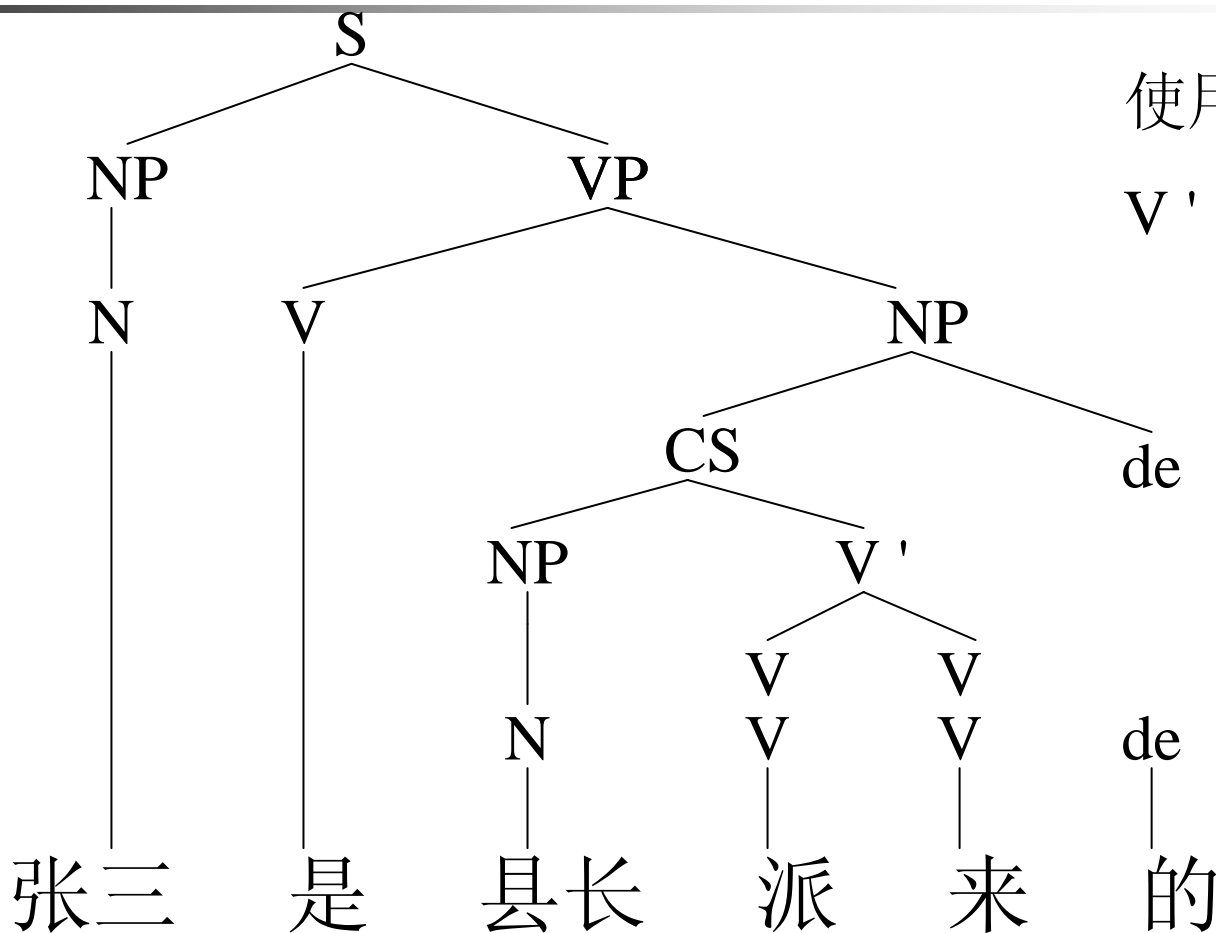


县长: N

词典匹配成功

- (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) $V' \rightarrow V V$

自顶向下分析法示例-14

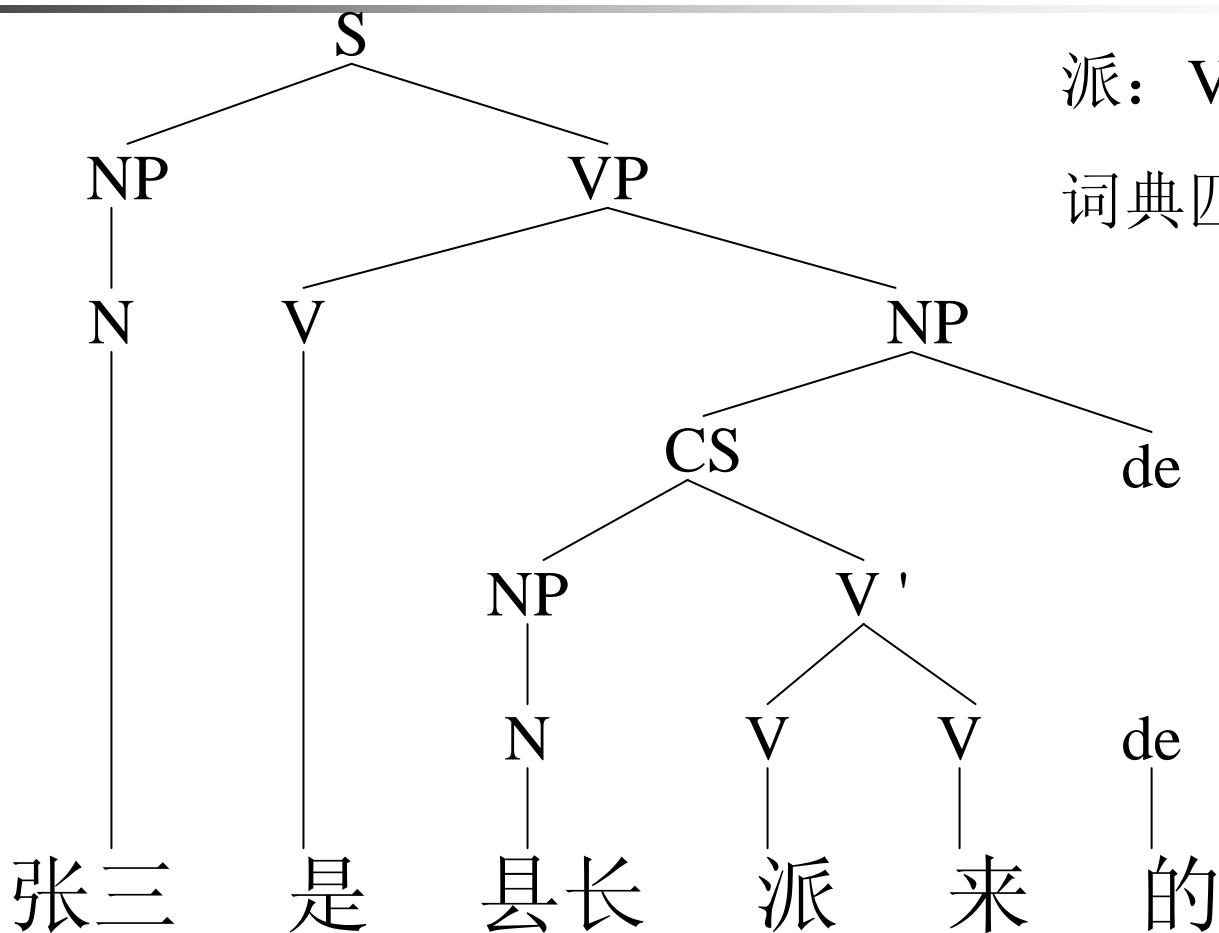


使用规则:

$V' \rightarrow V V$

- (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) $V' \rightarrow V V$

自顶向下分析法示例-15

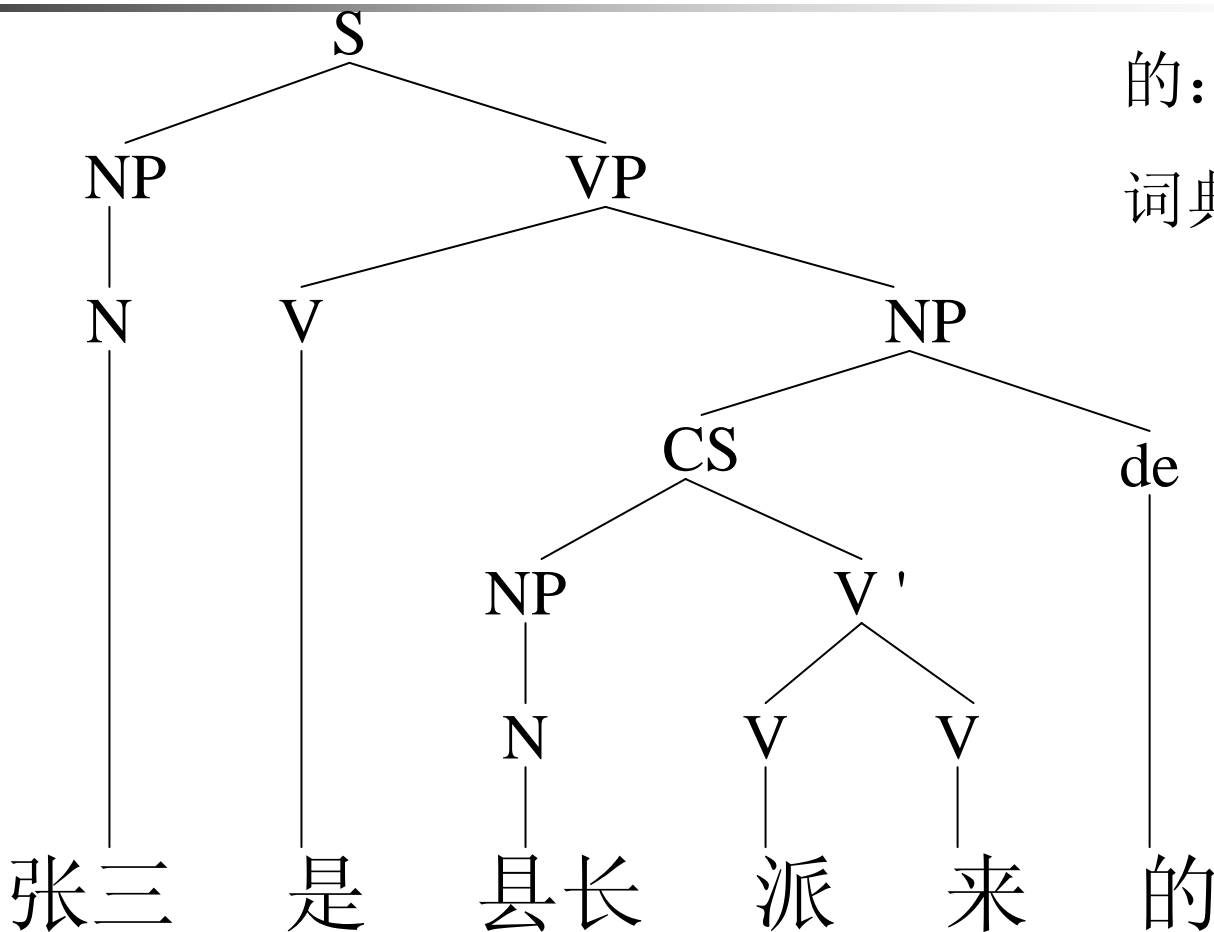


派: V 来: V

词典匹配成功

- (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) $V' \rightarrow V V$

自顶向下分析法示例-16



的: de

词典匹配成功

句法树扩展完毕

恰好句子匹配完成

分析成功



3 自底向上分析法

- 自底向上的方法也叫基于**归约**的方法。

这种方法是先逐步吃进待分析字符串，把它们从局部到整体层层归约为可能的成分。

- 如果整个待分析字符串被归约为开始符号 S ，那么分析成功。
- 如果在某个局部证明不可能有任何从这里把整个待分析字符串归约为句子的方案，那么就需要回溯。
- 如果经过回溯始终无法将待分析字符串归约为 S ，那么分析失败。

- (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) $V' \rightarrow V V$

自底向上分析法示例-1

查词典

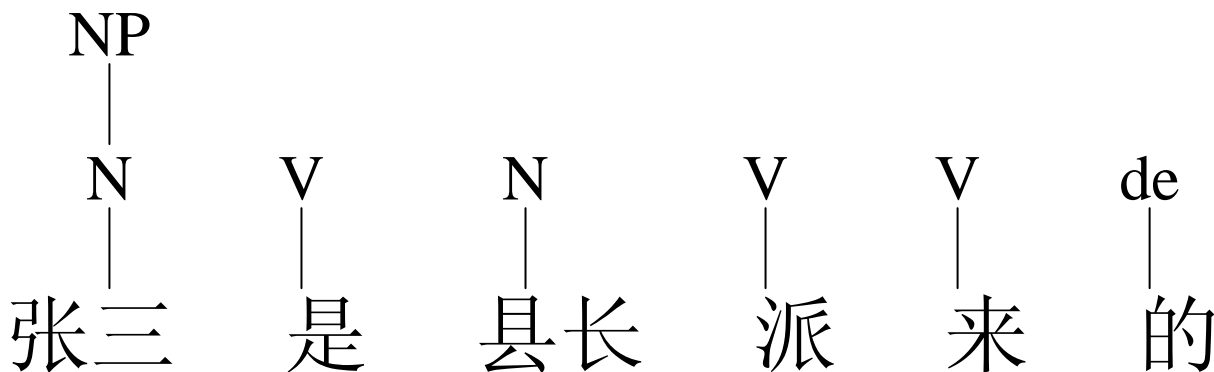
N V N V V de
| | | | | |
张三 是 县长 派 来 的

自底向上分析法示例-2

- (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) $V' \rightarrow V V$

使用规则:

$NP \rightarrow N$

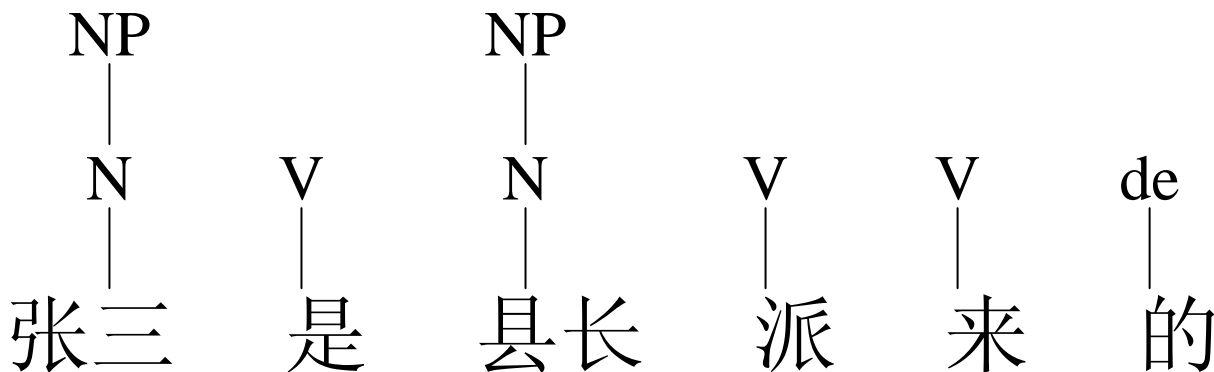


自底向上分析法示例-3

- (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) $V' \rightarrow V V$

使用规则:

$NP \rightarrow N$

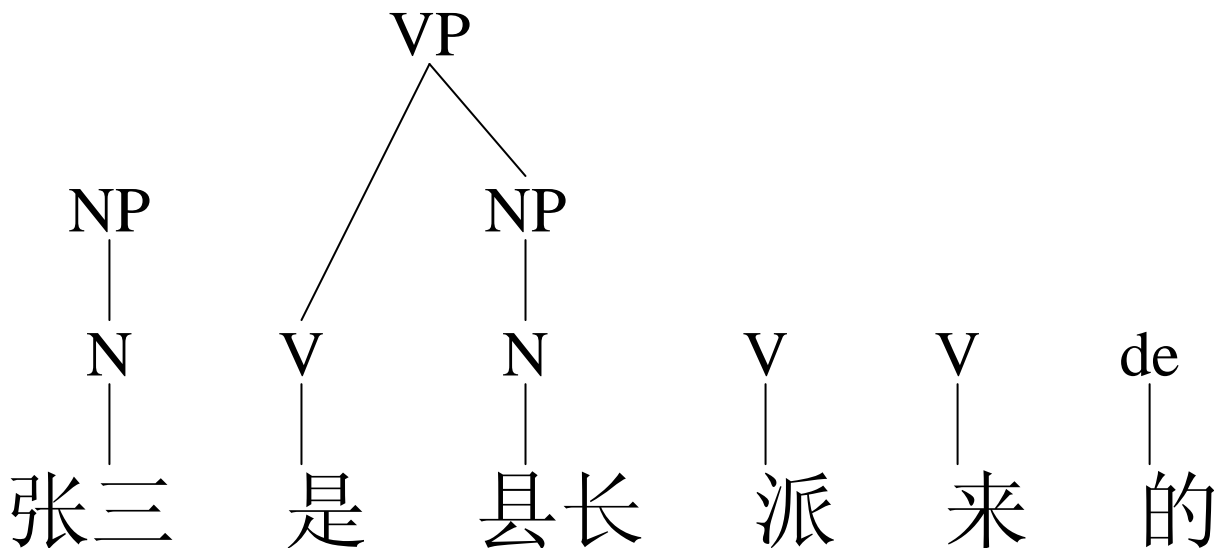


自底向上分析法示例-4

- (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) $V' \rightarrow V V$

使用规则:

$VP \rightarrow V NP$

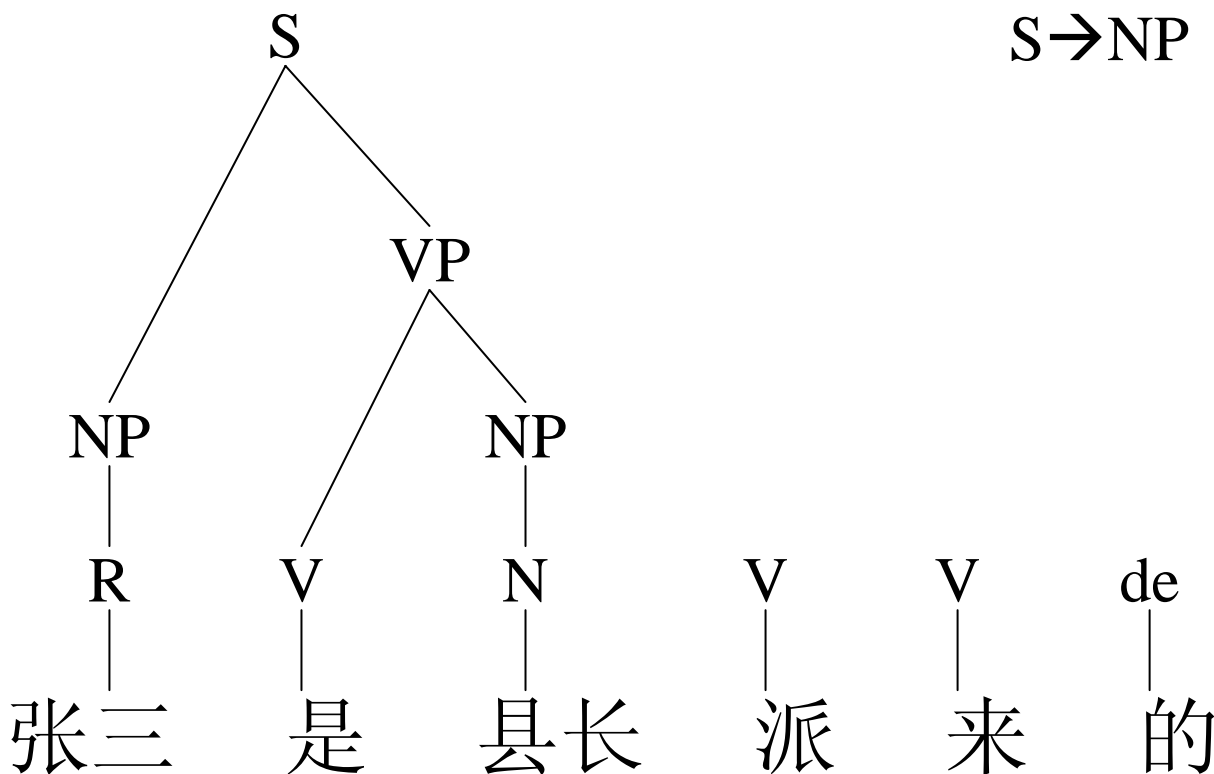


- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

自底向上分析法示例-5

使用规则:

$S \rightarrow NP VP$

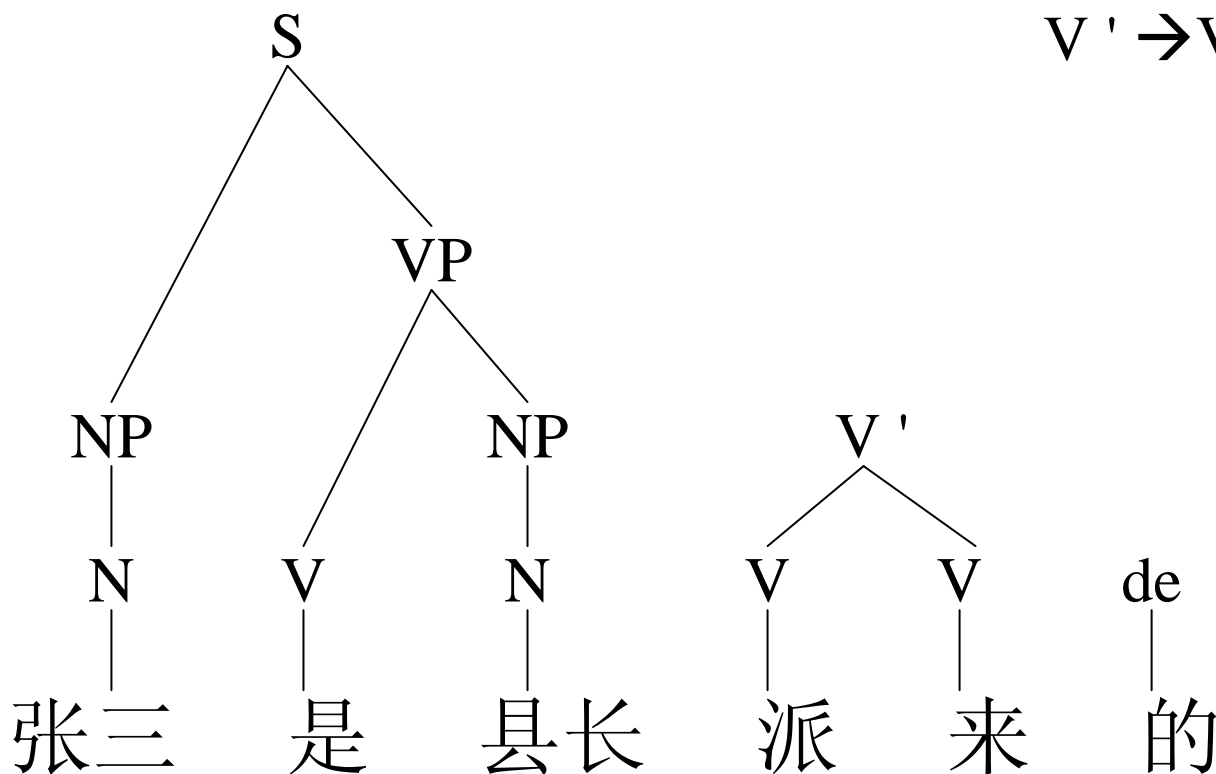


- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

自底向上分析法示例-6

使用规则:

$V' \rightarrow V V$

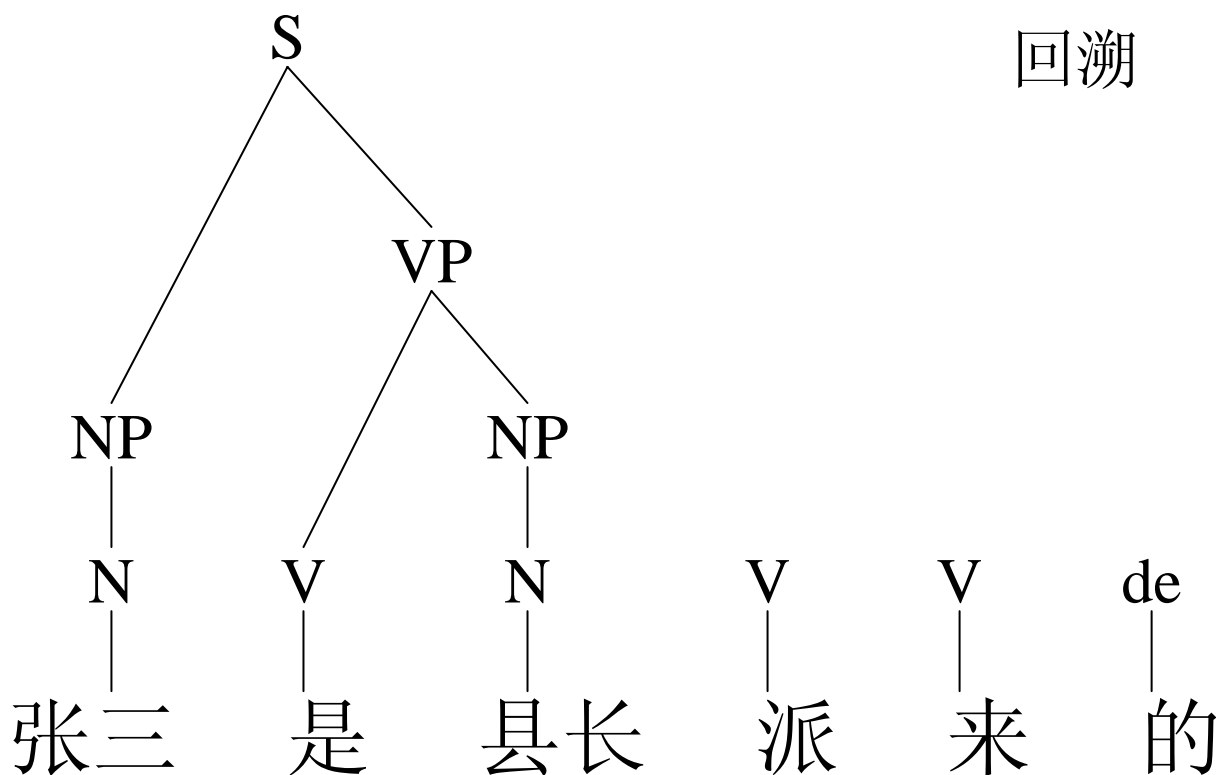


- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

自底向上分析法示例-7

无规则可用

回溯

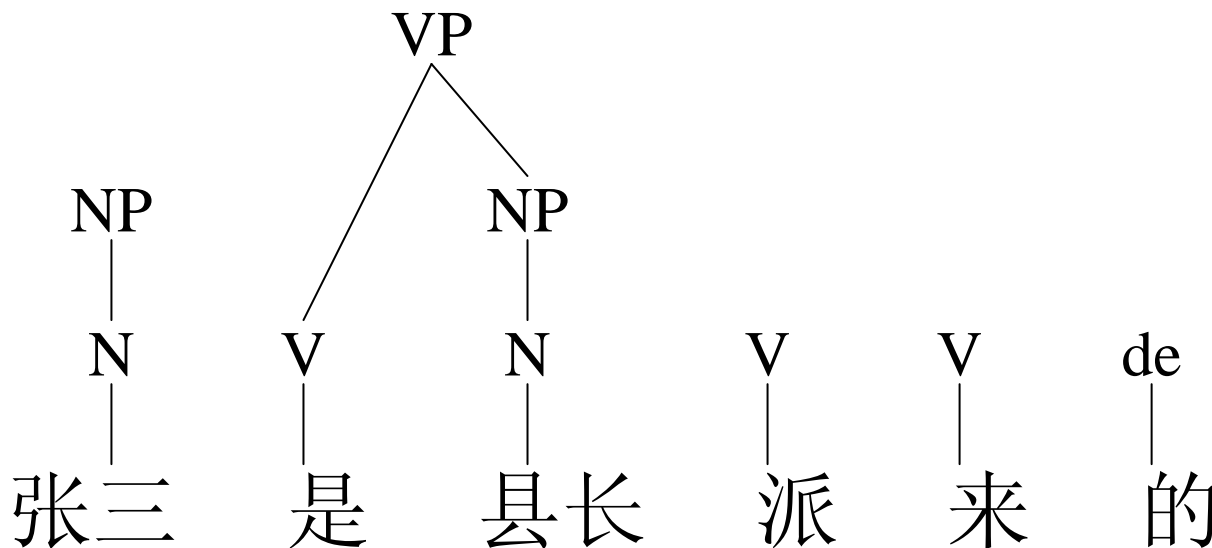


自底向上分析法示例-8

- (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) $V' \rightarrow V V$

无规则可用

回溯

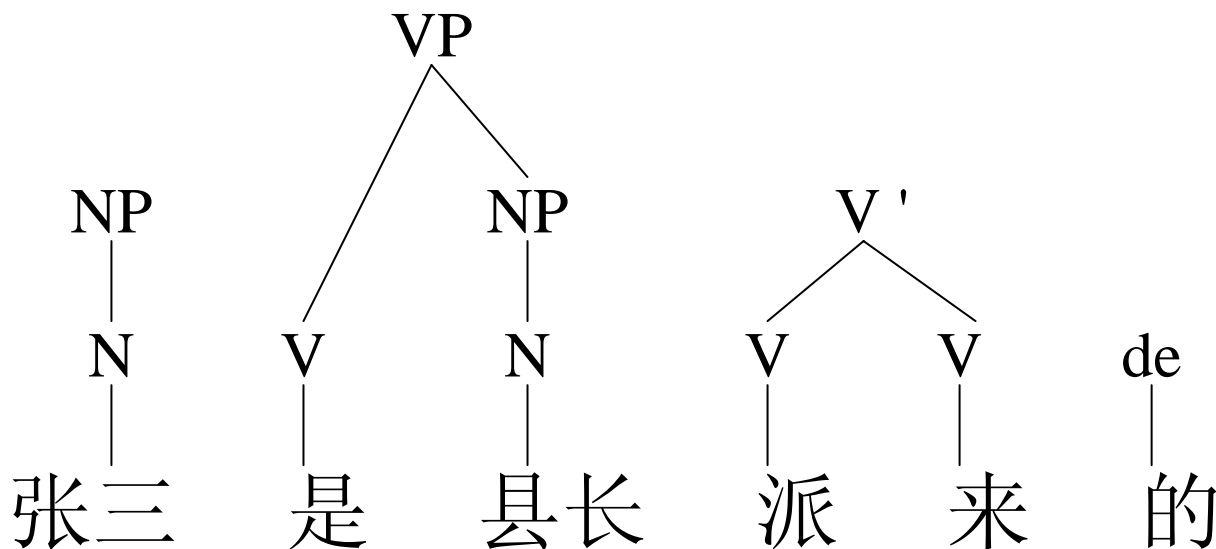


- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

自底向上分析法示例-9

使用规则:

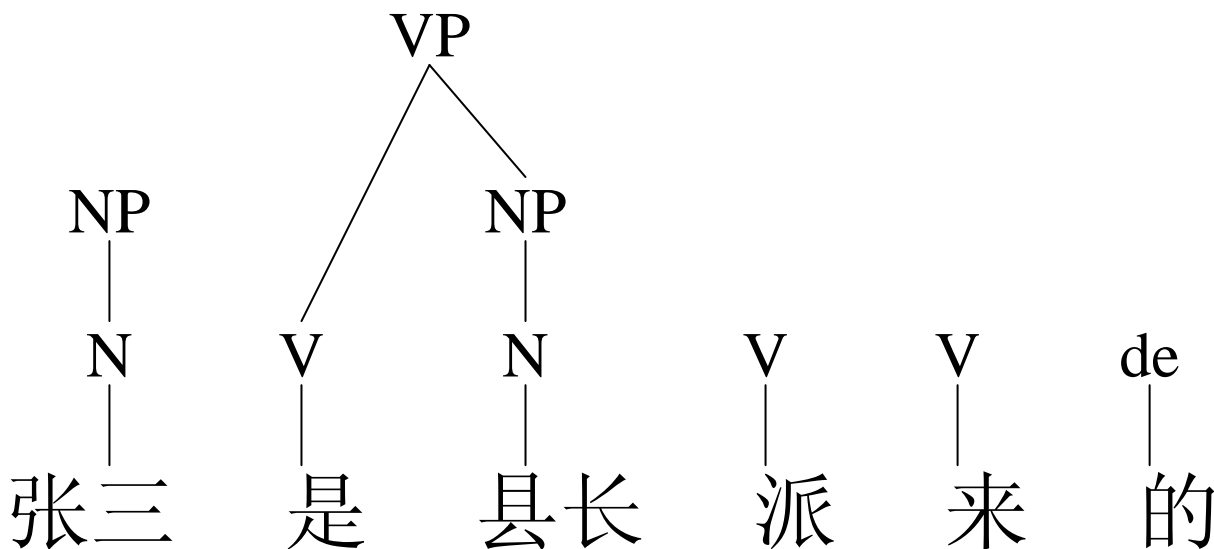
$V' \rightarrow V V$



- (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) $V' \rightarrow V V$

自底向上分析法示例-10

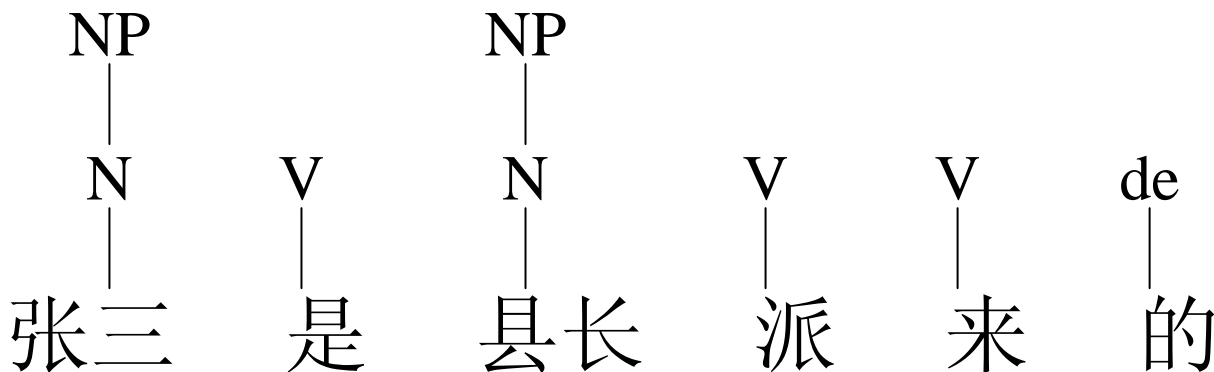
无规则可用，
回溯



自底向上分析法示例-11

- (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) $V' \rightarrow V V$

无规则可用，
回溯

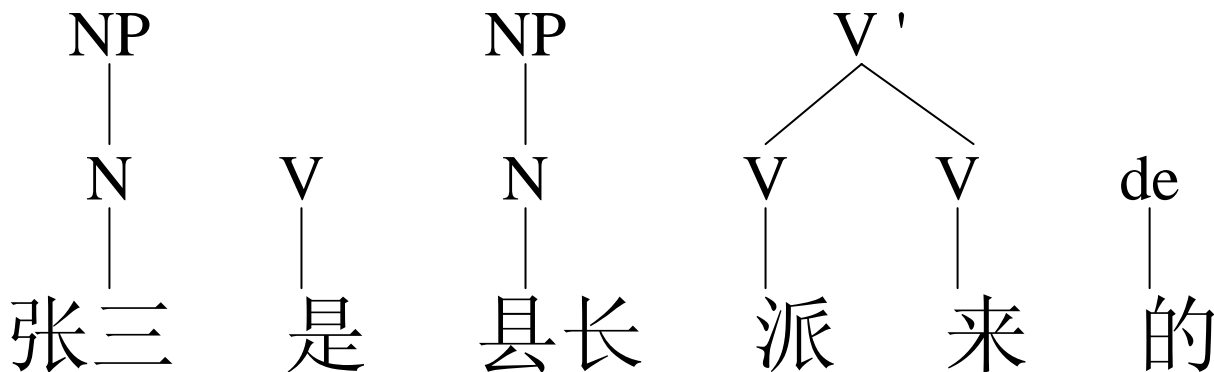


- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

自底向上分析法示例-12

使用规则:

$V' \rightarrow V V$

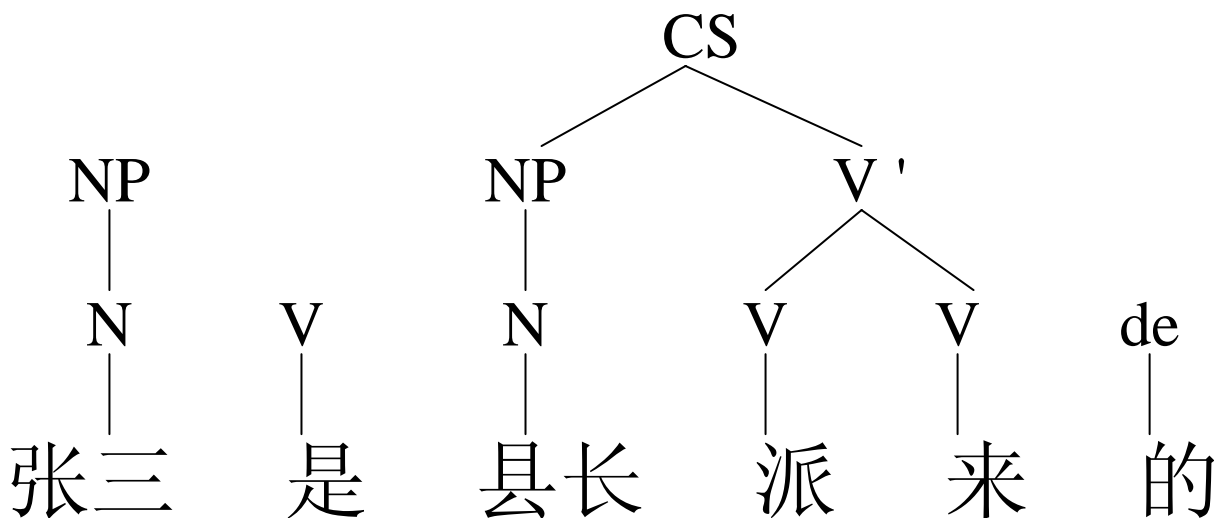


- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

自底向上分析法示例-13

使用规则:

$CS \rightarrow NP V'$

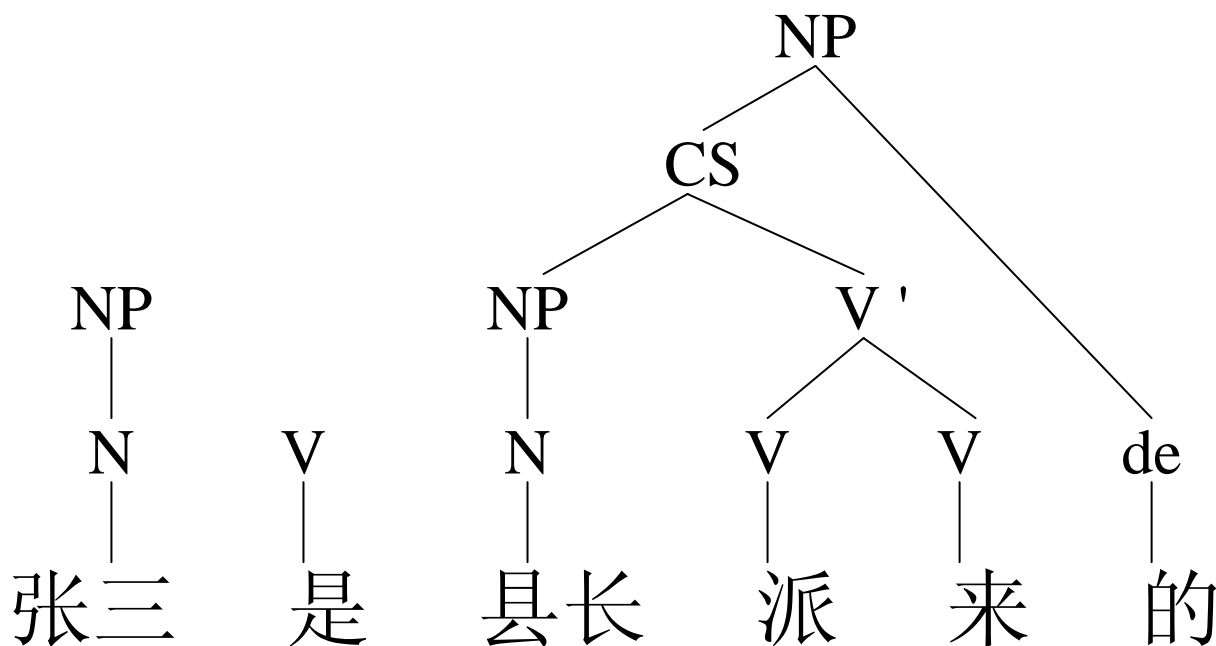


- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

自底向上分析法示例-14

使用规则:

$NP \rightarrow CS$ de

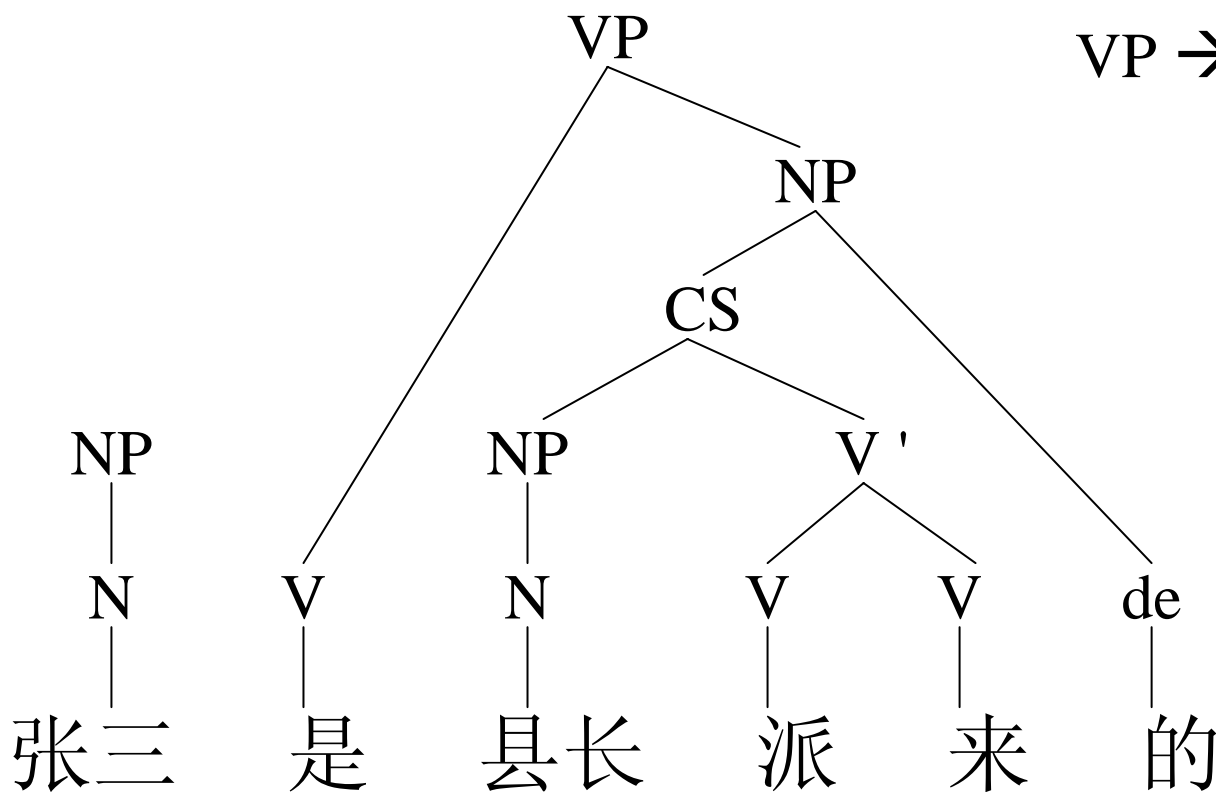


- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

自底向上分析法示例-15

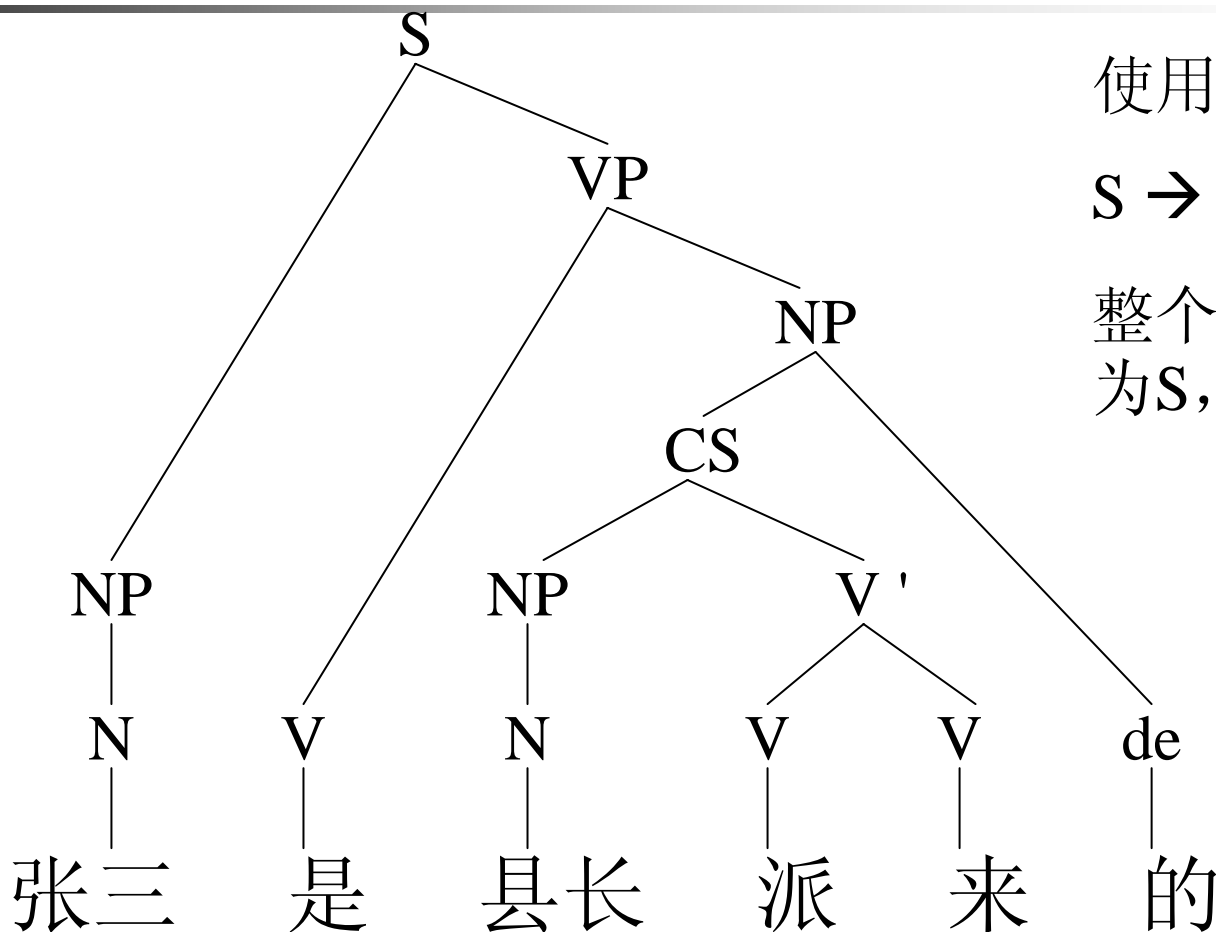
使用规则:

$VP \rightarrow V NP$



- (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) $V' \rightarrow V V$

自底向上分析法示例-16



使用规则:

$S \rightarrow NP VP$

整个句子归结为S，分析成功



自顶向下 vs. 自底向上

对于 $A \rightarrow B C$ 这样的CFG规则， A 为左部， $B C$ 为右部：

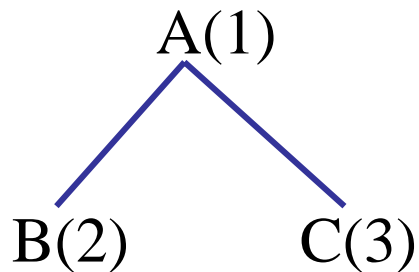
如果是根据规则右部（已匹配部分）来调用新规则，就是自底向上；

如果是根据规则左部（待匹配标记）来调用新规则，就是自顶向下。

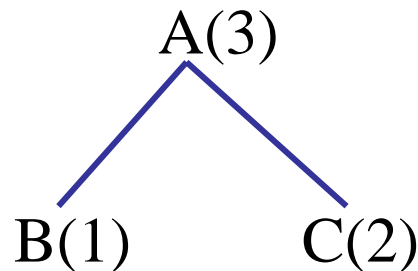
3 左角分析法—概述

- 左角分析法是一种自顶向下和自底向上相结合的方法
- 所谓“左角(Left Corner)”是指任何一个句法子树中左下角的那个符号
- 比较：

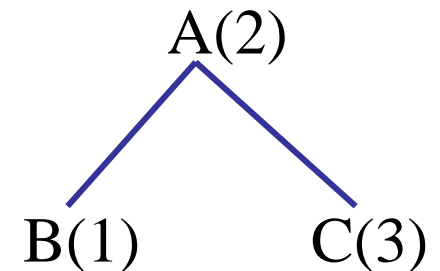
自顶向下分析法



自底向上分析法



左角分析法





左角分析法-基本思想

- 1 由箭头右部最左的非终结符去调用规则（自底向上，然后向下）
- 2 由终结符触发归约（即当碰到适合的终结符时，一条规则被满足）；
- 3 只有被满足的非终结符节点，才能进一步触发归约；
- 4 从历史上形成的离当前处理节点最近的最大的子树根节点开始回溯
- 5 回溯就是断开已经形成的子树的右分支（**right branch**），由右分支节点（非终结符或终结符）开始，重新调用规则进行预测（等待归约）

规则被满足指的是规则根节点覆盖了终结符序列中的一个子串

左角分析法示例-1

- (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) $V' \rightarrow V V$

查词典

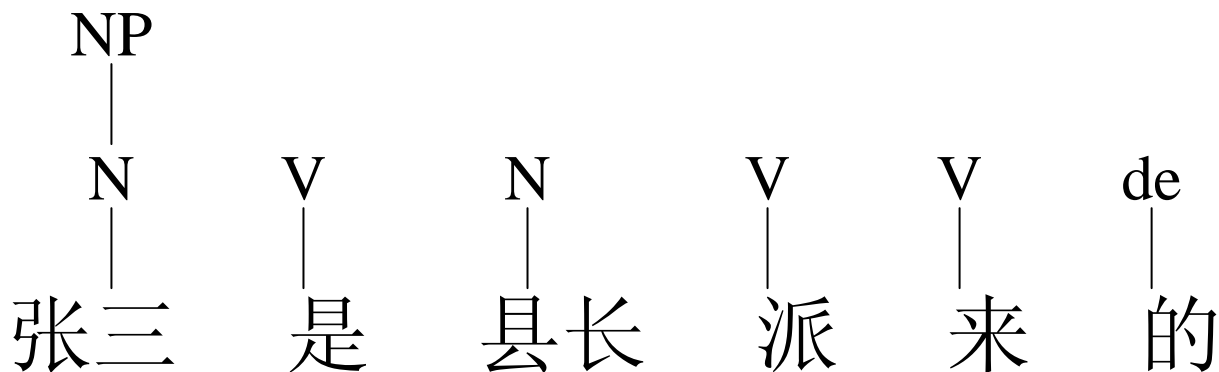
N V N V V de
| | | | | |
张三 是 县长 派 来 的

左角分析法示例-2

- (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) $V' \rightarrow V V$

使用规则:

$NP \rightarrow N$

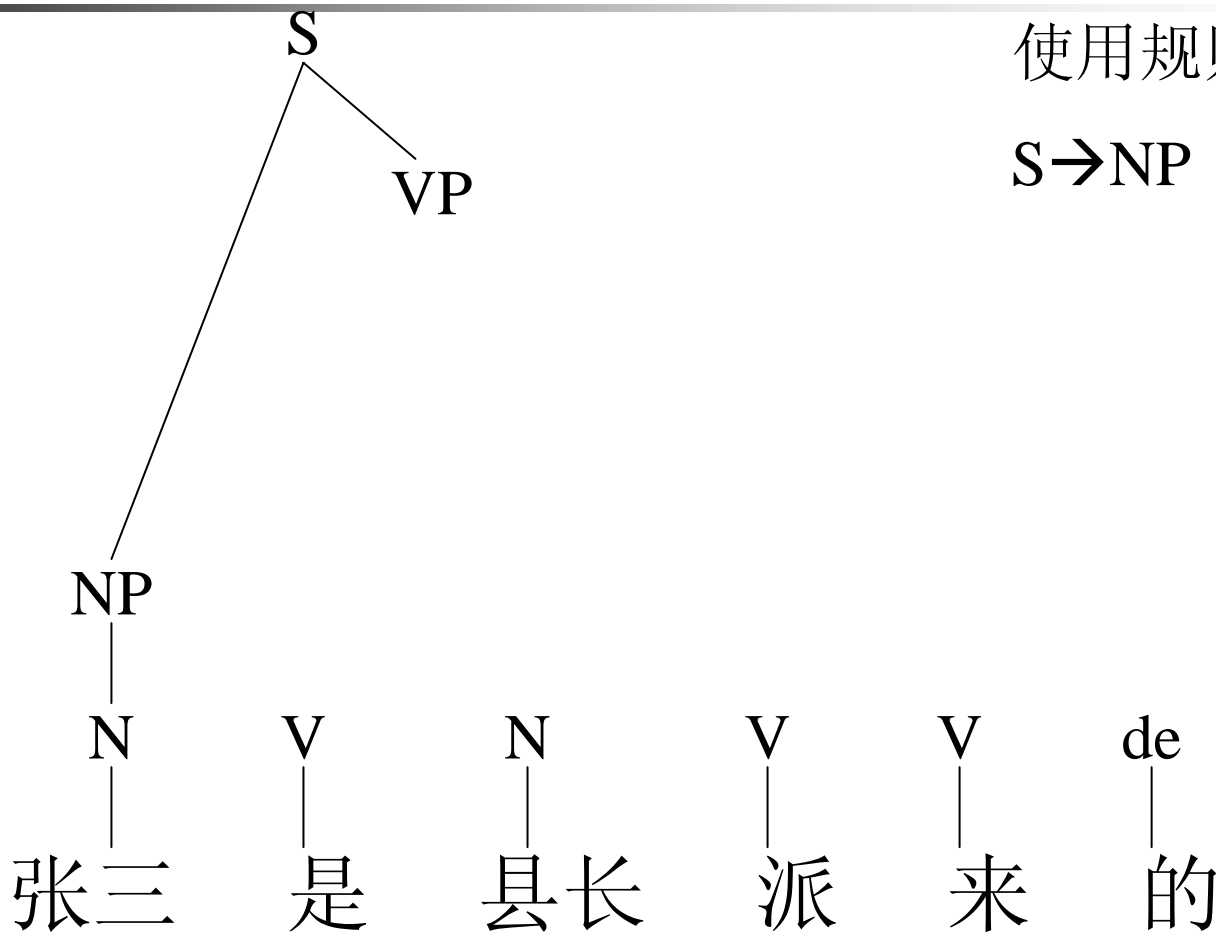


左角分析法示例-3

- (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) $V' \rightarrow V V$

使用规则:

$S \rightarrow NP VP$

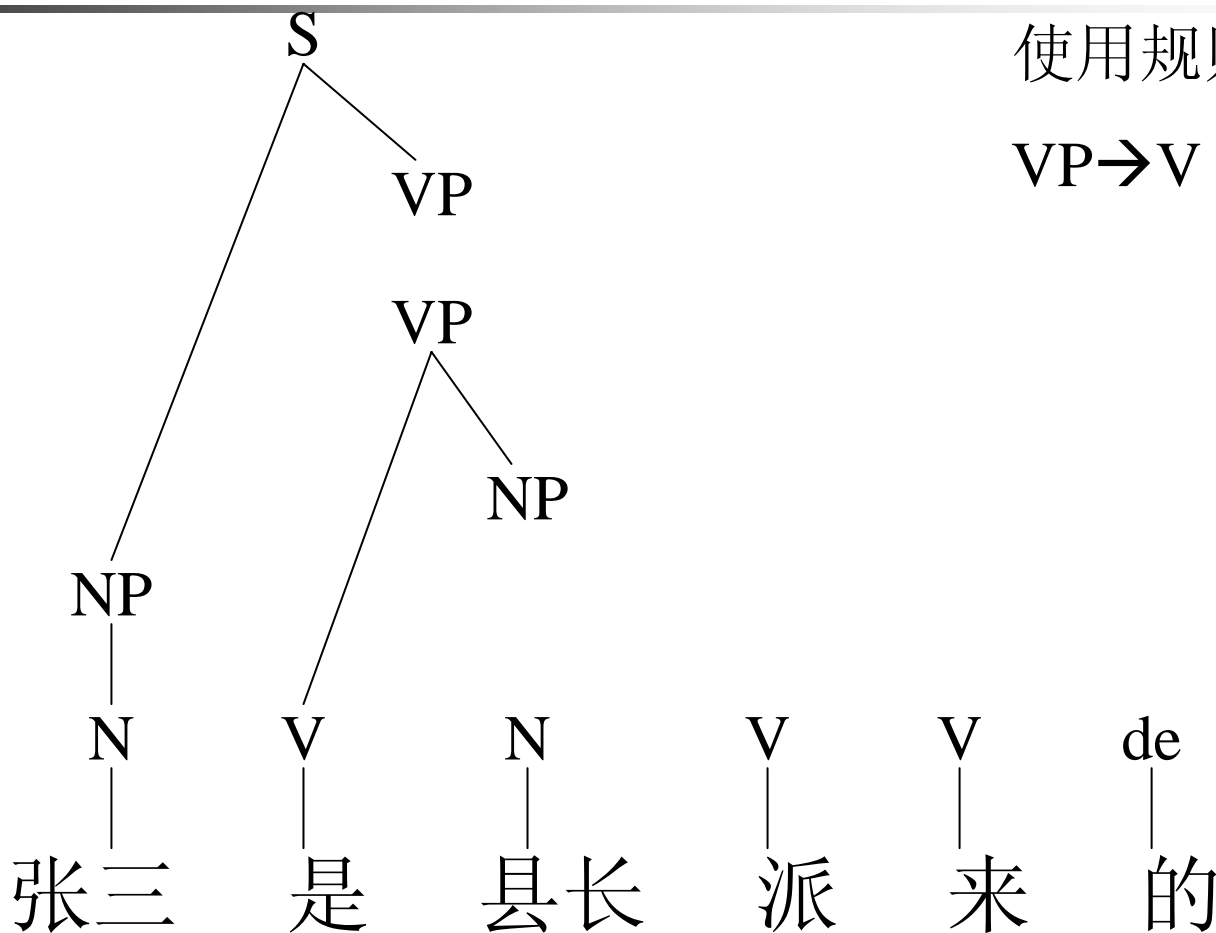


左角分析法示例-4

- (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) $V' \rightarrow V \ V$

使用规则:

$VP \rightarrow V \ NP$

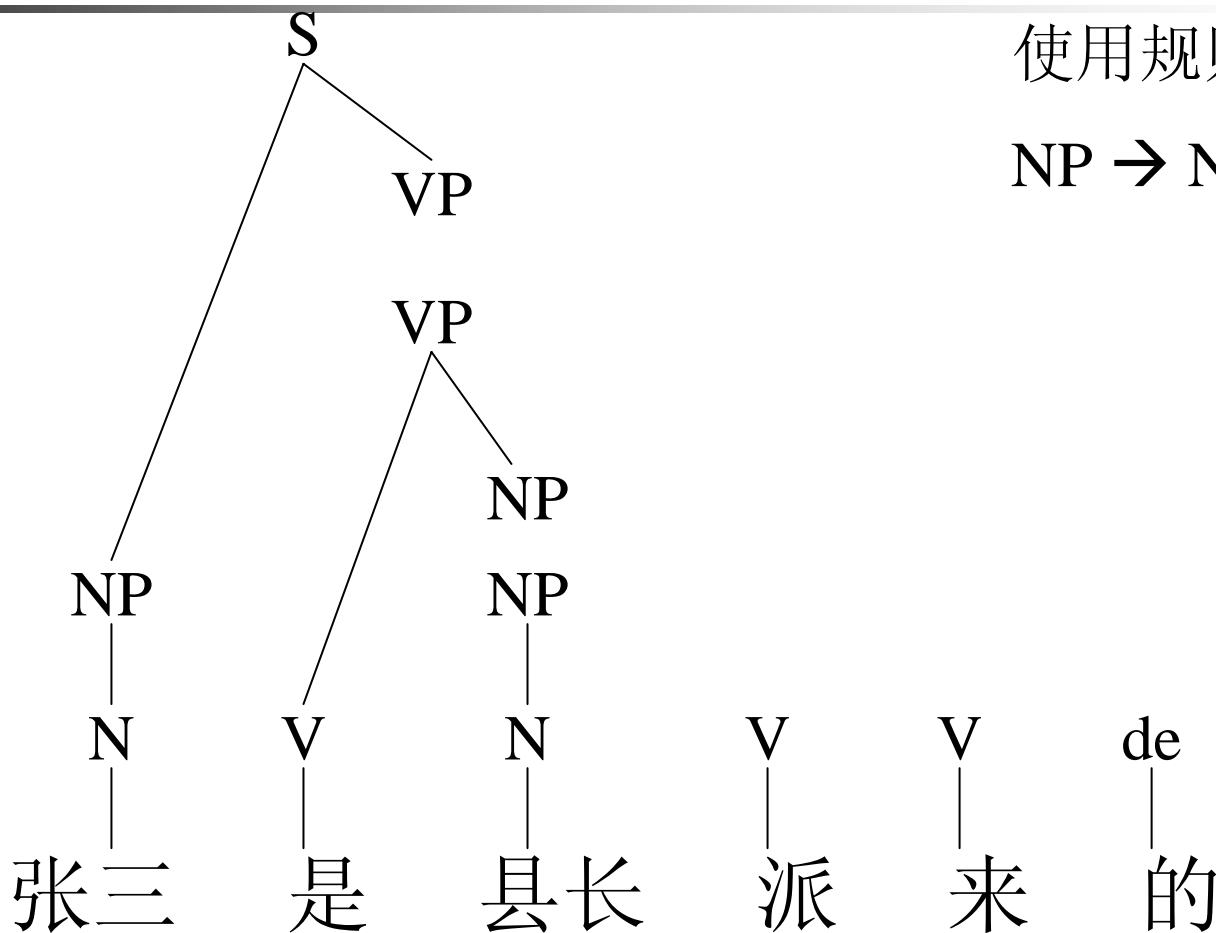


左角分析法示例-5

- (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) $V' \rightarrow V V$

使用规则:

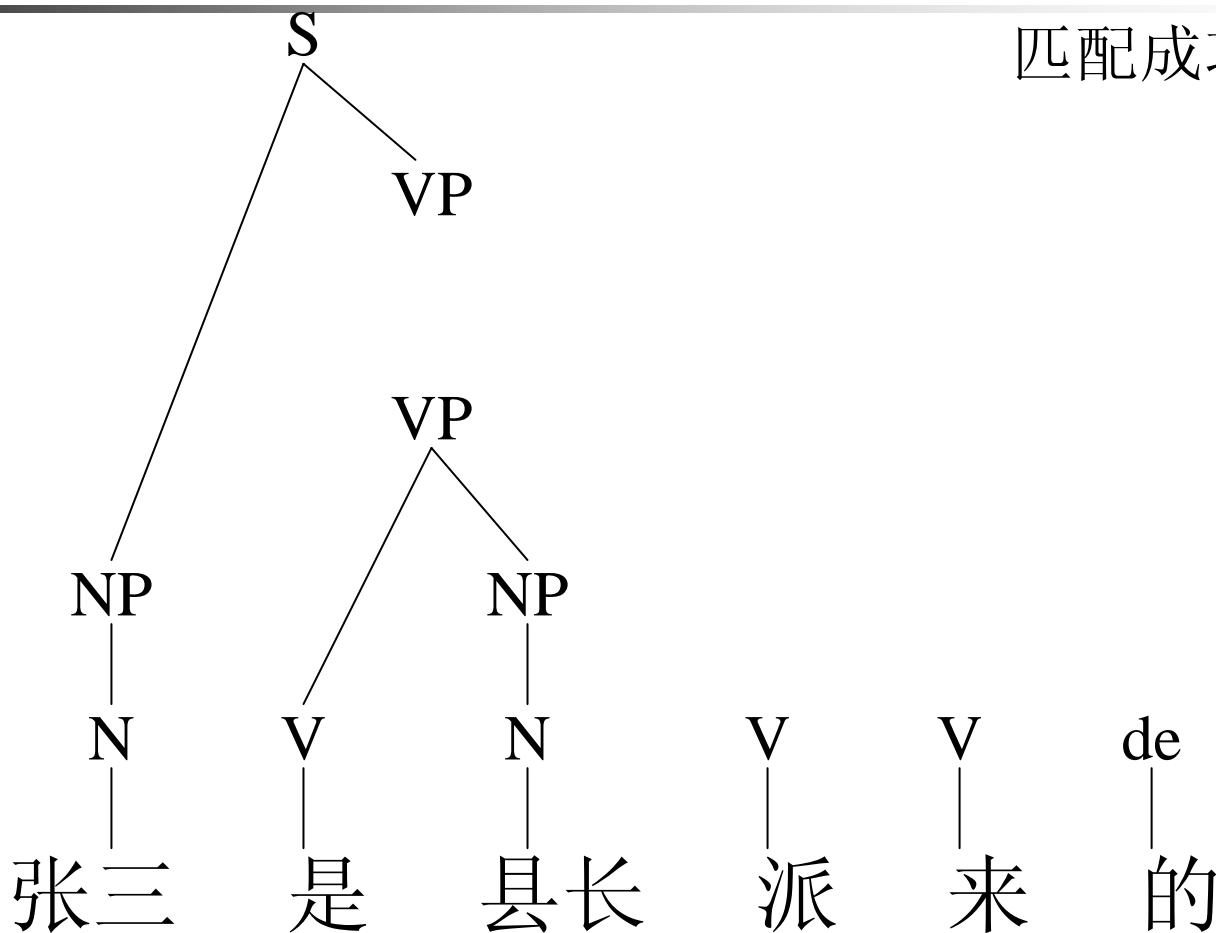
$NP \rightarrow N$



- (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) $V' \rightarrow V V$

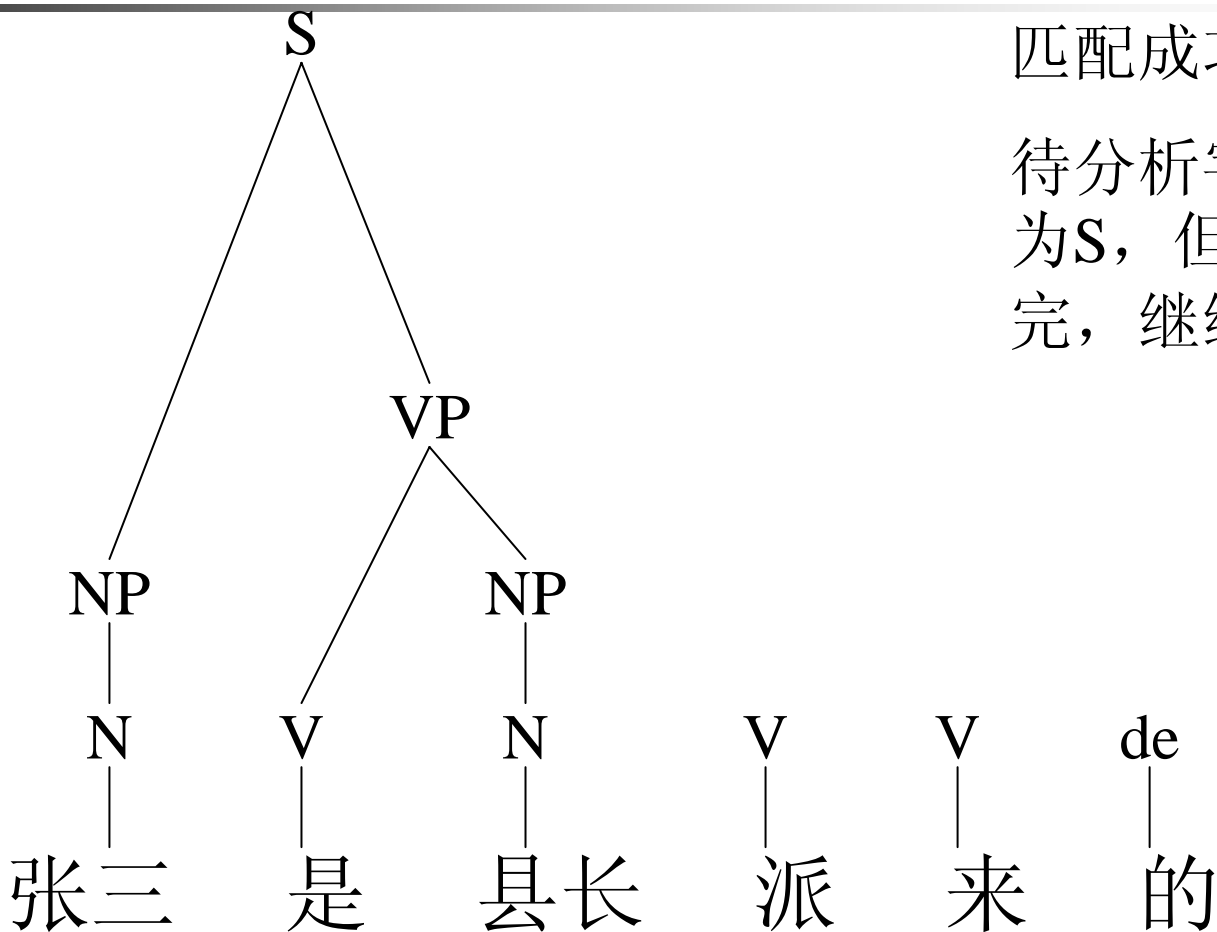
左角分析法示例-6

匹配成功



- (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) $V' \rightarrow V V$

左角分析法示例-7



匹配成功

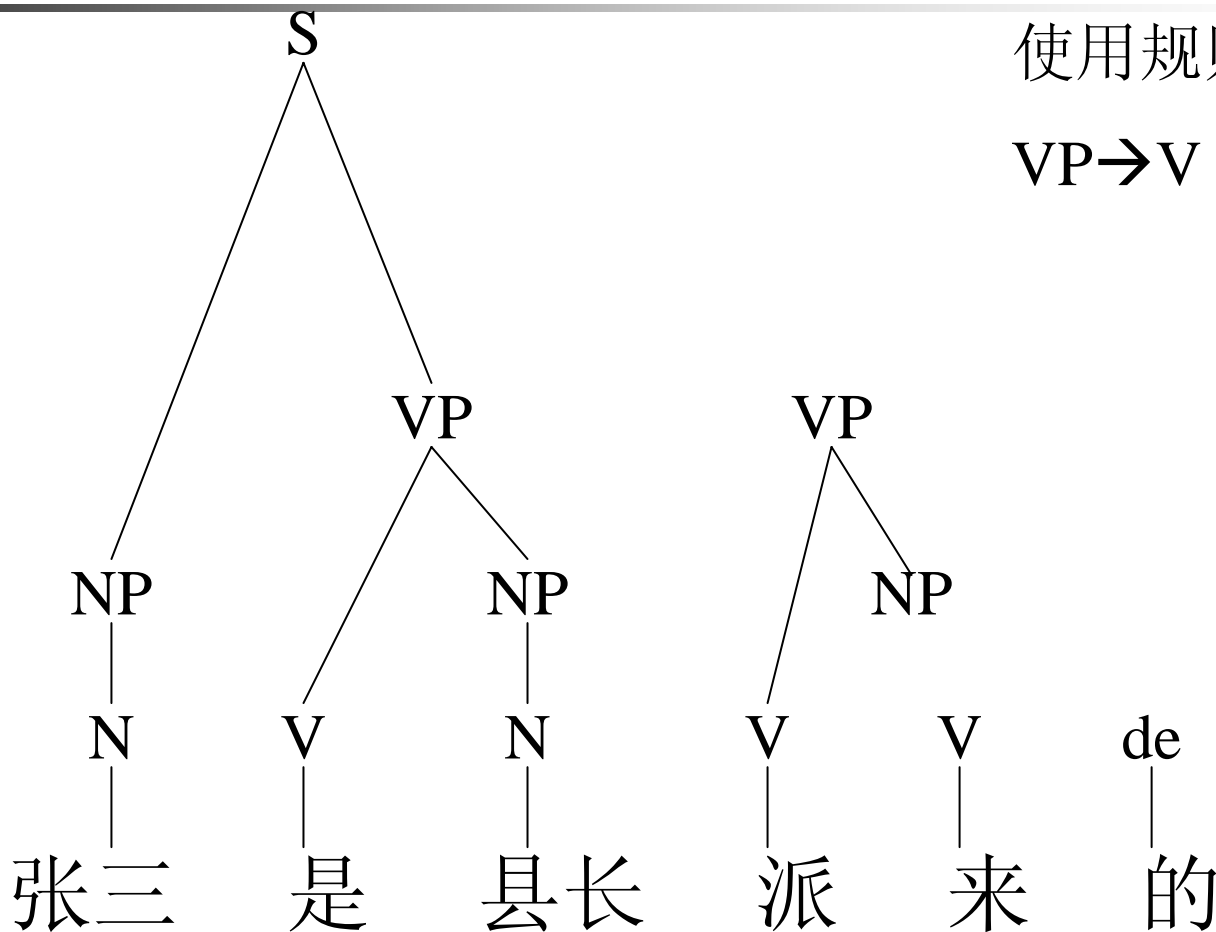
待分析字符串归结为S，但没有分析完，继续扫描

左角分析法示例-8

- (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) $V' \rightarrow V V$

使用规则:

$VP \rightarrow V NP$

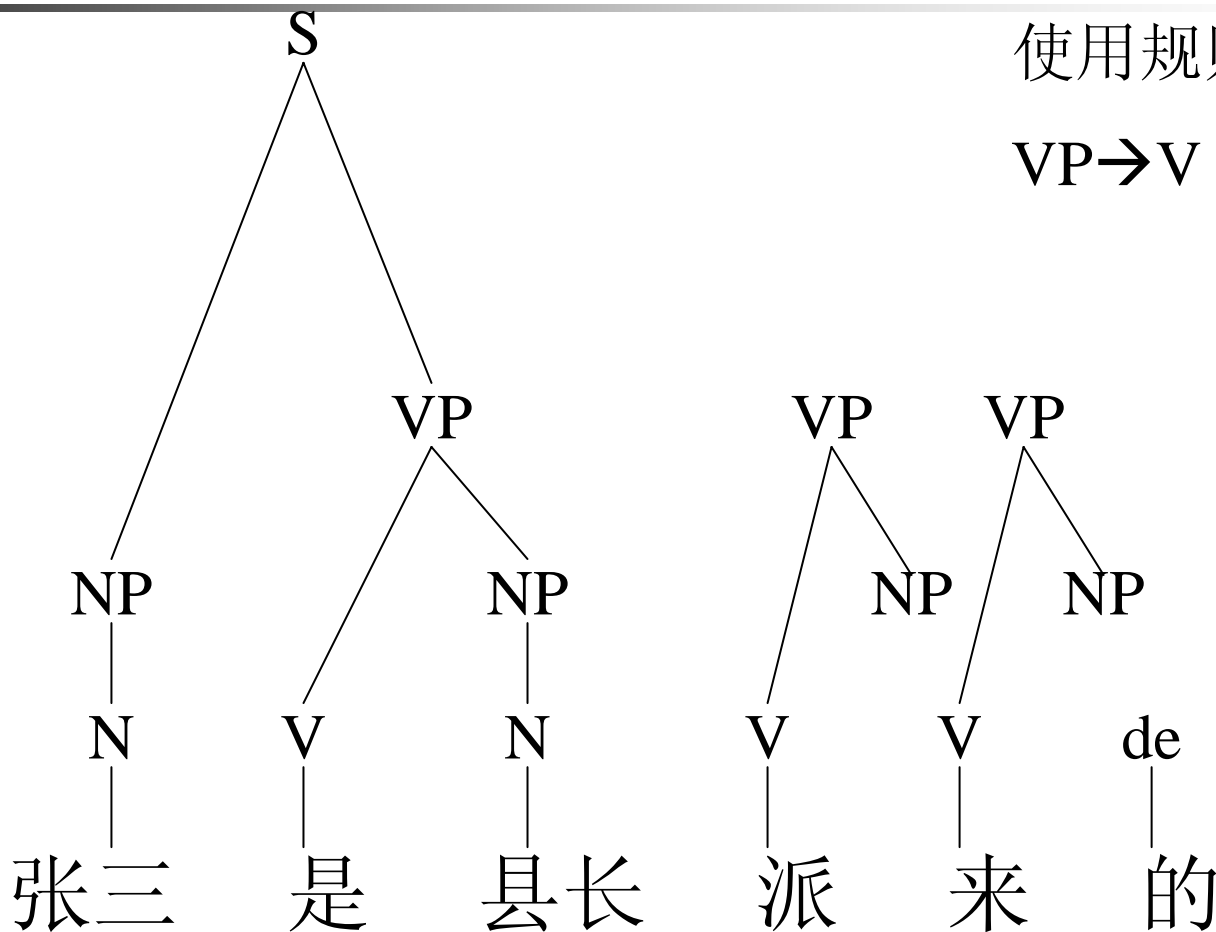


- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

左角分析法示例-9

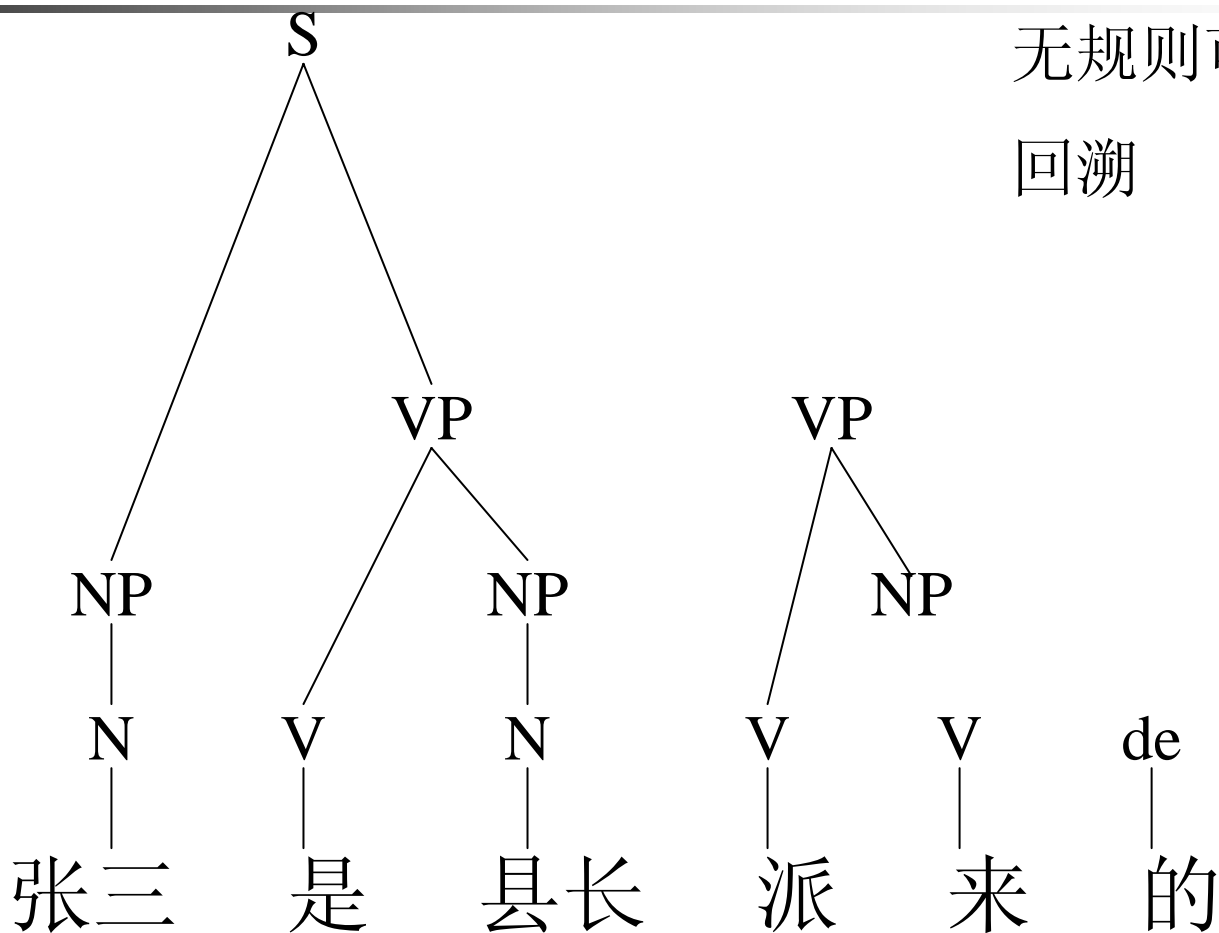
使用规则:

$VP \rightarrow V NP$



- (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) $V' \rightarrow V V$

左角分析法示例-10

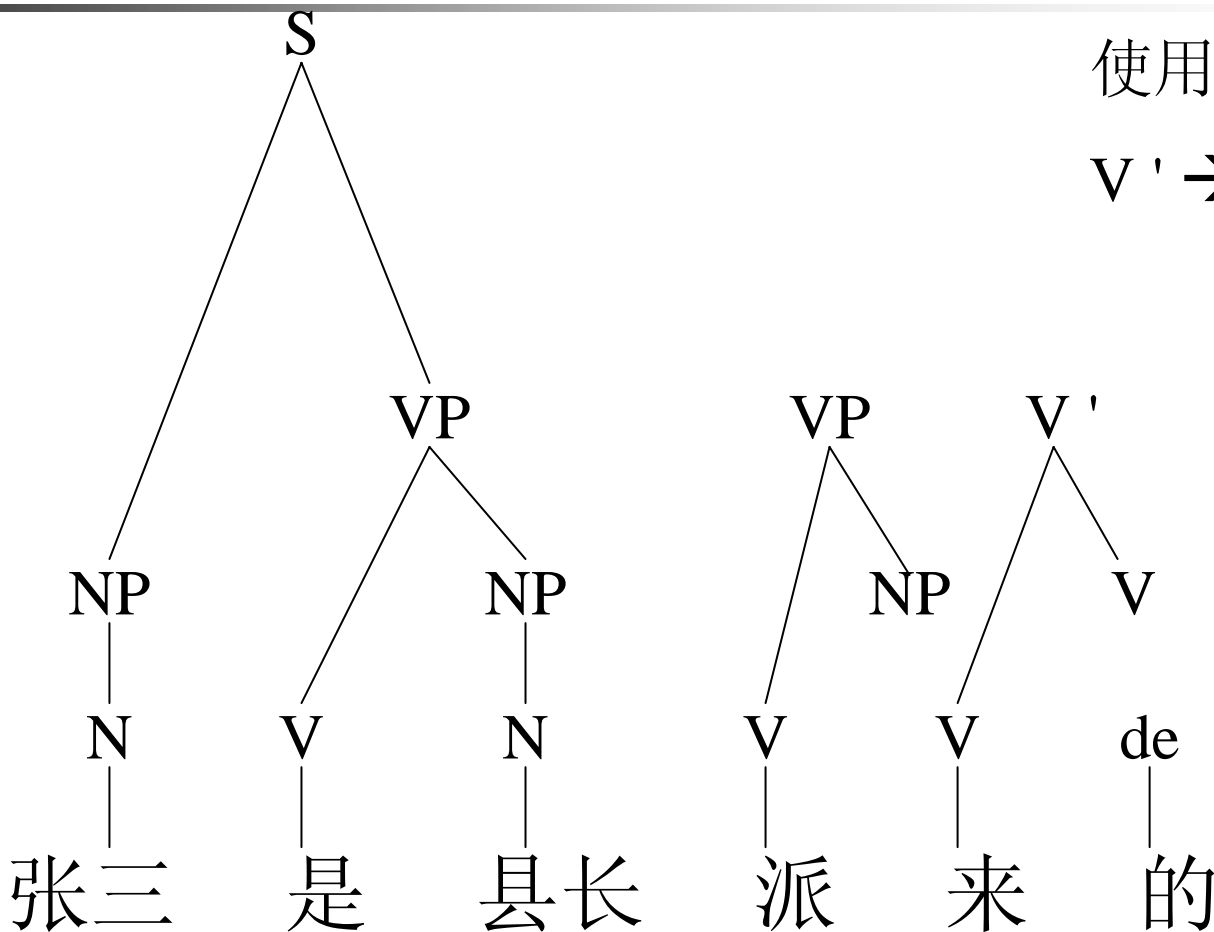


左角分析法示例-11

- (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) $V' \rightarrow V V$

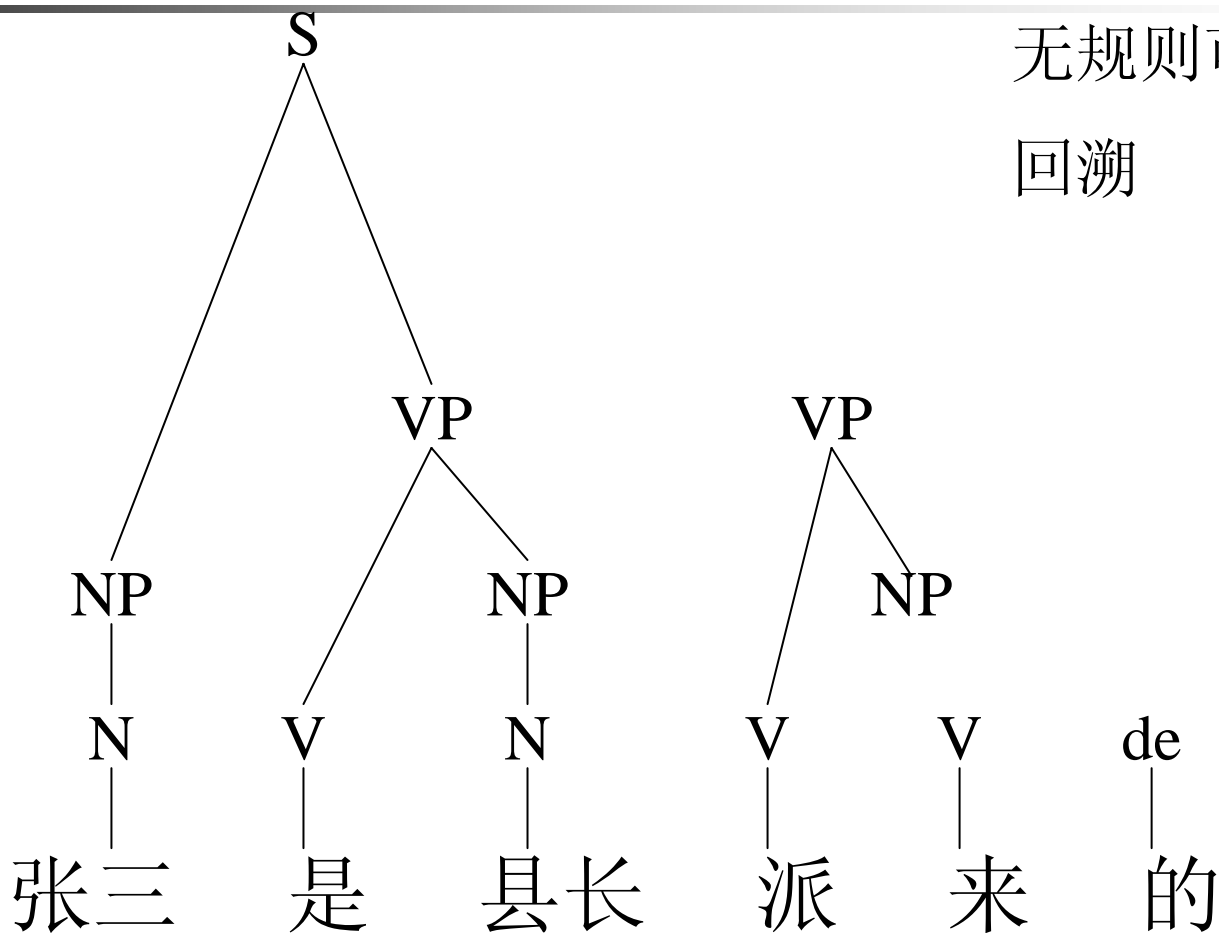
使用规则

$V' \rightarrow V V$



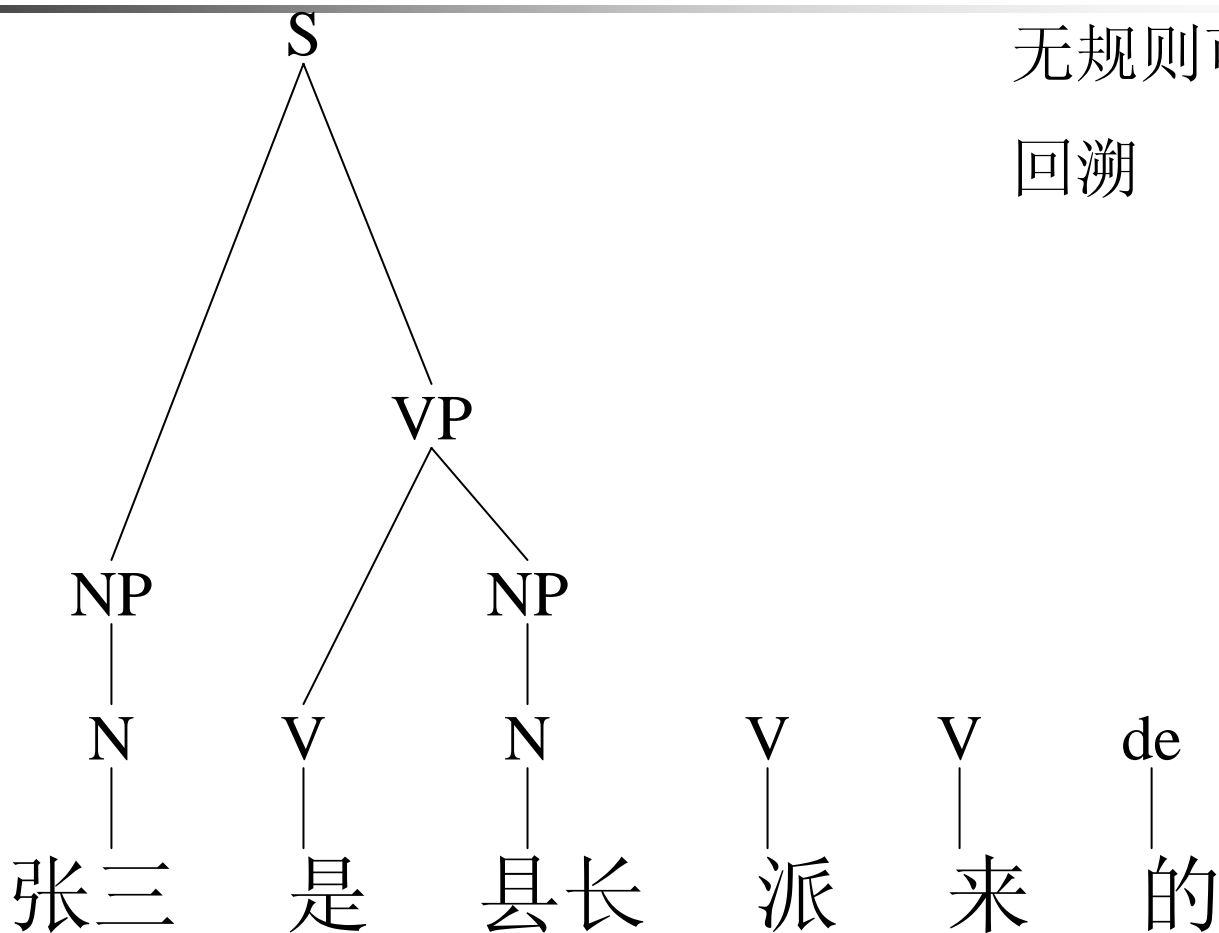
- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

左角分析法示例-12



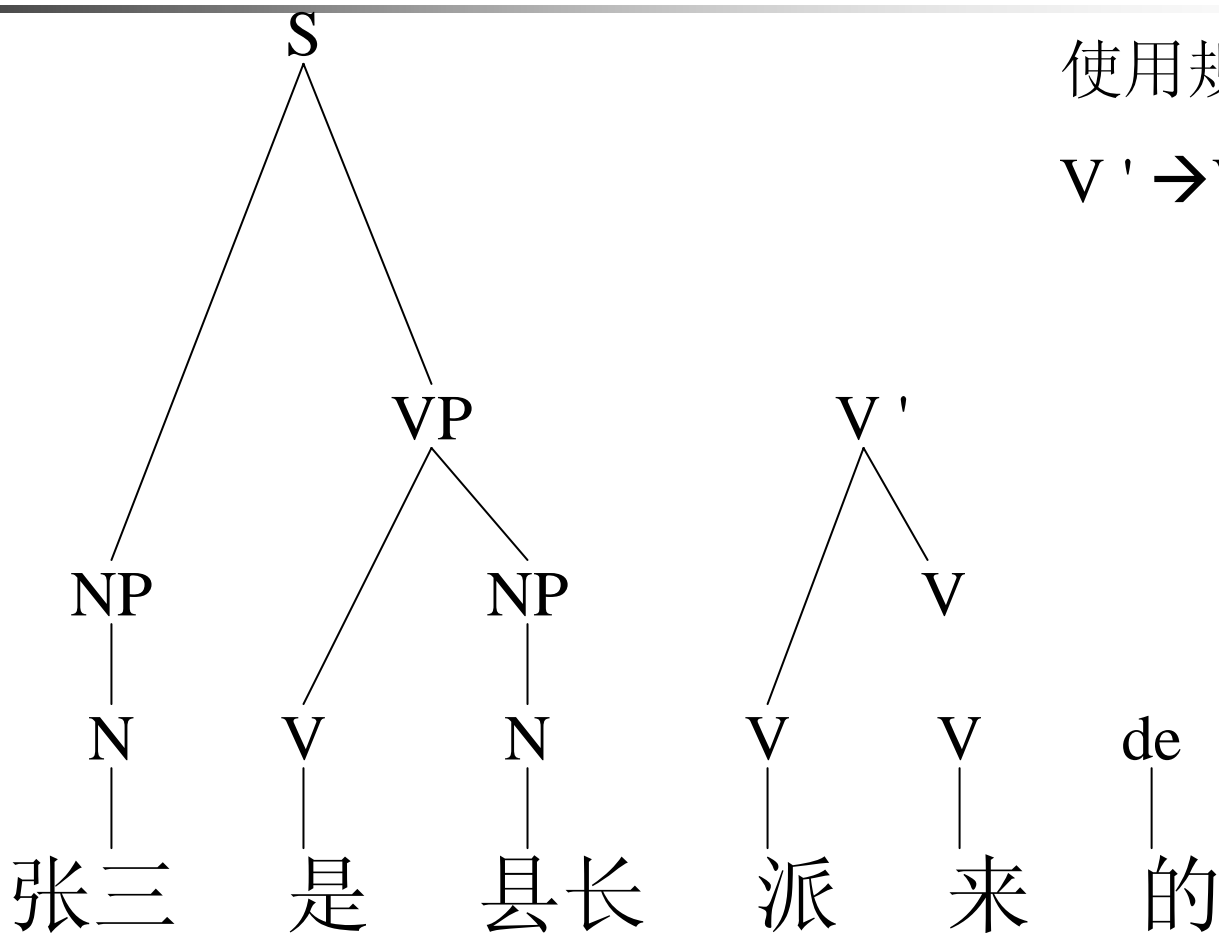
- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

左角分析法示例-13



- (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) $V' \rightarrow V V$

左角分析法示例-14

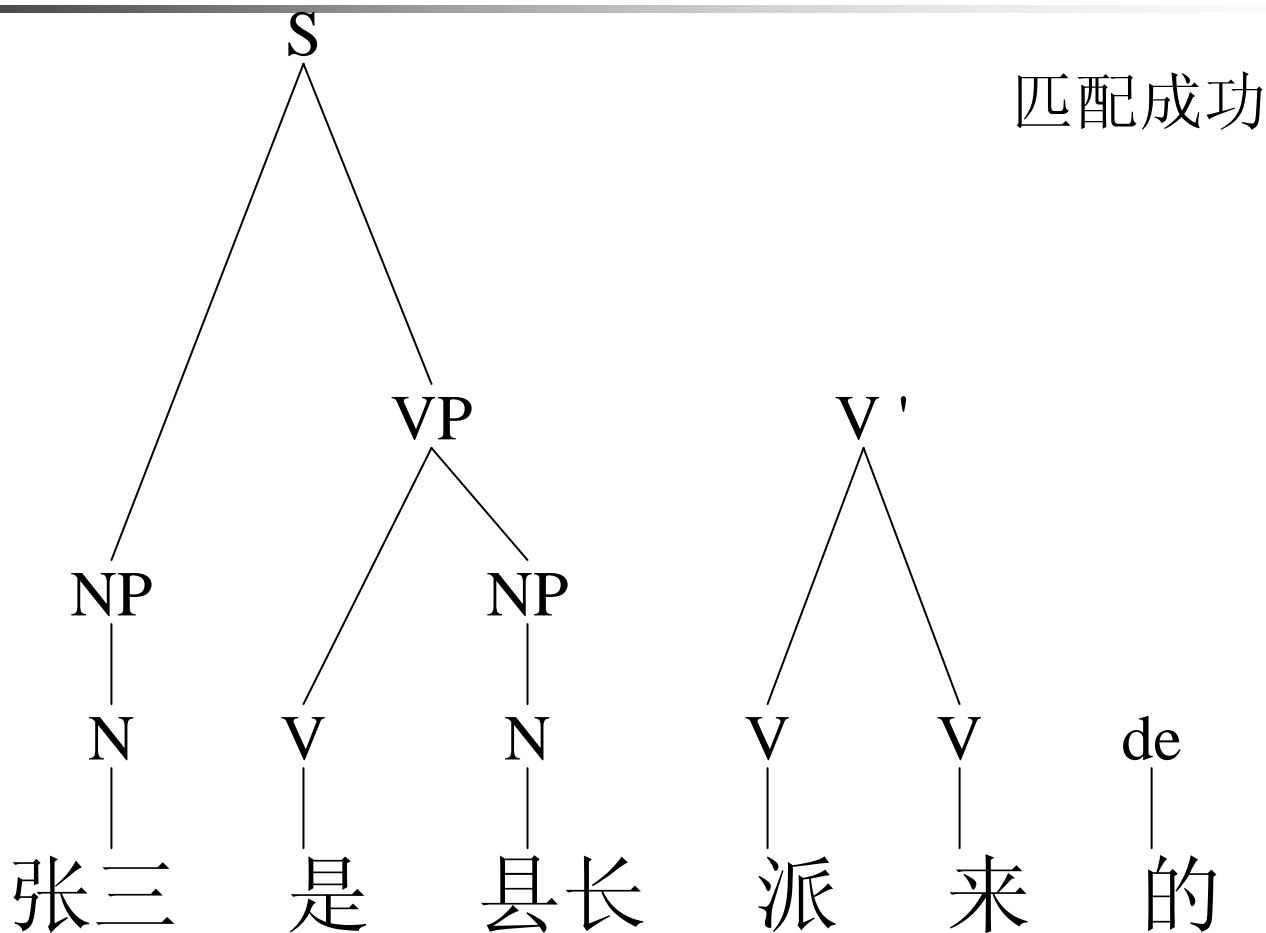


使用规则

$V' \rightarrow V V$

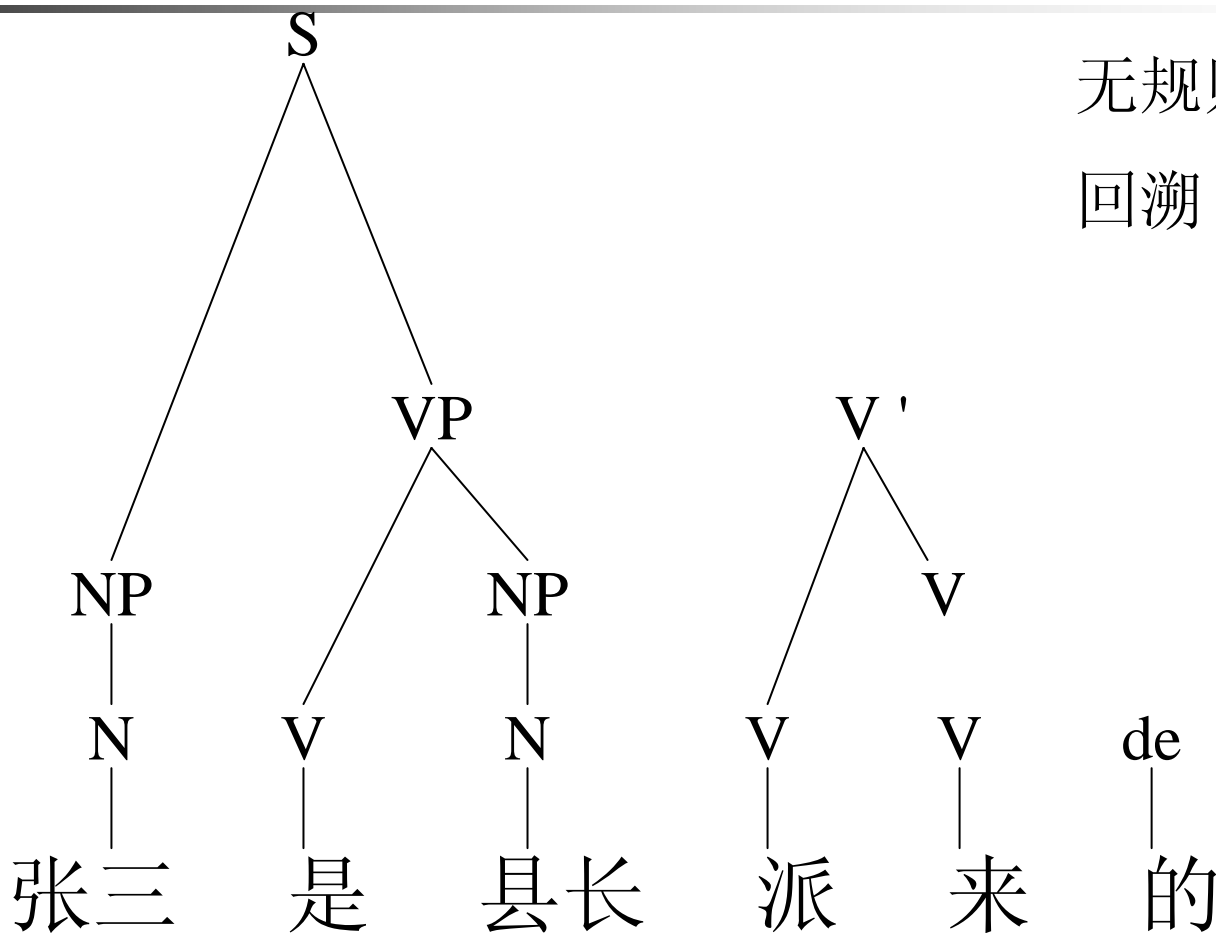
- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

左角分析法示例-15



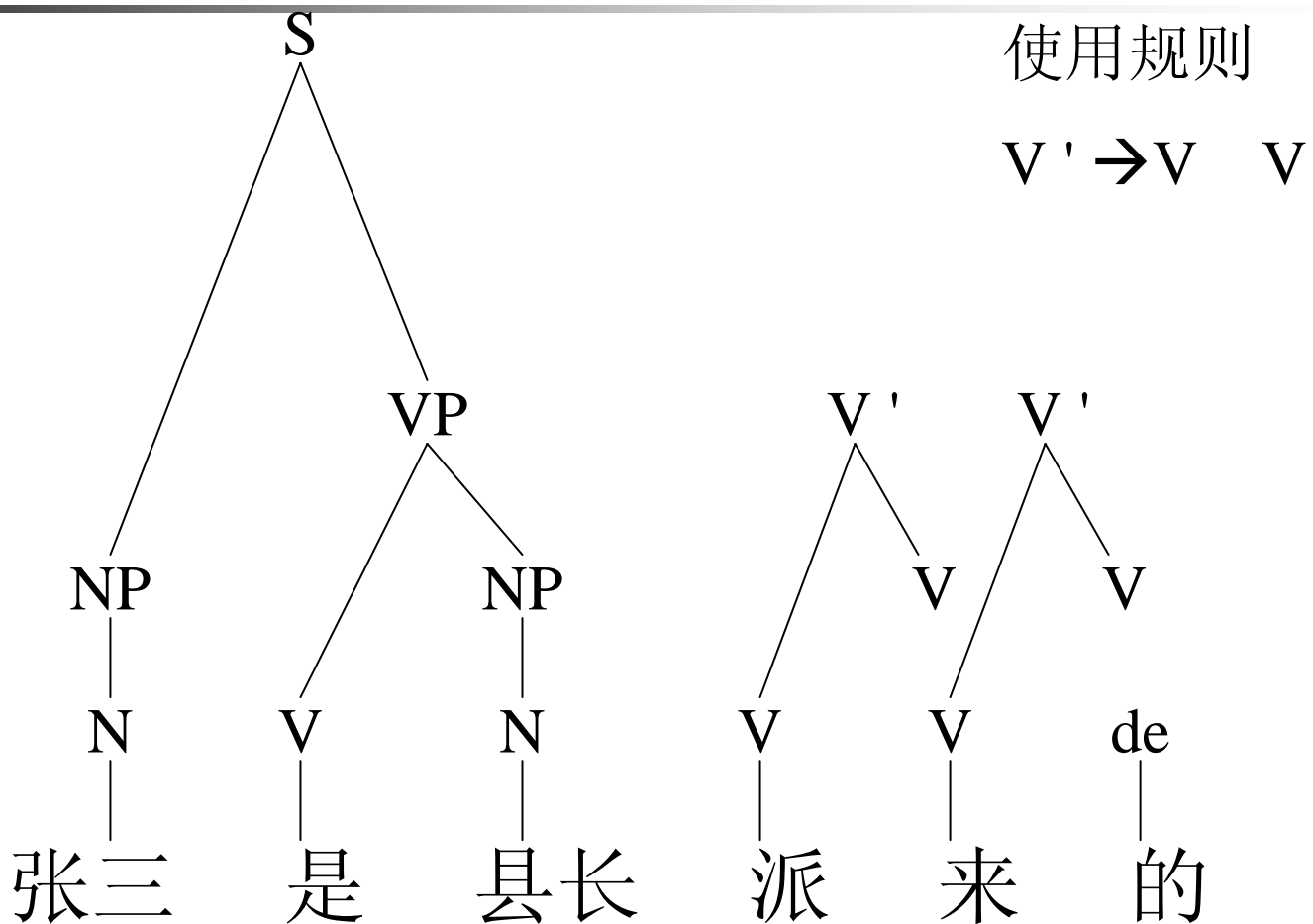
左角分析法示例-16

- (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) $V' \rightarrow V V$



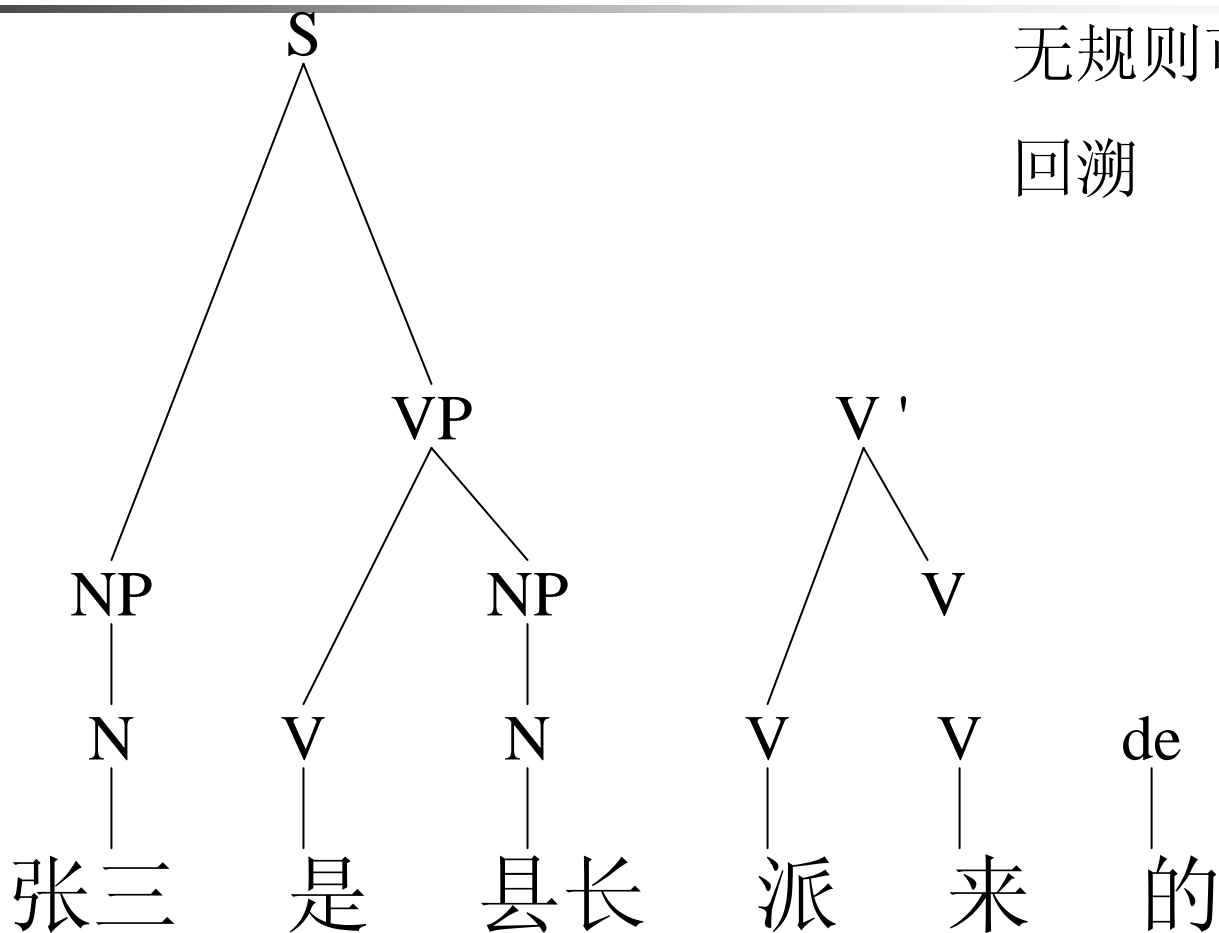
- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

左角分析法示例-17



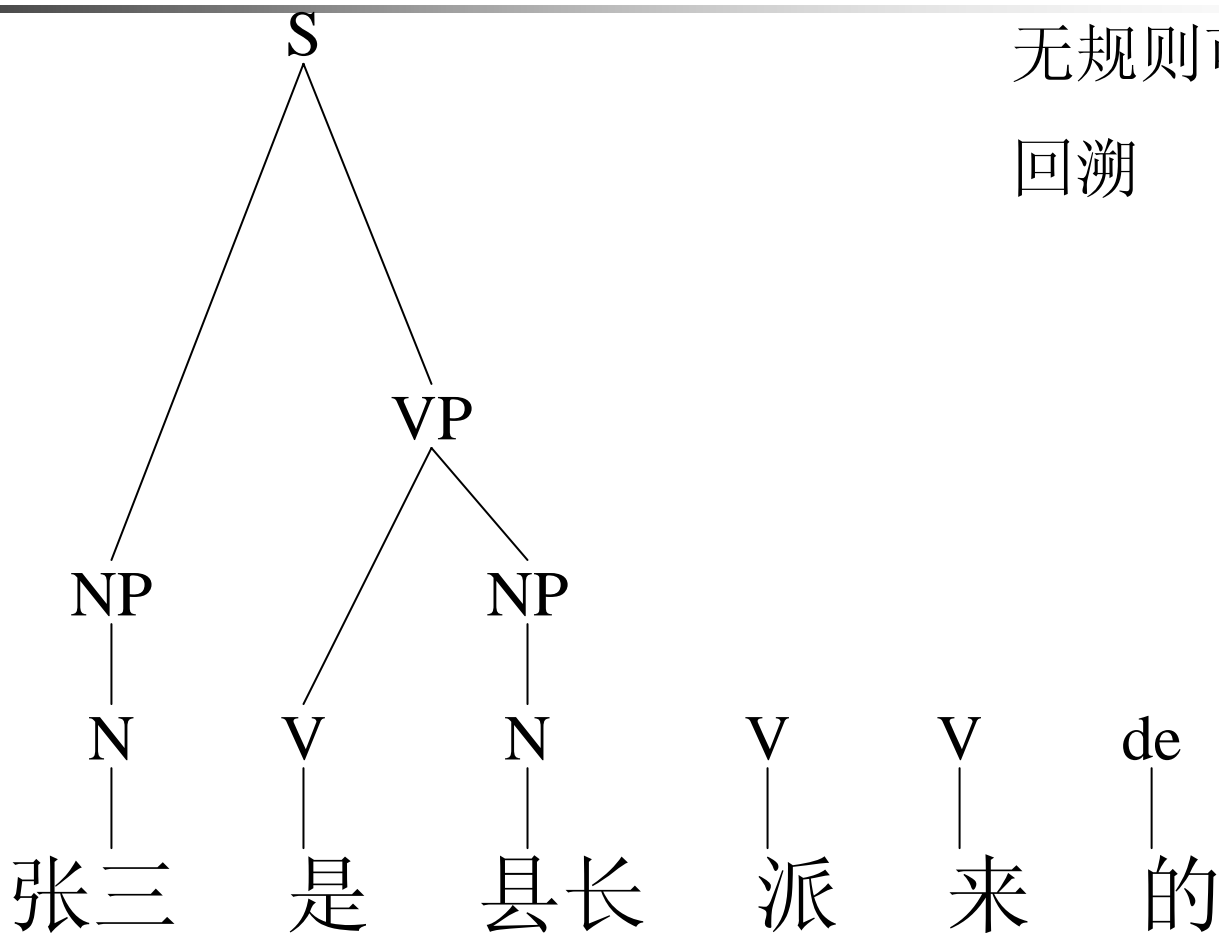
- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

左角分析法示例-18



- (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) $V' \rightarrow V V$

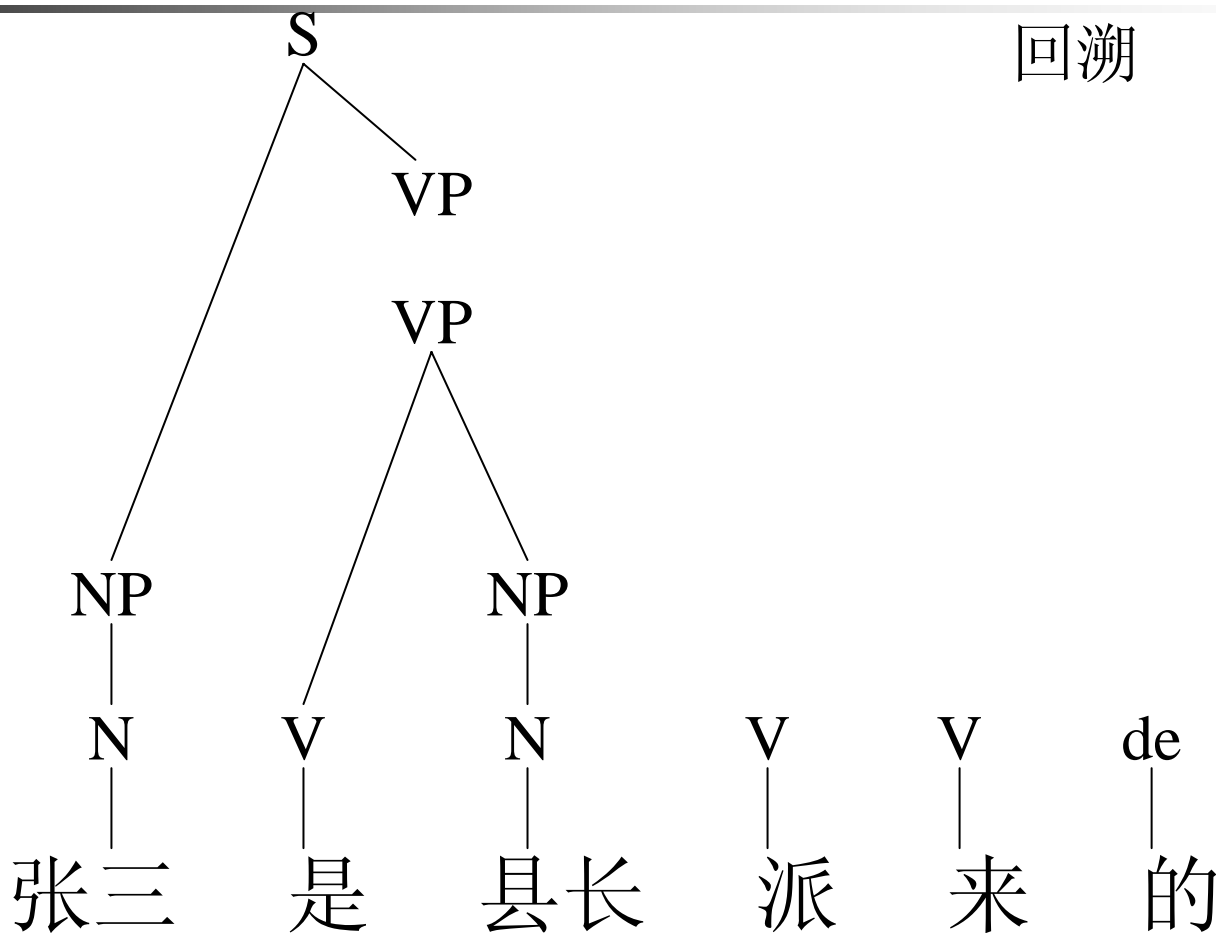
左角分析法示例-19



- (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) $V' \rightarrow V V$

左角分析法示例-20

回溯

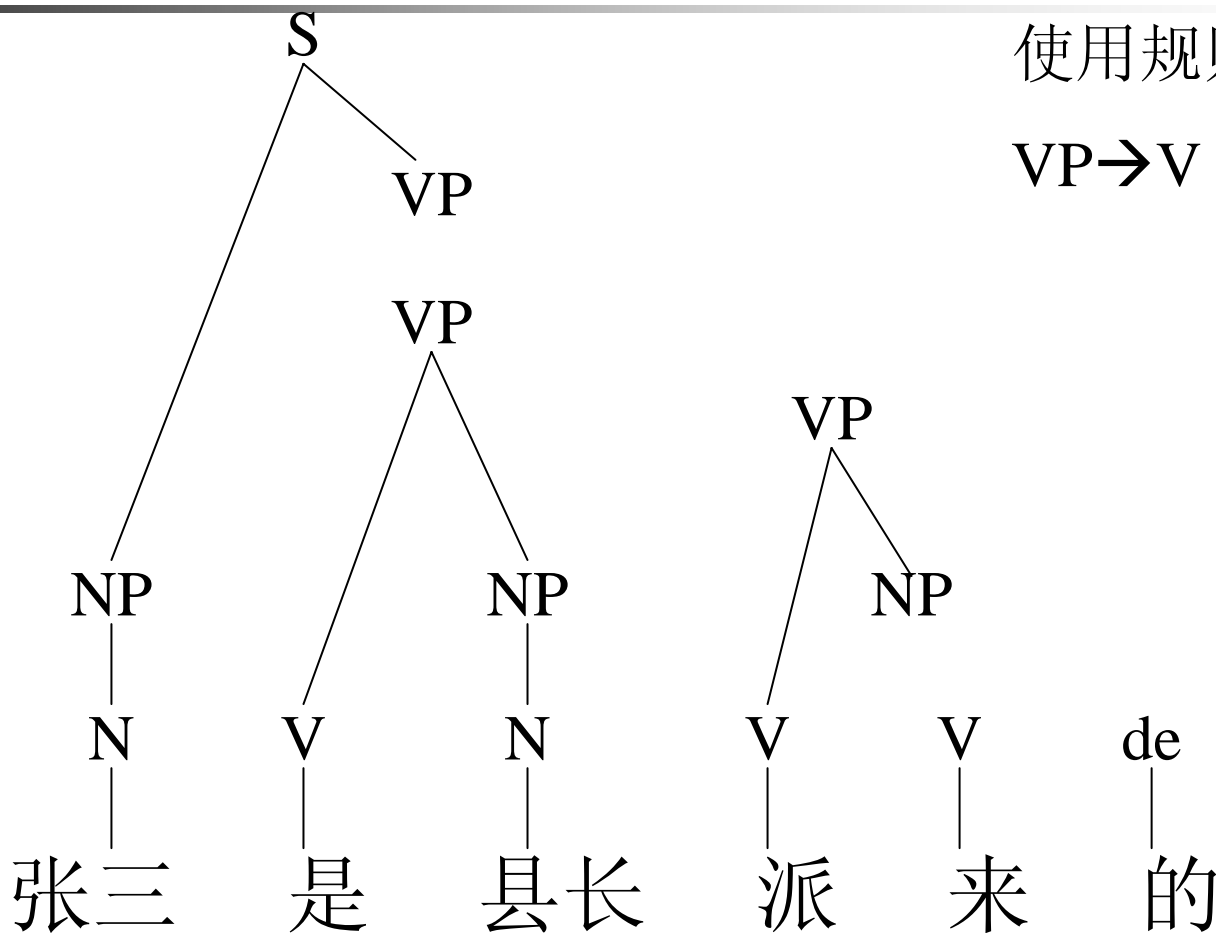


- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

左角分析法示例-21

使用规则:

$VP \rightarrow V NP$

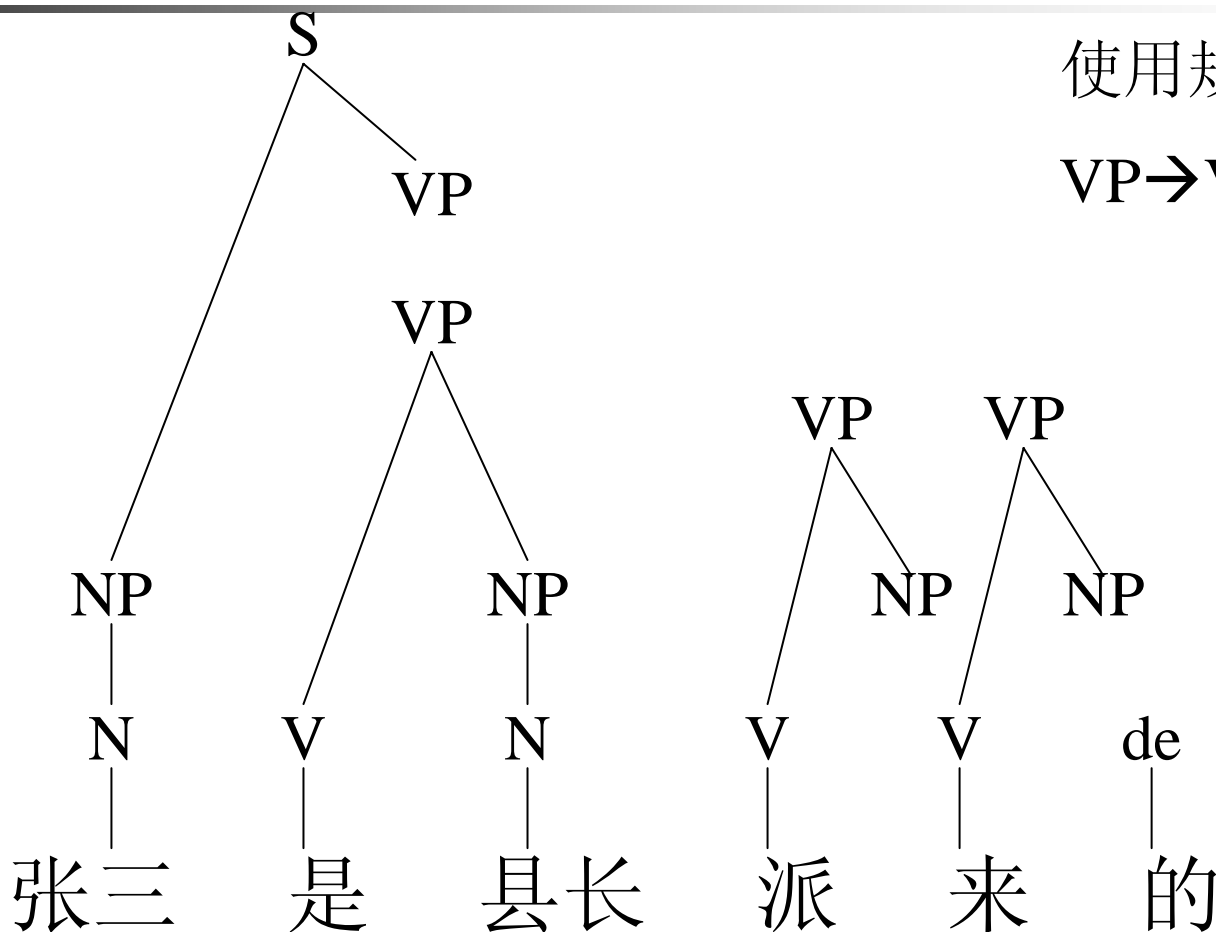


- (1) $S \rightarrow NP \ VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS \text{ 的}$
- (4) $CS \rightarrow NP \ V'$
- (5) $VP \rightarrow V \ NP$
- (6) $V' \rightarrow V \ V$

左角分析法示例-22

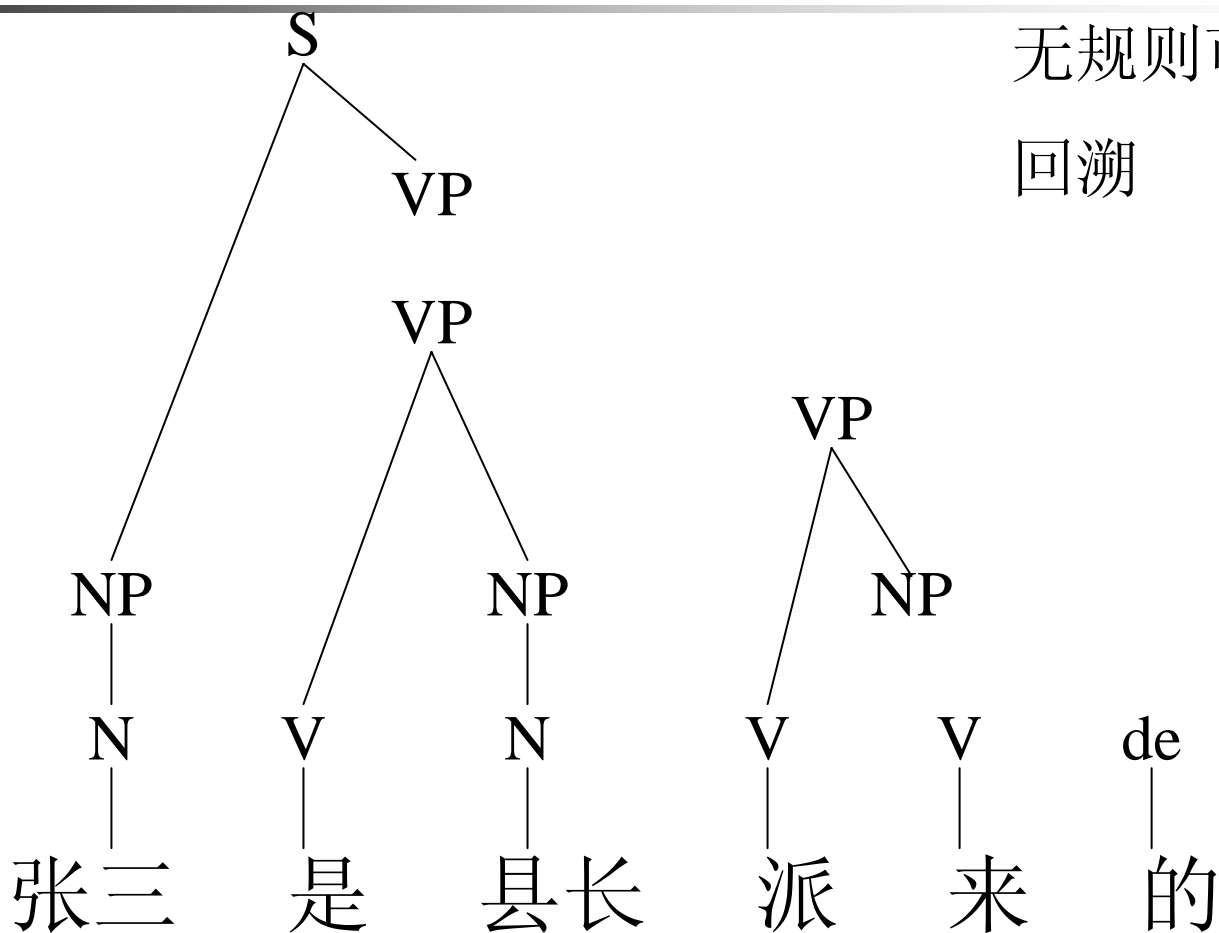
使用规则:

$VP \rightarrow V \ NP$



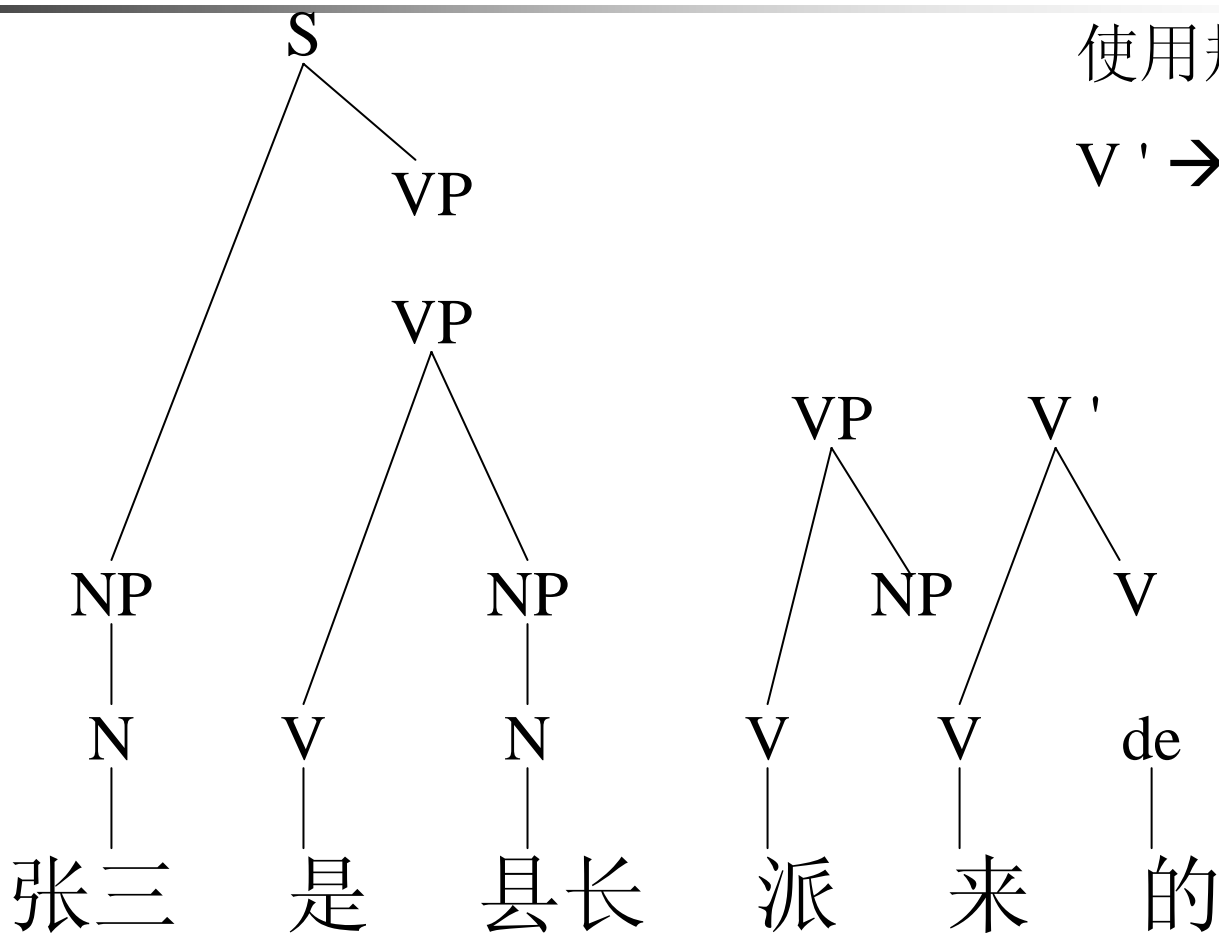
- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

左角分析法示例-23



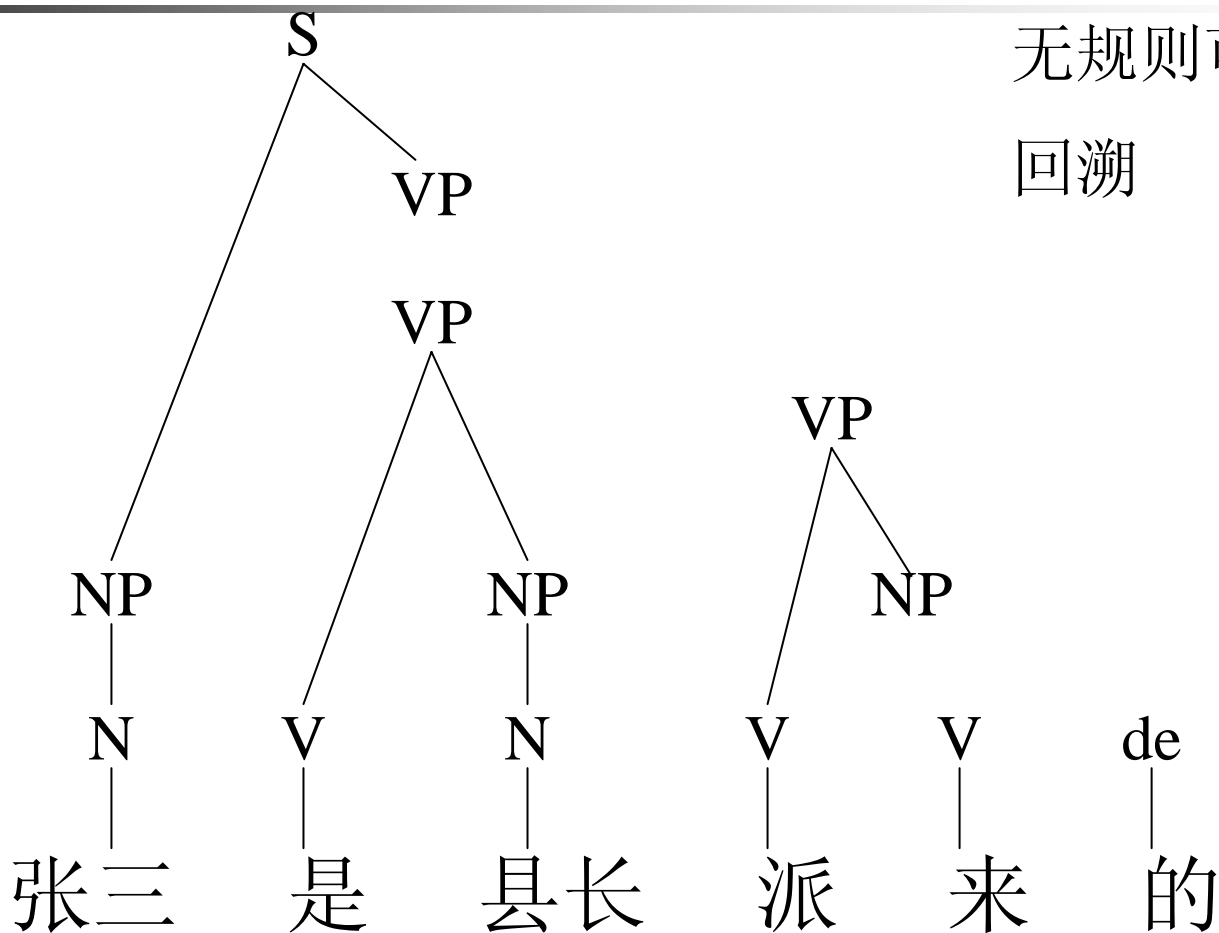
- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

左角分析法示例-24



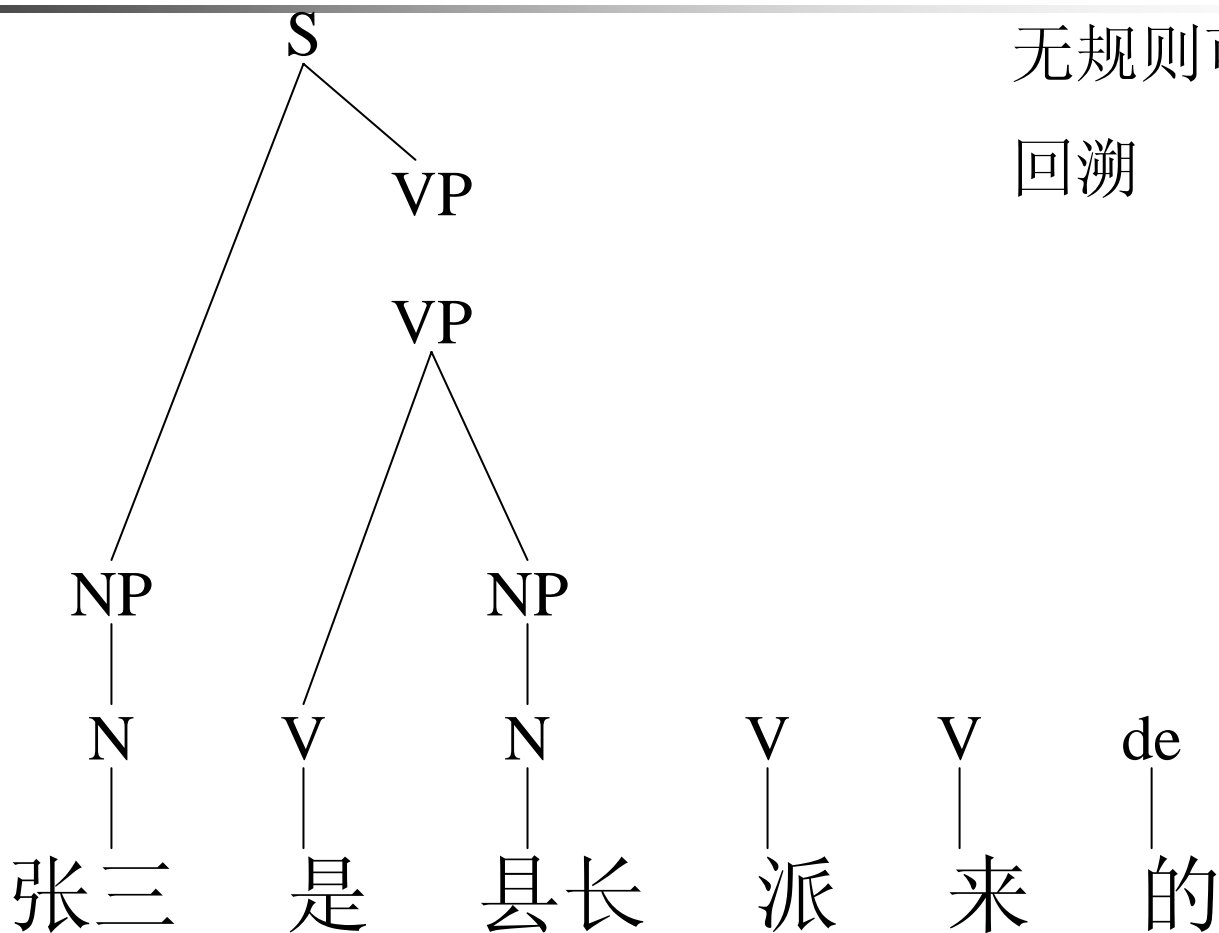
- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

左角分析法示例-25



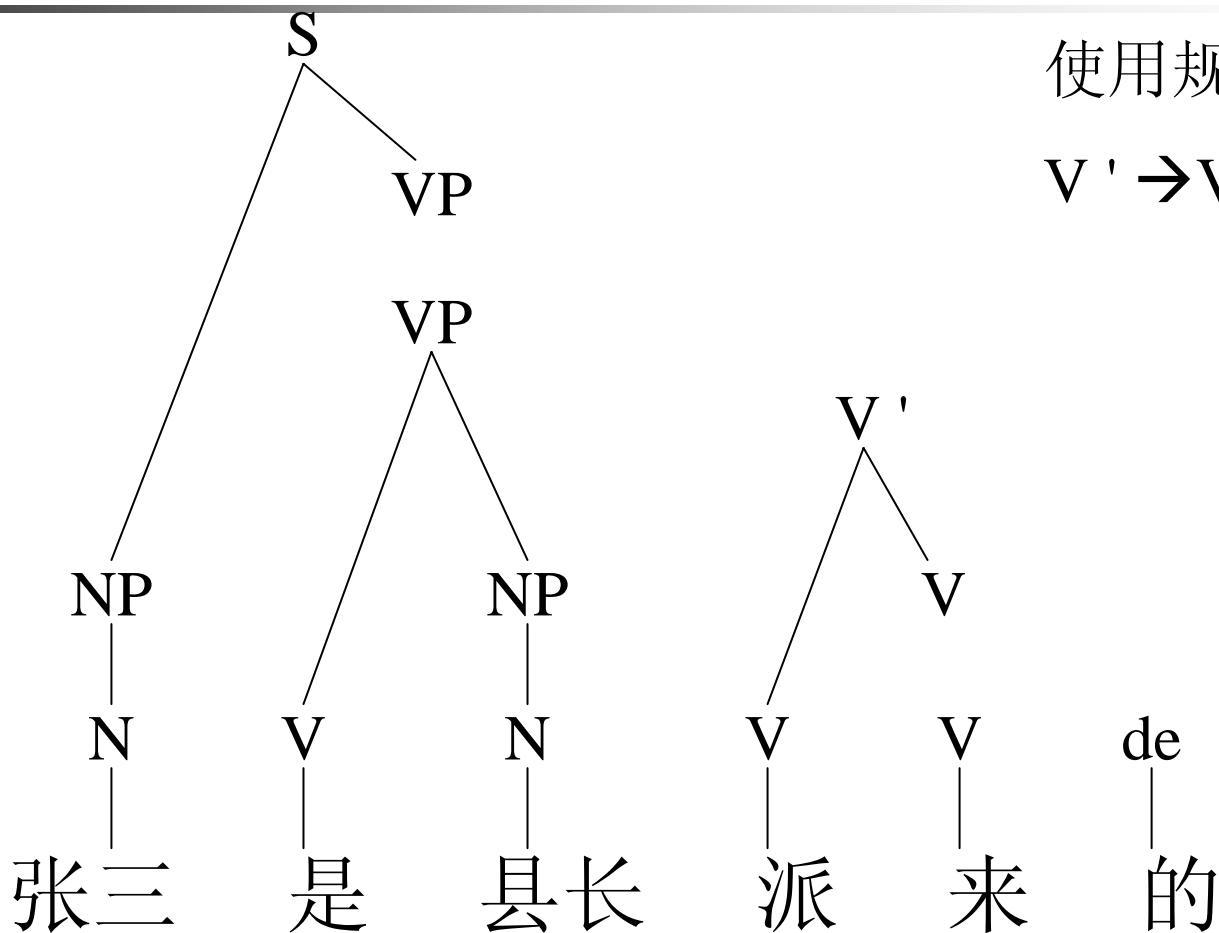
- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

左角分析法示例-26



- (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) $V' \rightarrow V V$

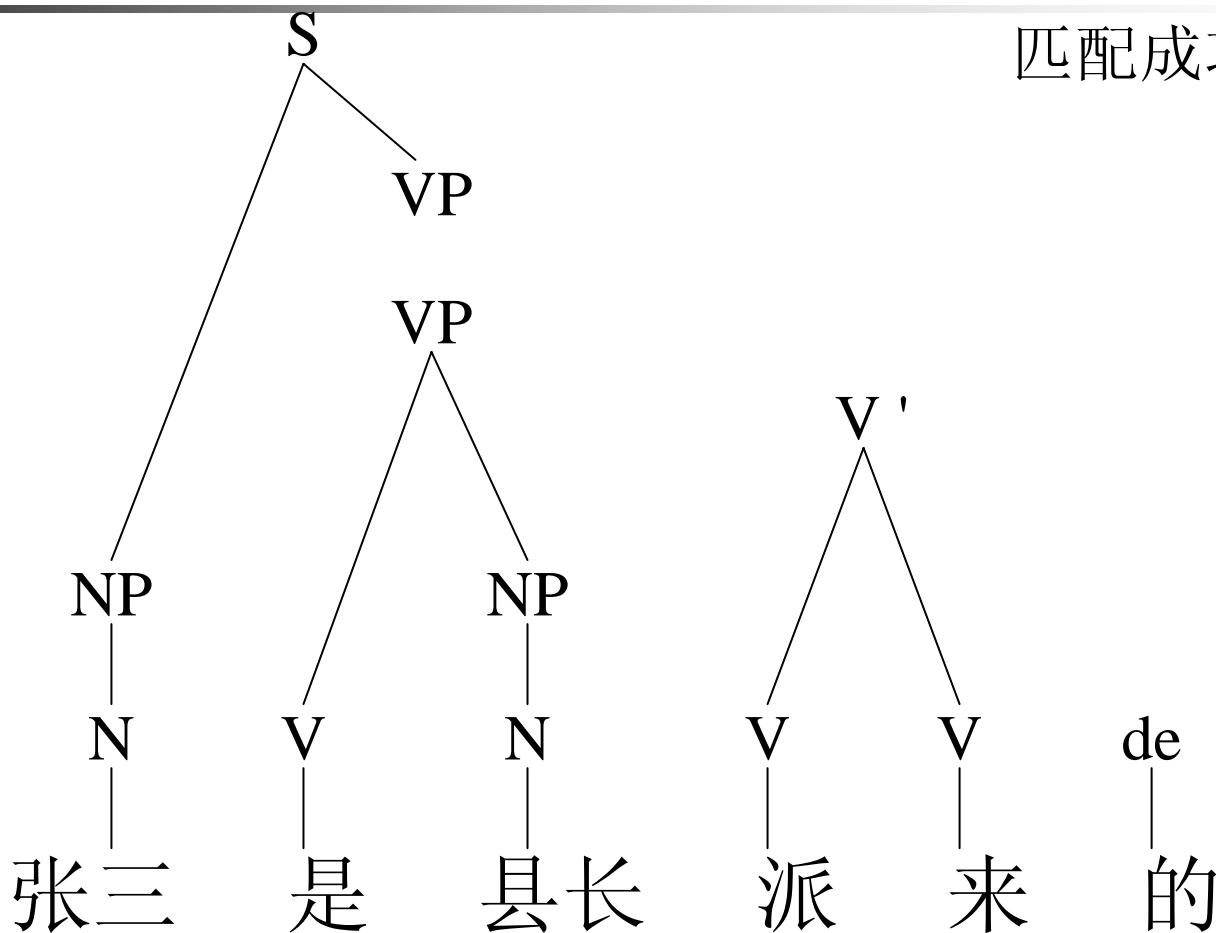
左角分析法示例-27



- (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) $V' \rightarrow V V$

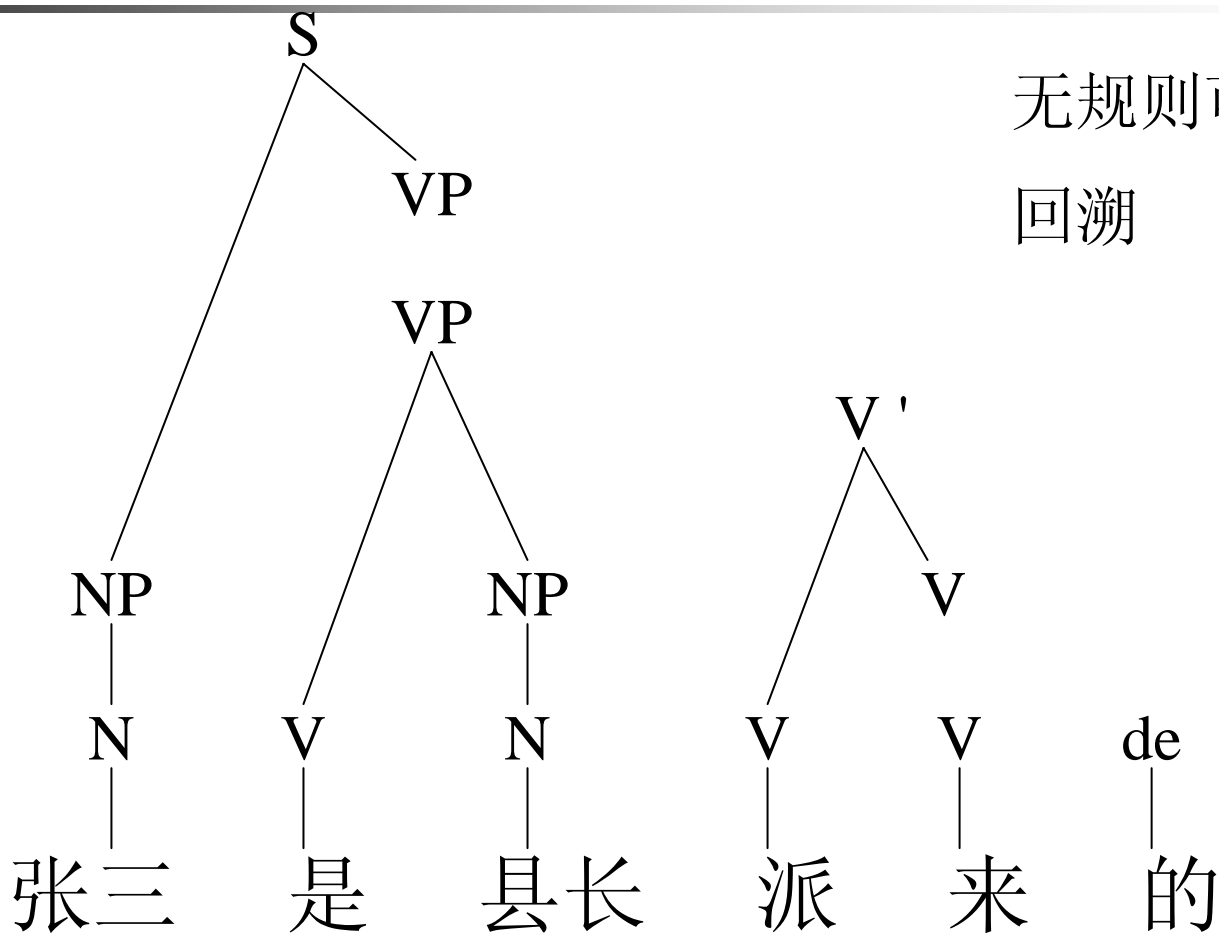
左角分析法示例-28

匹配成功



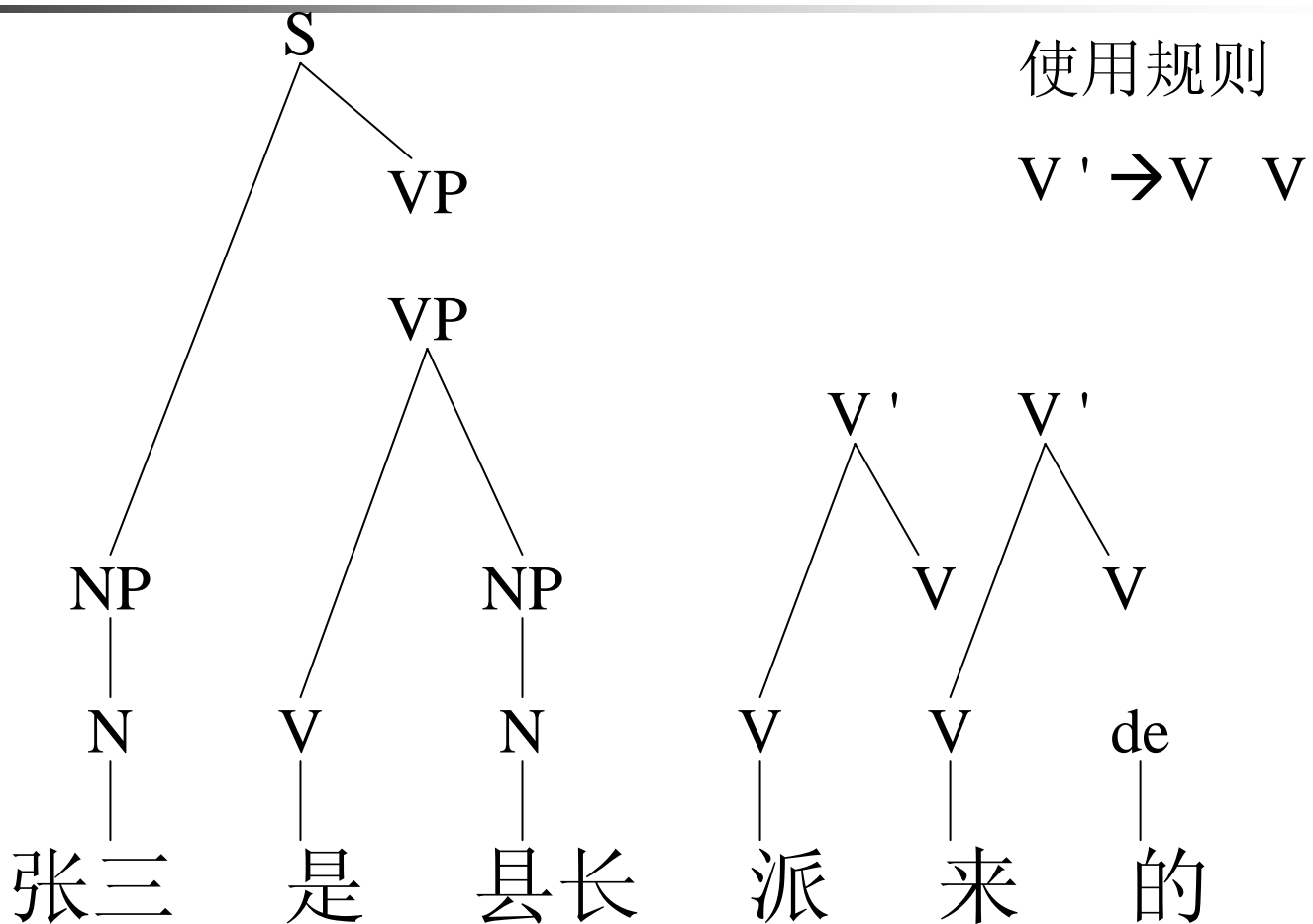
- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

左角分析法示例-29



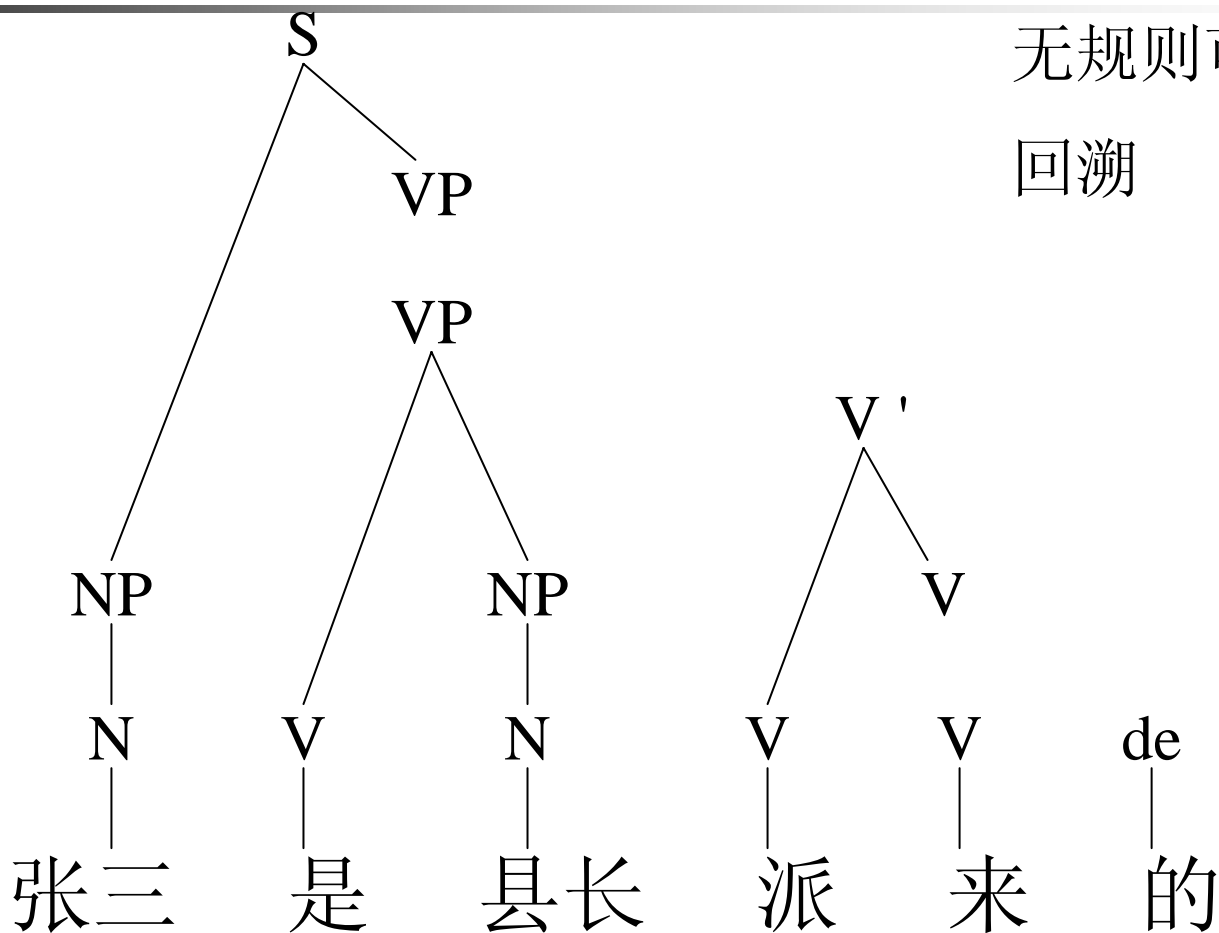
- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

左角分析法示例-30



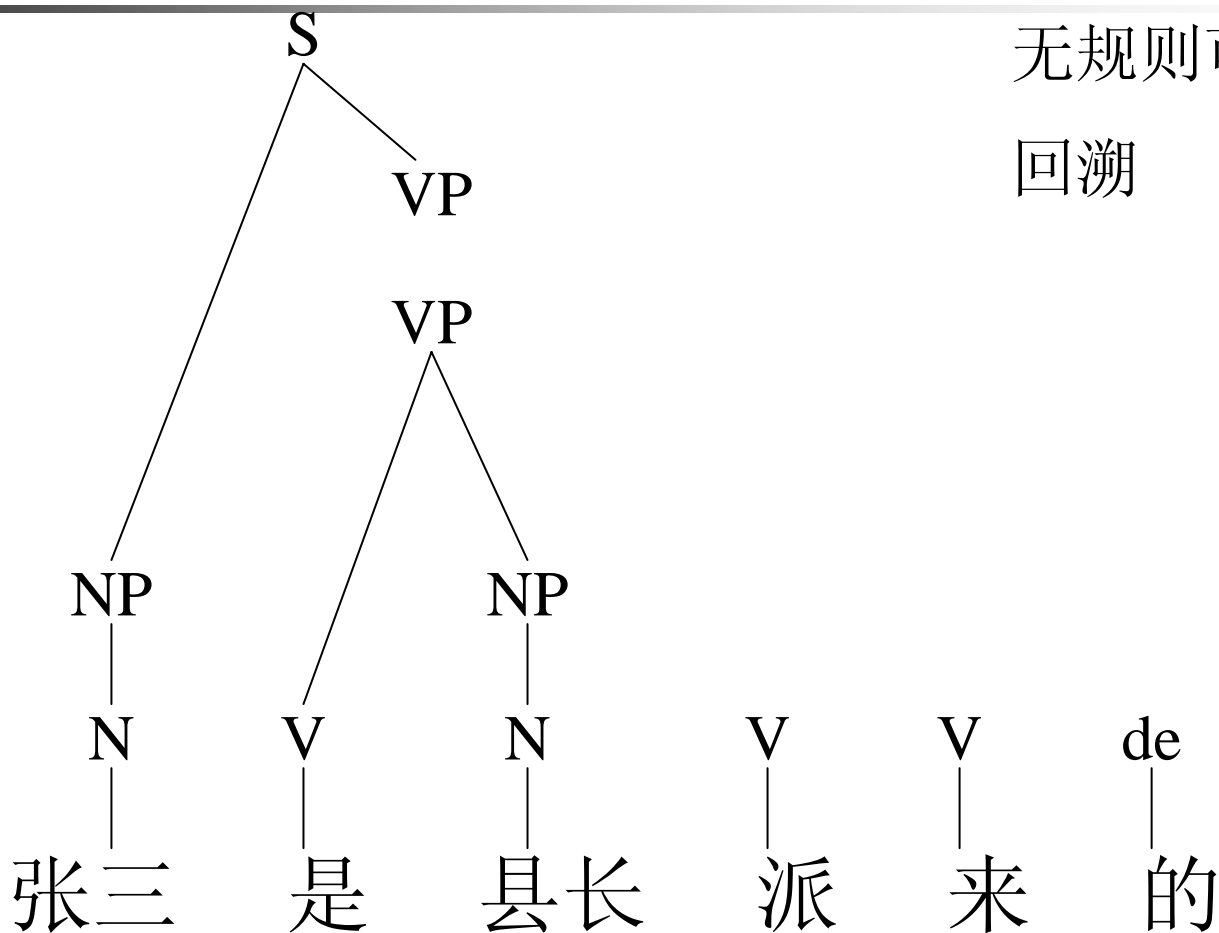
- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

左角分析法示例-31



- (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) $V' \rightarrow V V$

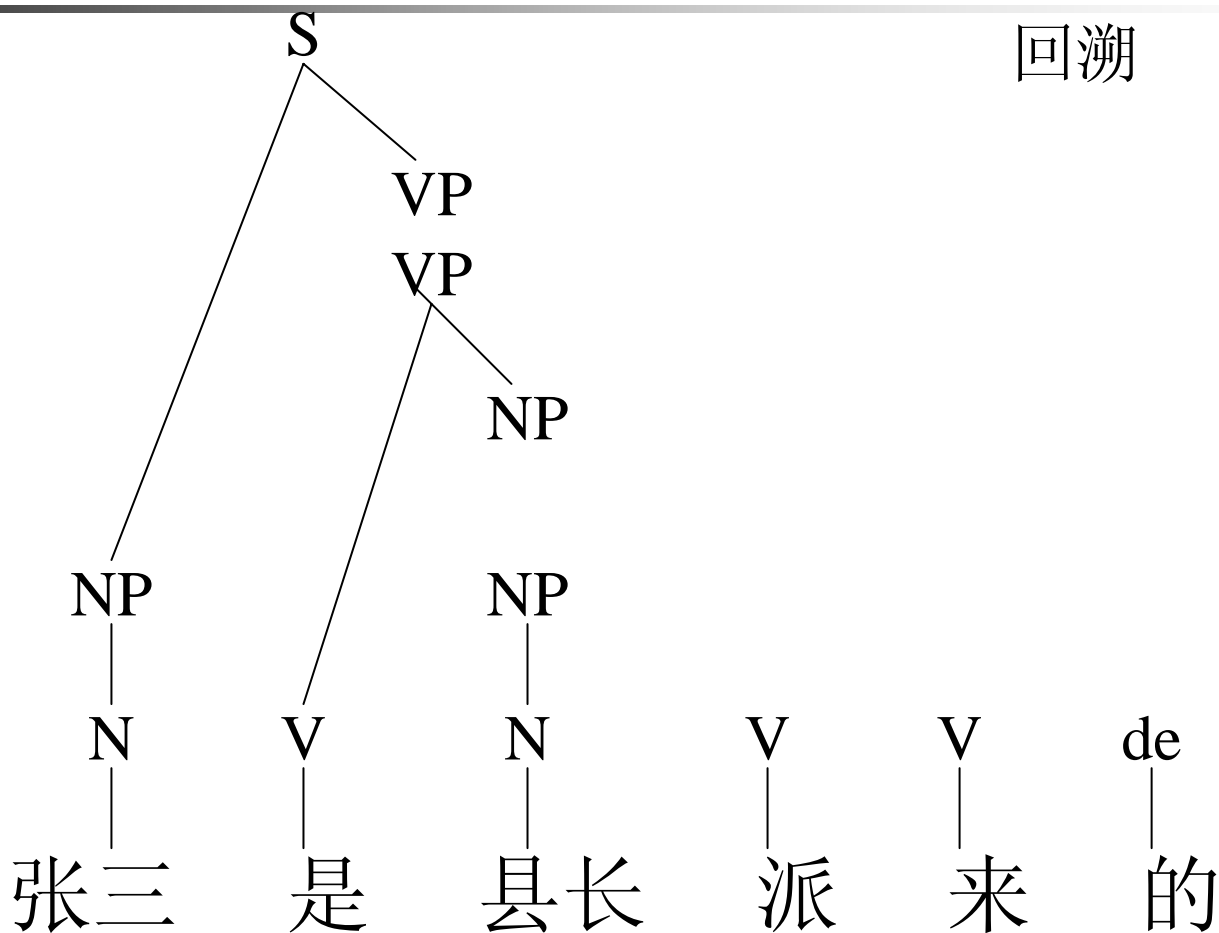
左角分析法示例-32



- (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) $V' \rightarrow V V$

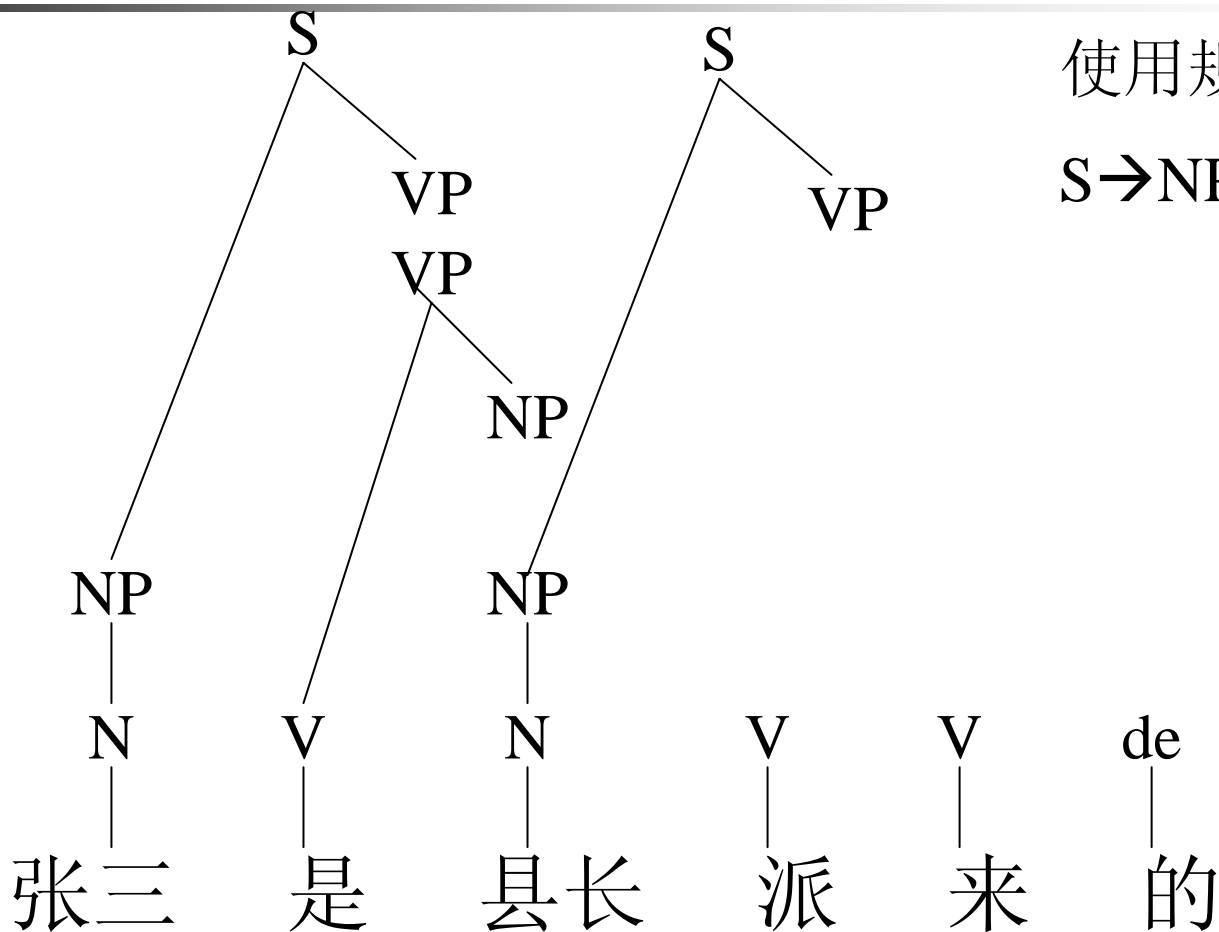
左角分析法示例-33

回溯



- (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) $V' \rightarrow V V$

左角分析法示例-34

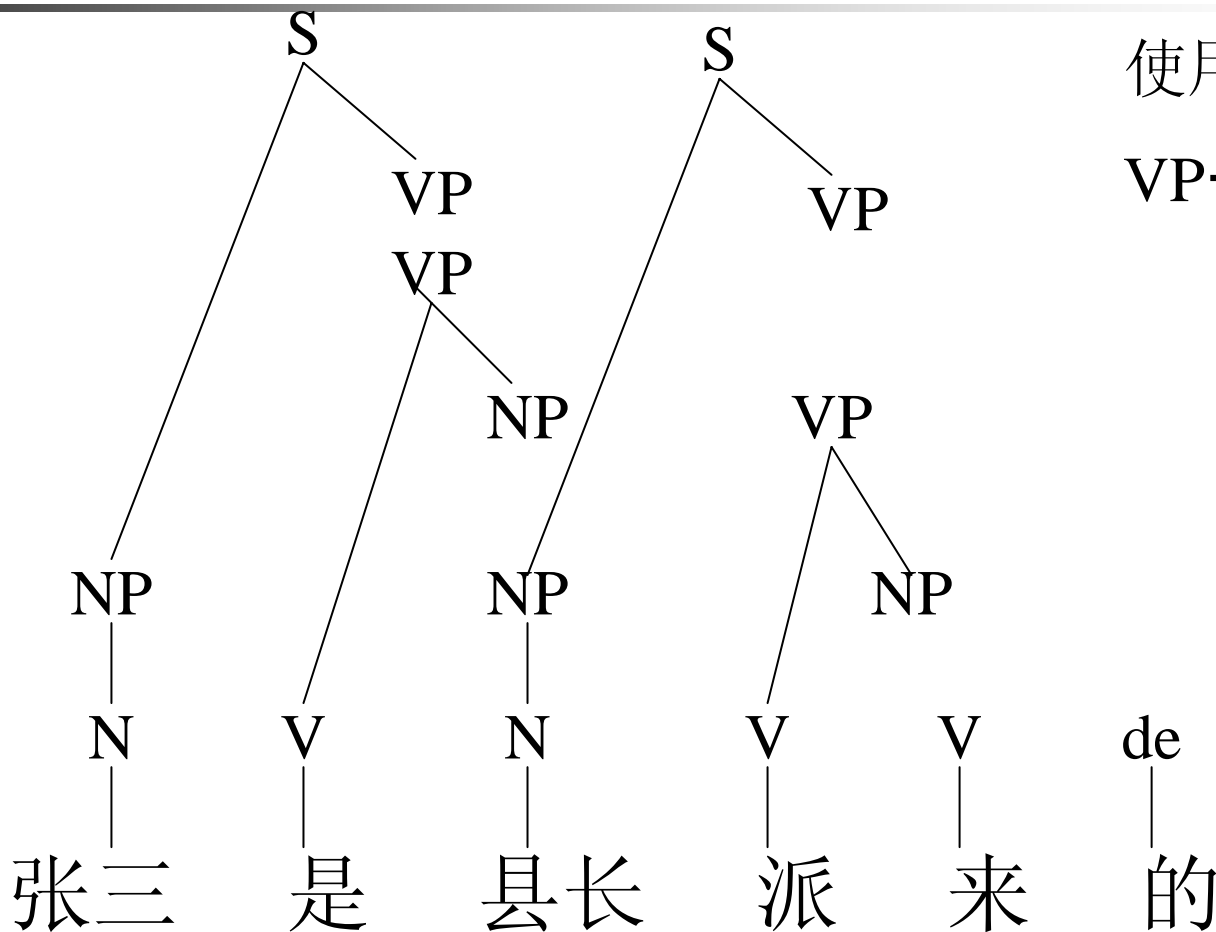


使用规则

$S \rightarrow NP VP$

- (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) $V' \rightarrow V V$

左角分析法示例-35

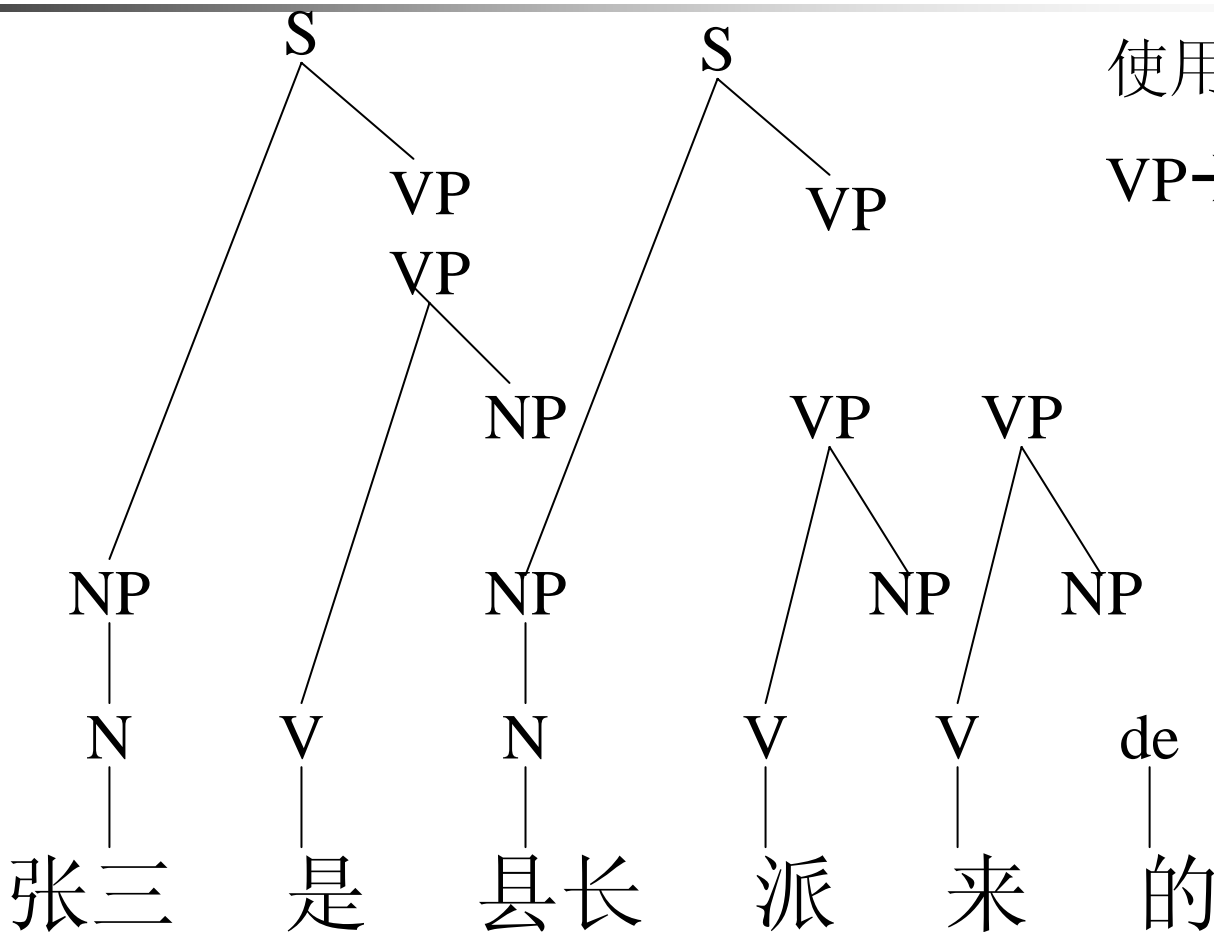


使用规则:

$VP \rightarrow V NP$

- (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) $V' \rightarrow V V$

左角分析法示例-36

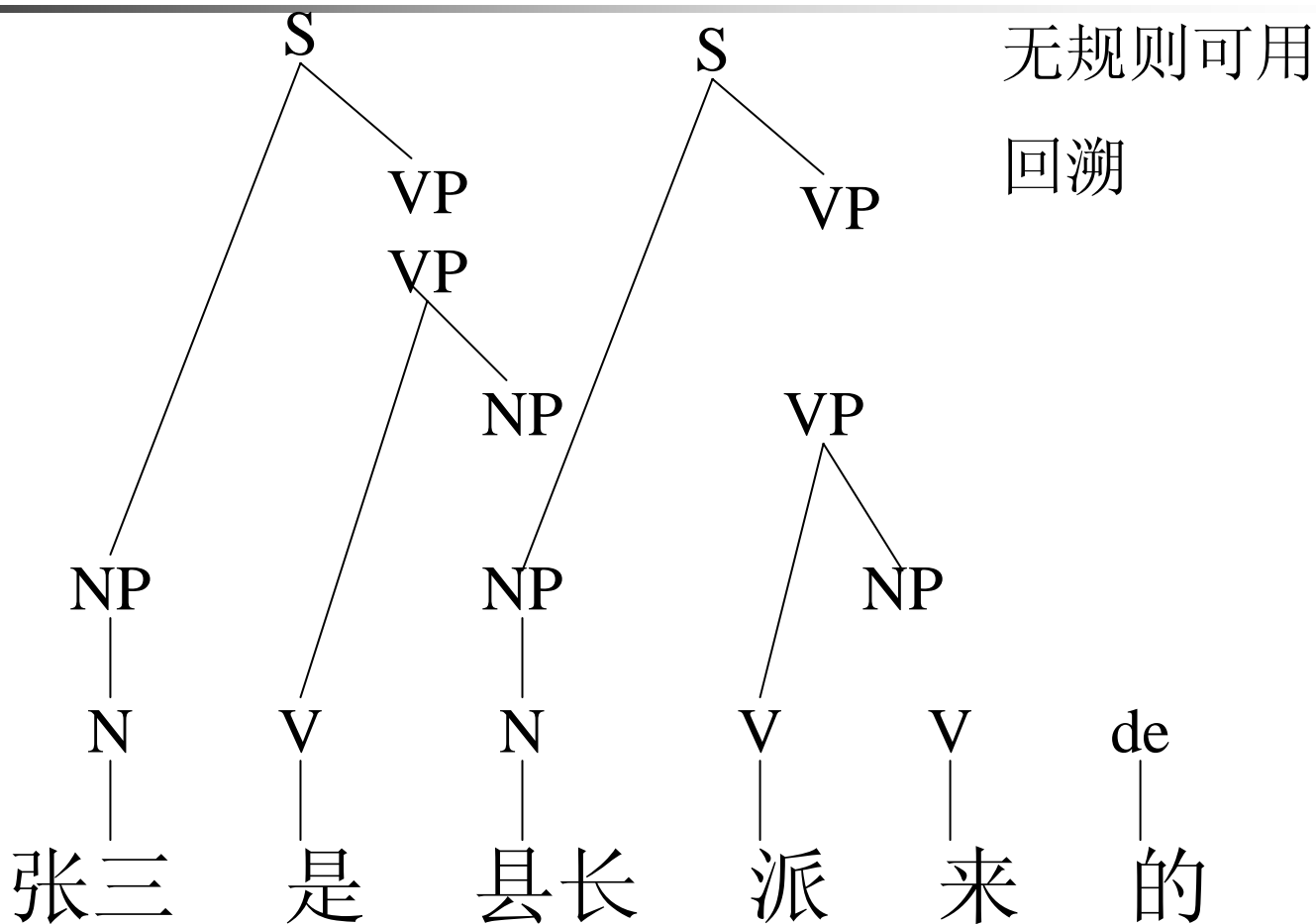


使用规则:

$VP \rightarrow V NP$

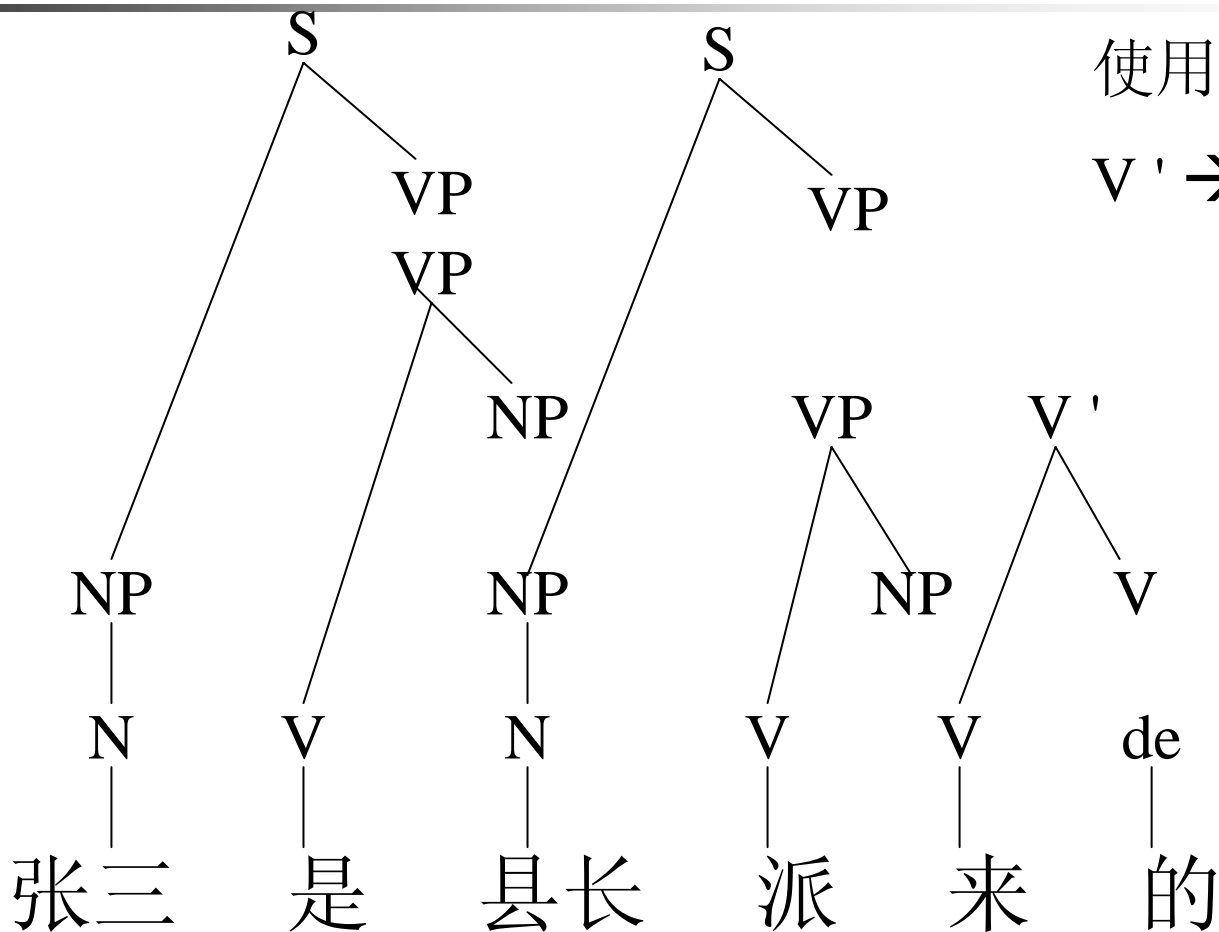
- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

左角分析法示例-37



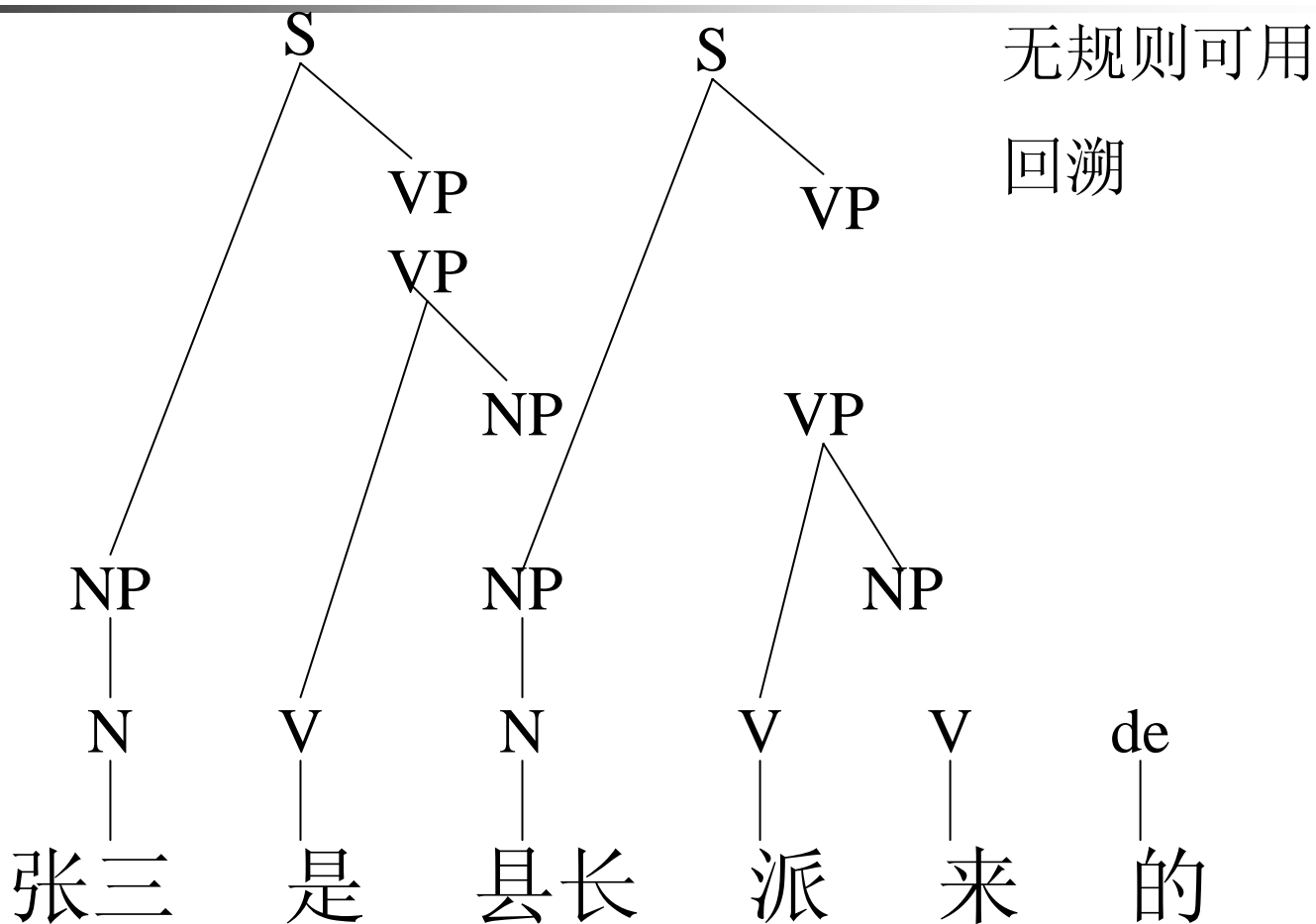
- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

左角分析法示例-38



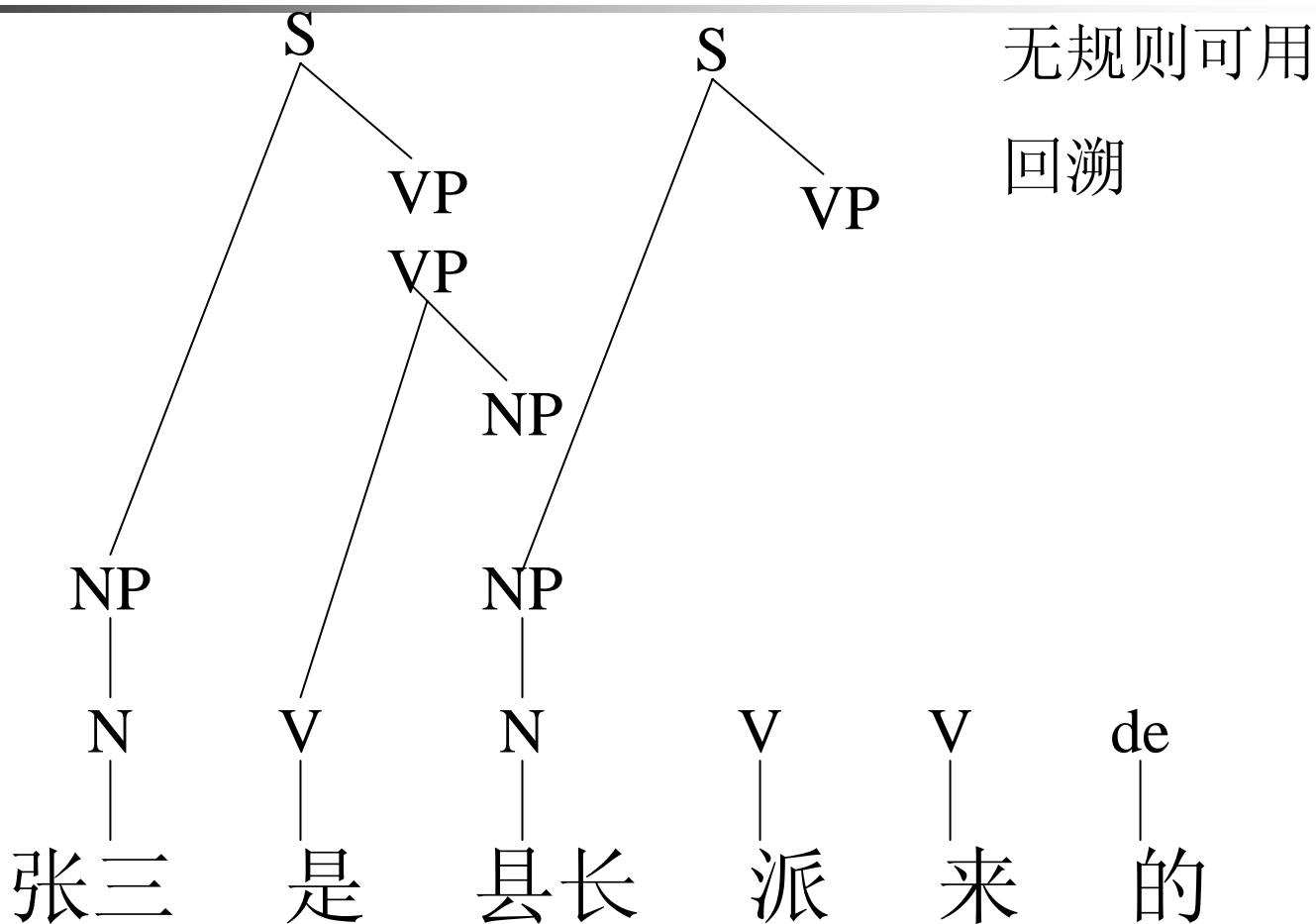
- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

左角分析法示例-39



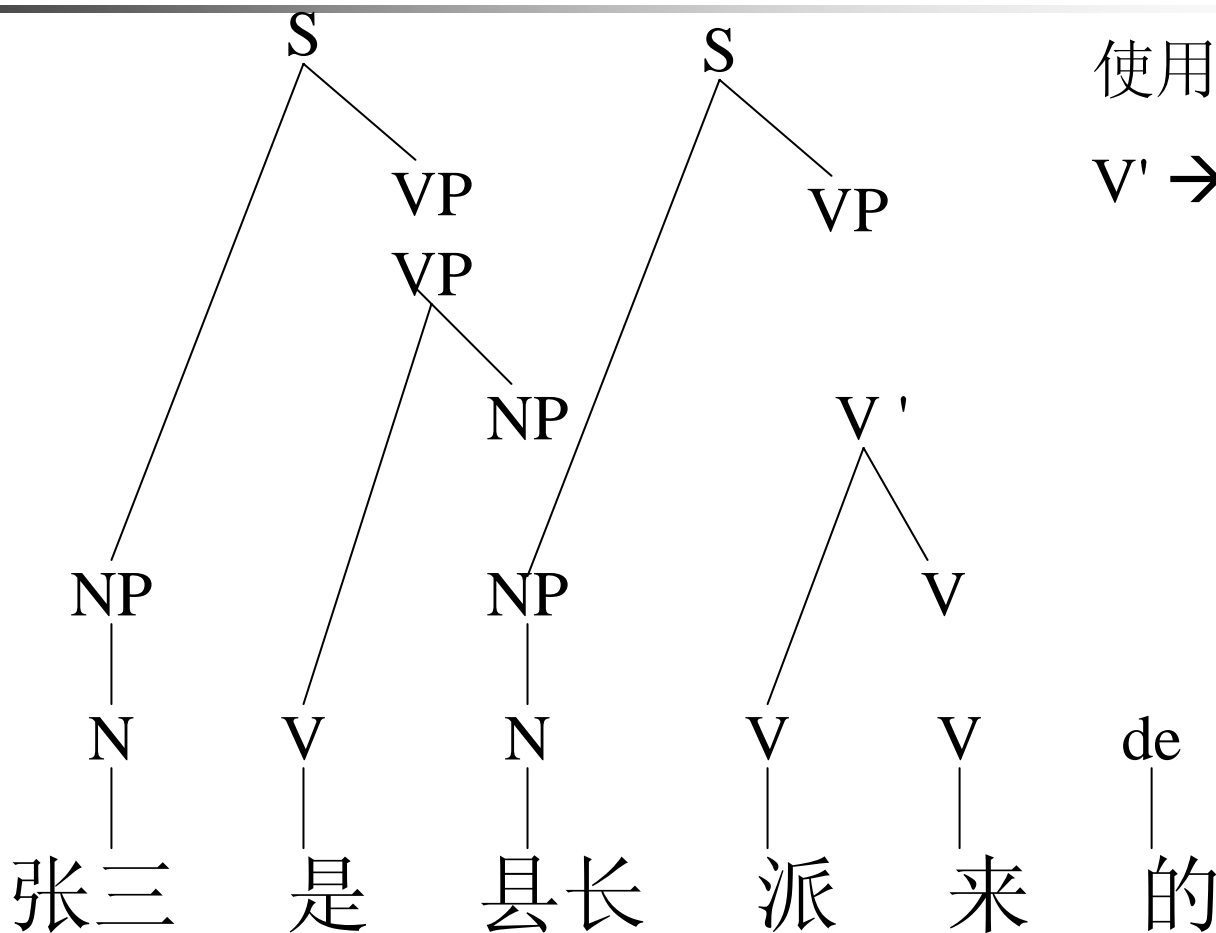
- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

左角分析法示例-40



- (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) $V' \rightarrow V V$

左角分析法示例-41

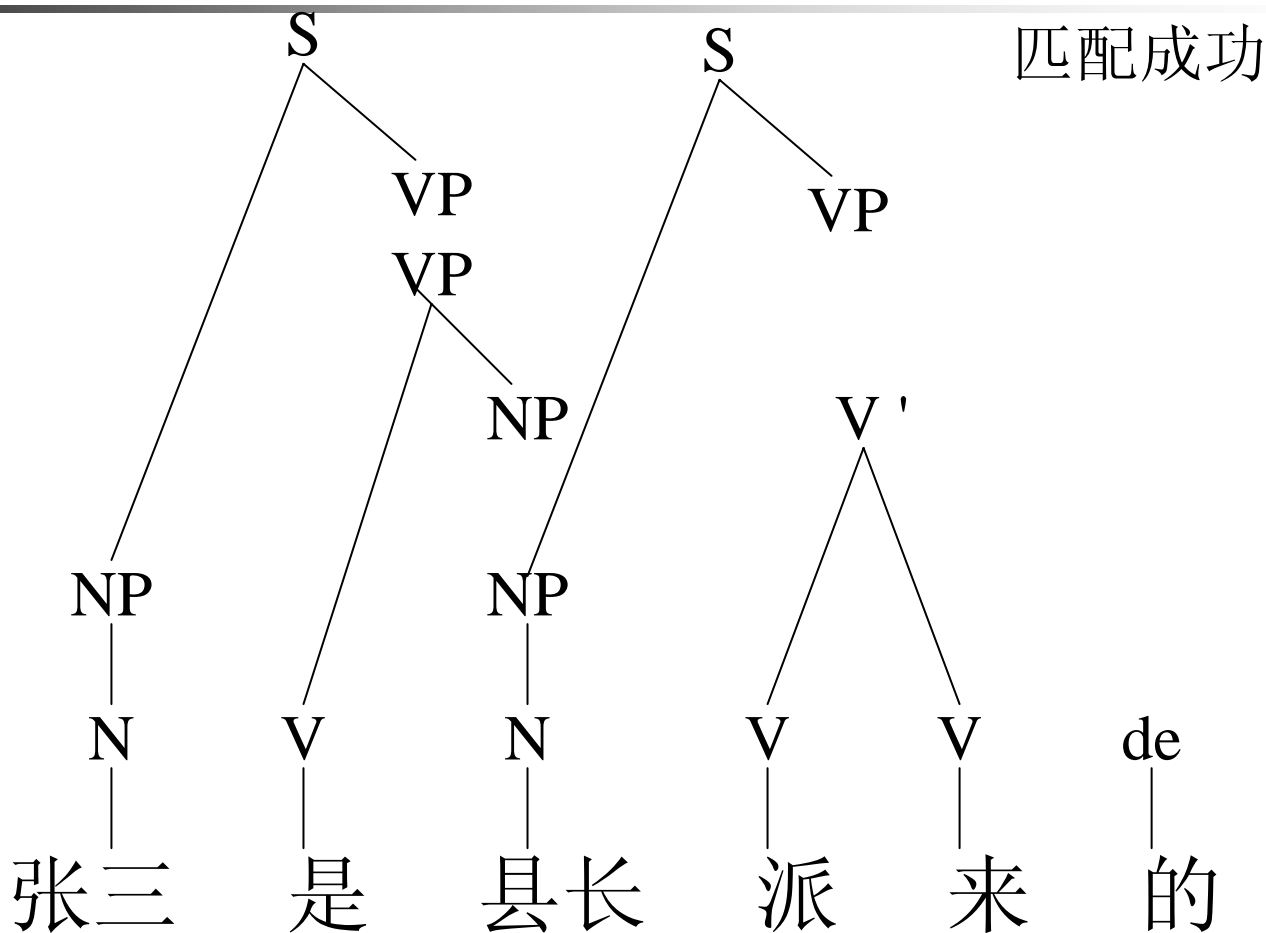


使用规则

$V' \rightarrow V V$

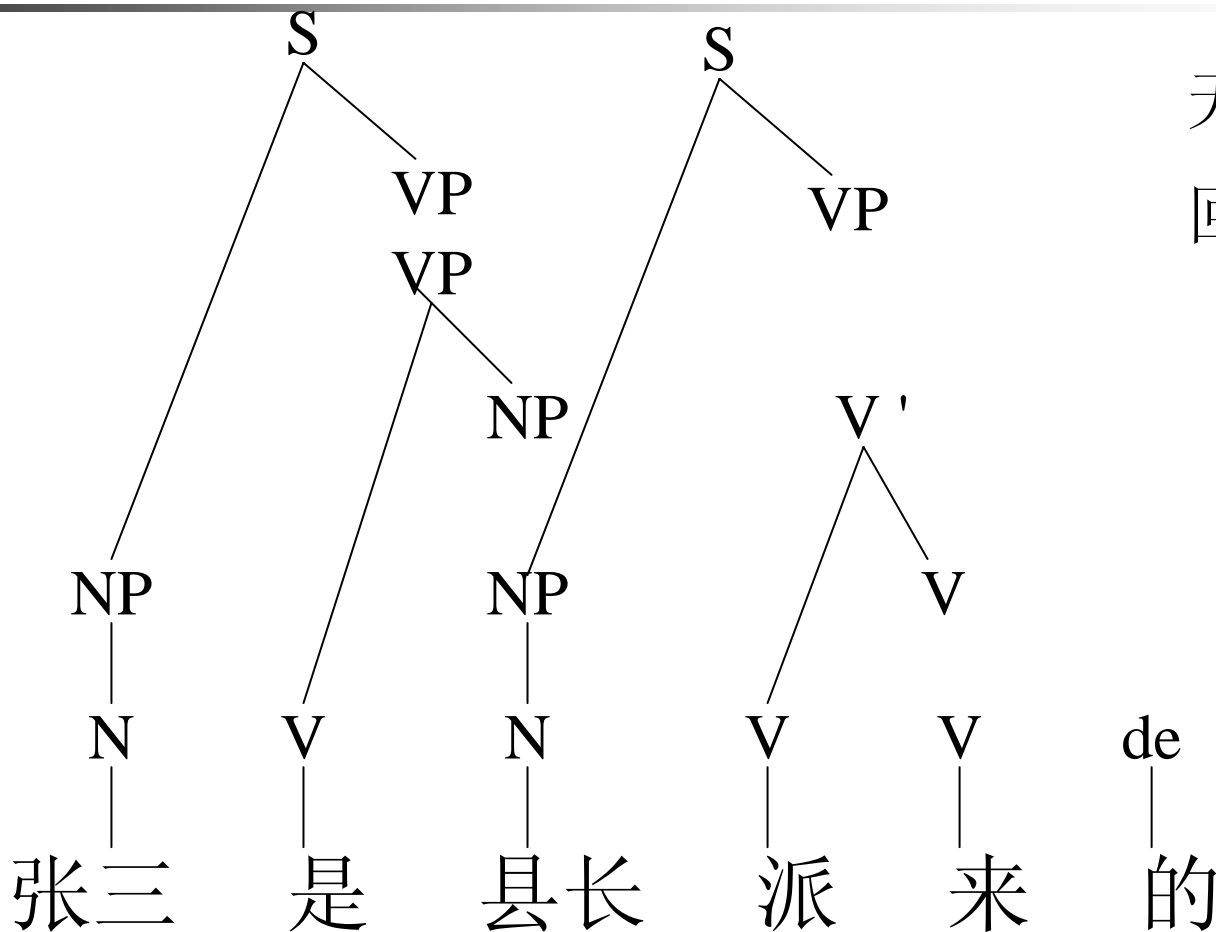
- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

左角分析法示例-42



- (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) $V' \rightarrow V V$

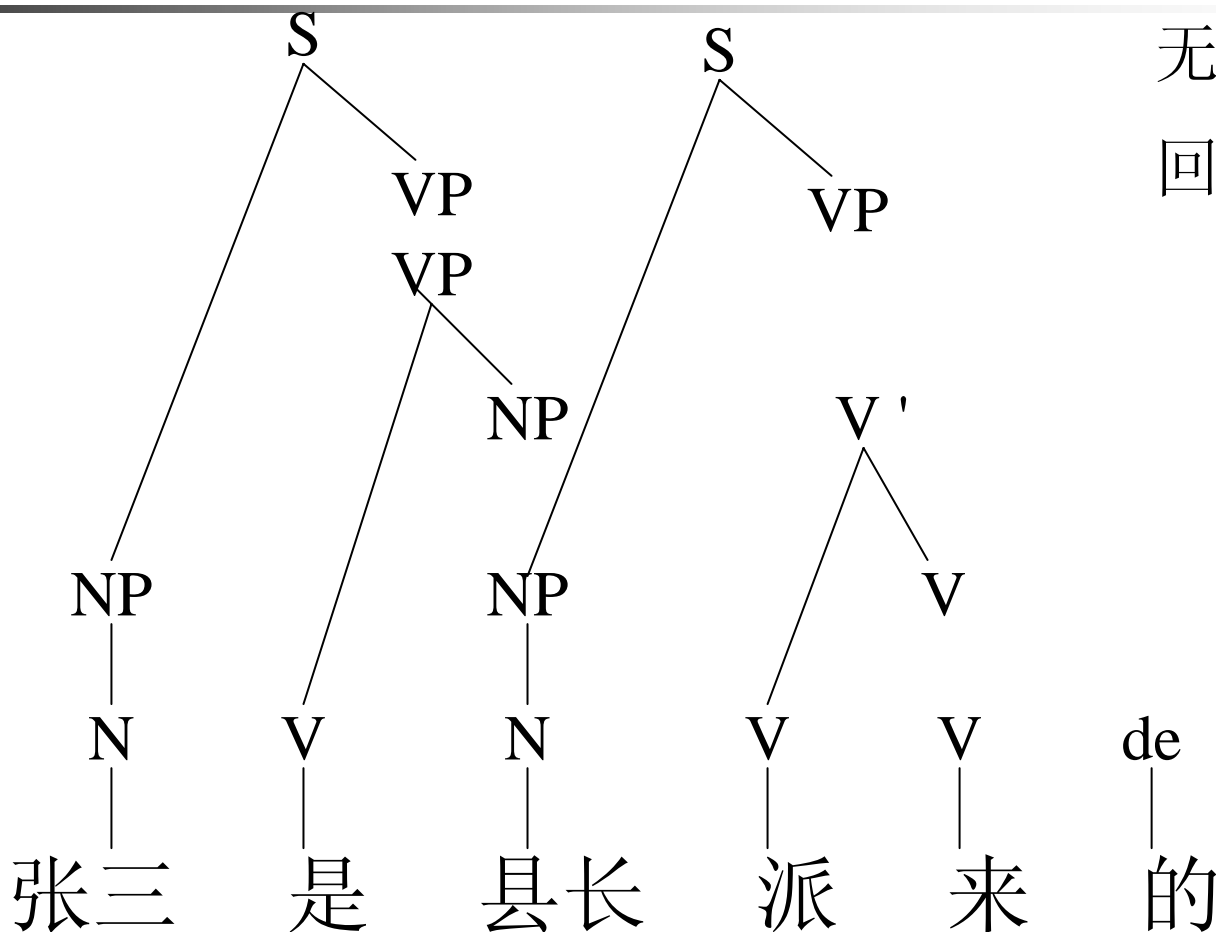
左角分析法示例-43



无规则可用
回溯

- (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) $V' \rightarrow V V$

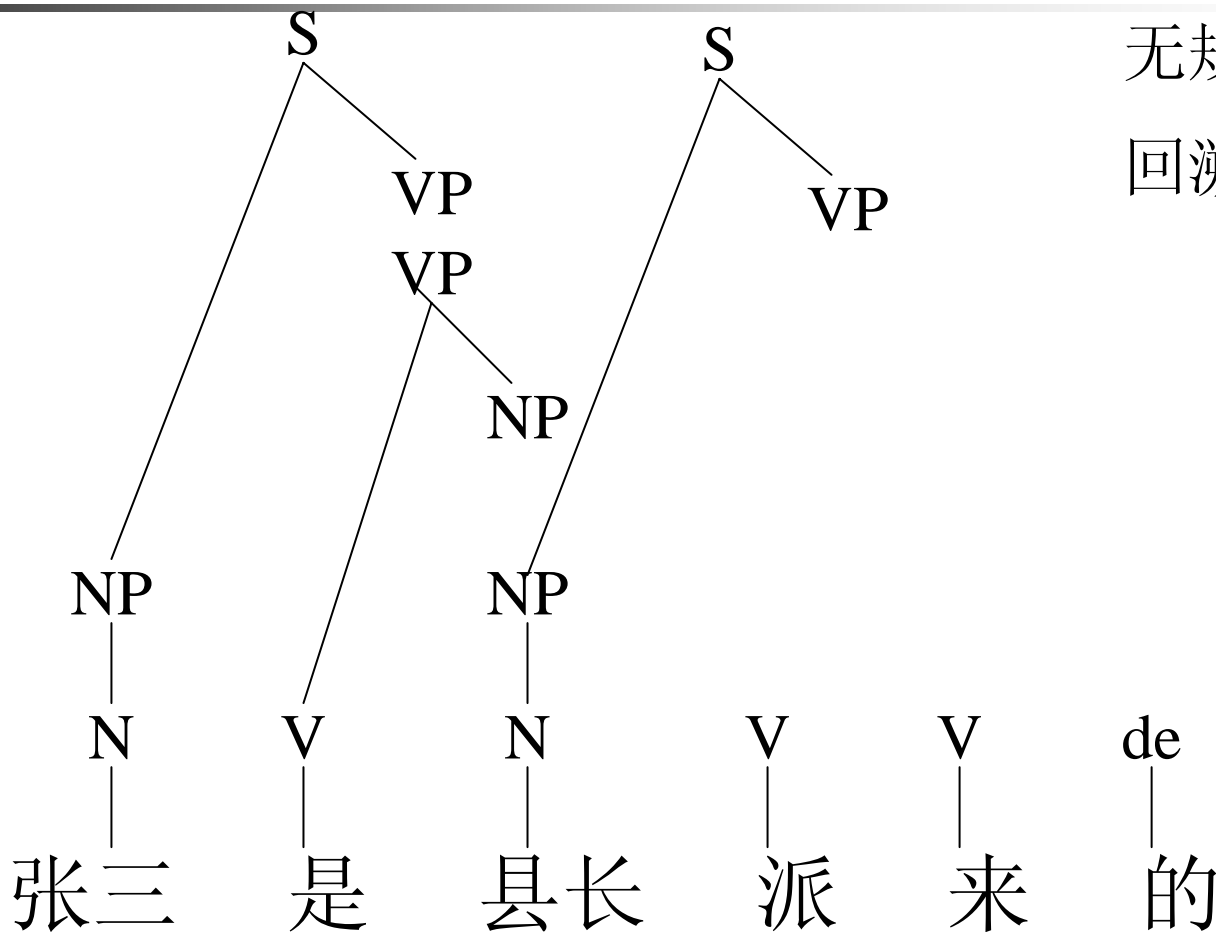
左角分析法示例-45



无规则可用
回溯

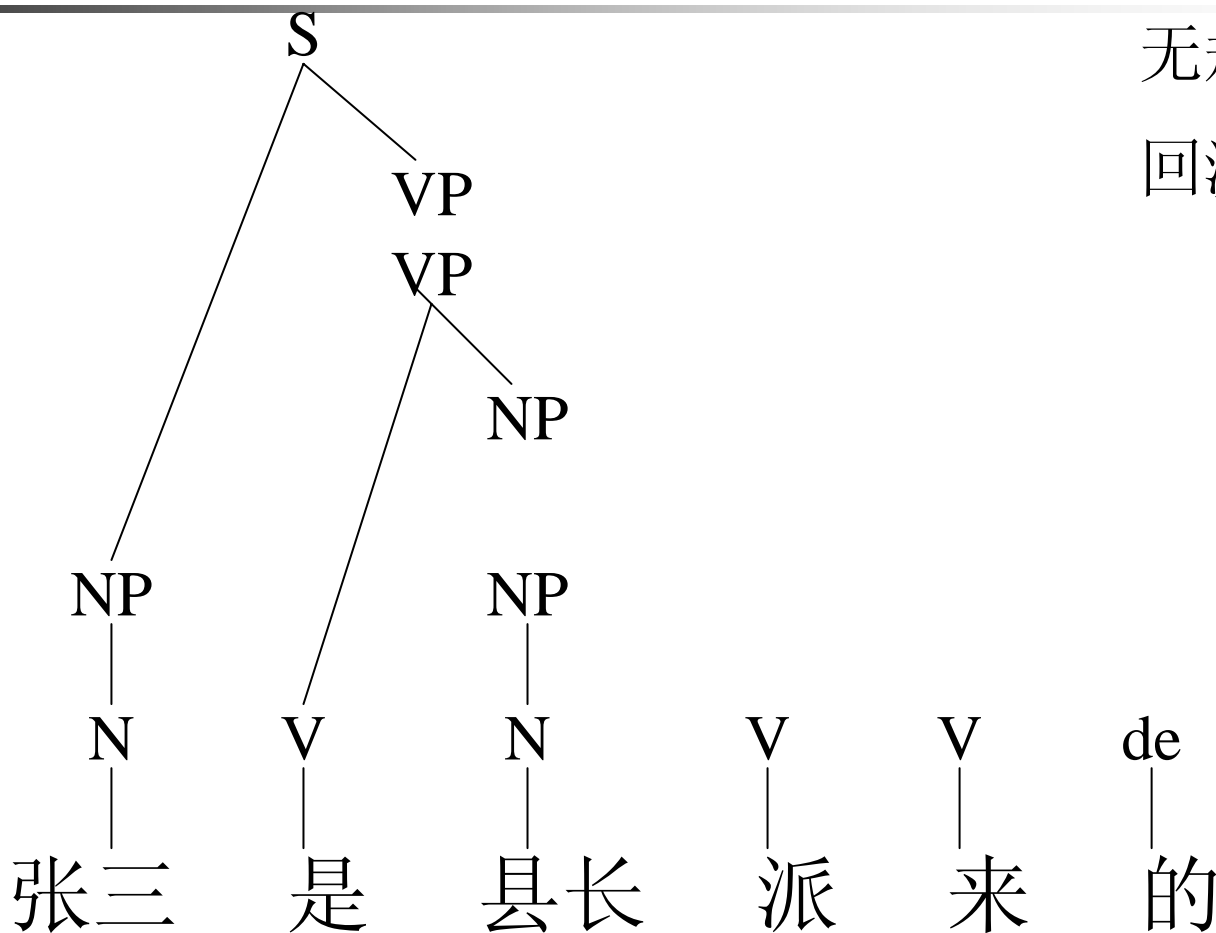
左角分析法示例-46

- (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) $V' \rightarrow V V$



左角分析法示例-47

- (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) $V' \rightarrow V V$

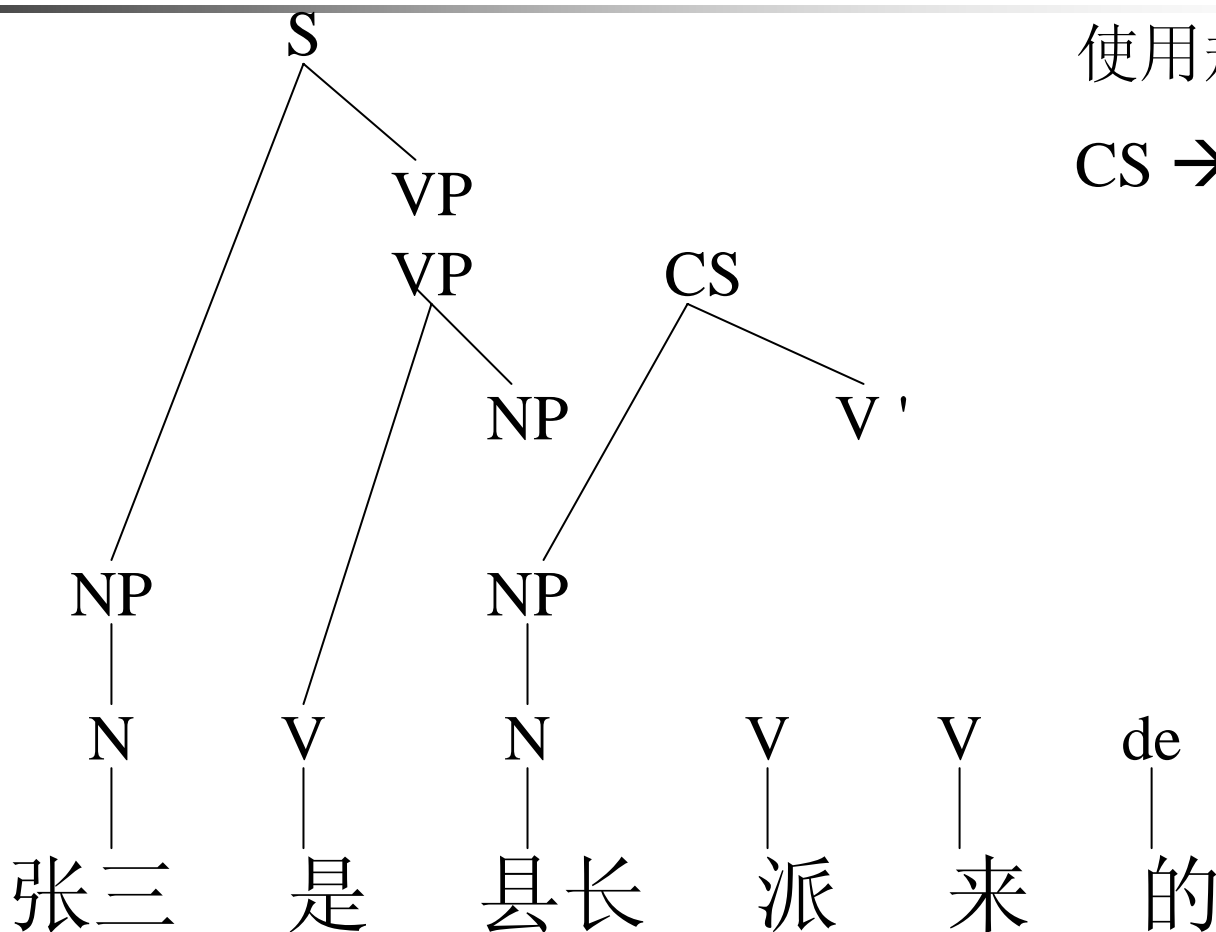


左角分析法示例-48

- (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) $V' \rightarrow V V$

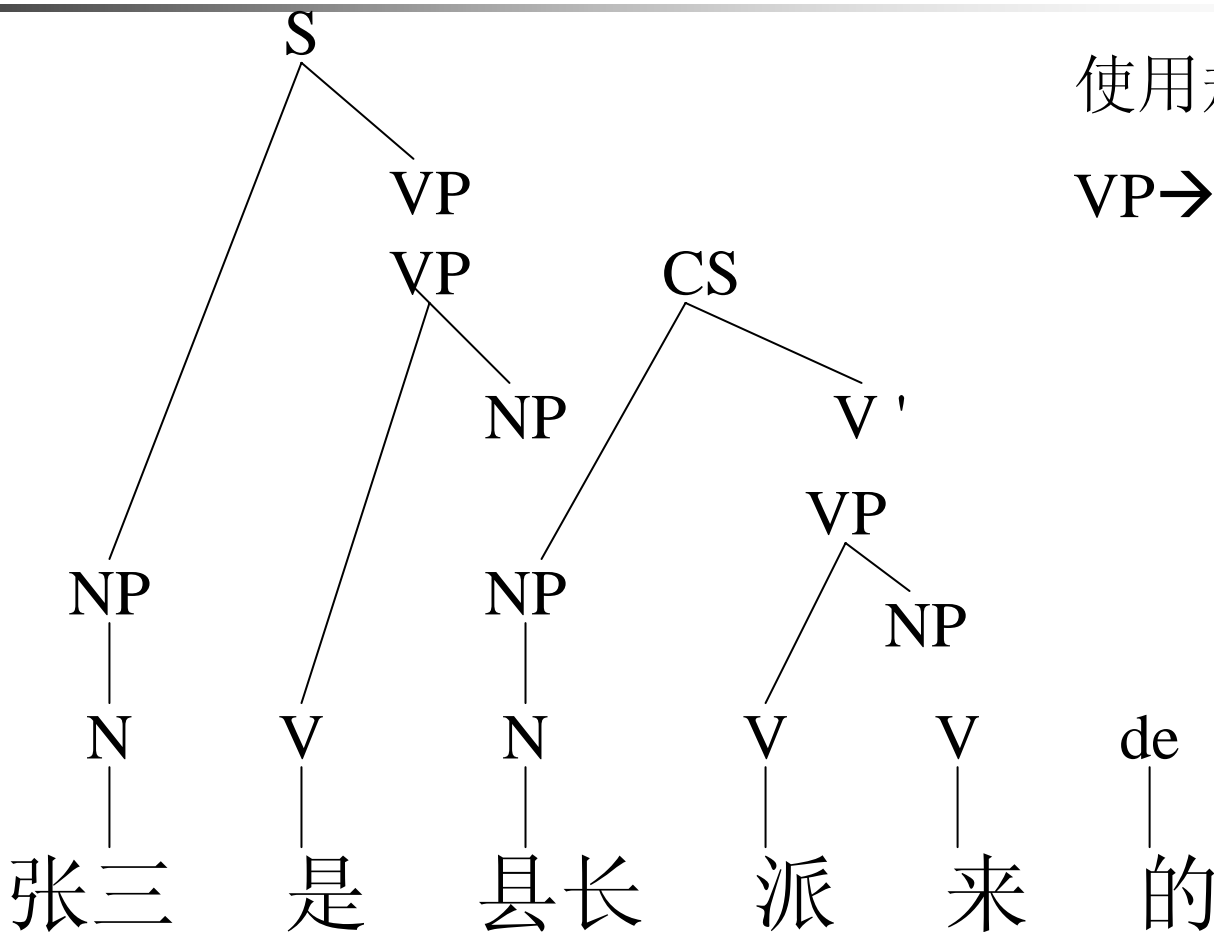
使用规则:

$CS \rightarrow NP V'$



- (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) $V' \rightarrow V V$

左角分析法示例-49

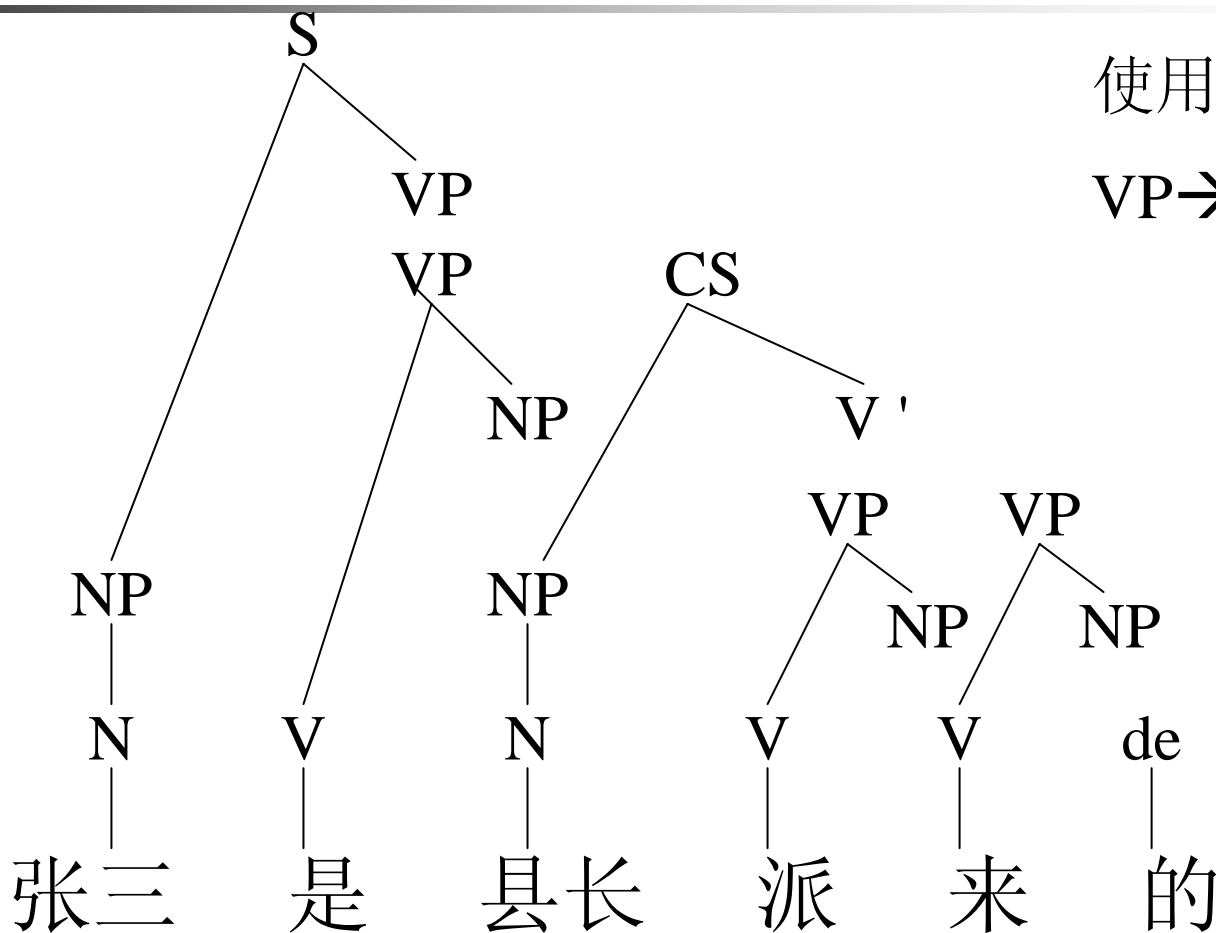


使用规则:

$VP \rightarrow V NP$

- (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) $V' \rightarrow V V$

左角分析法示例-50

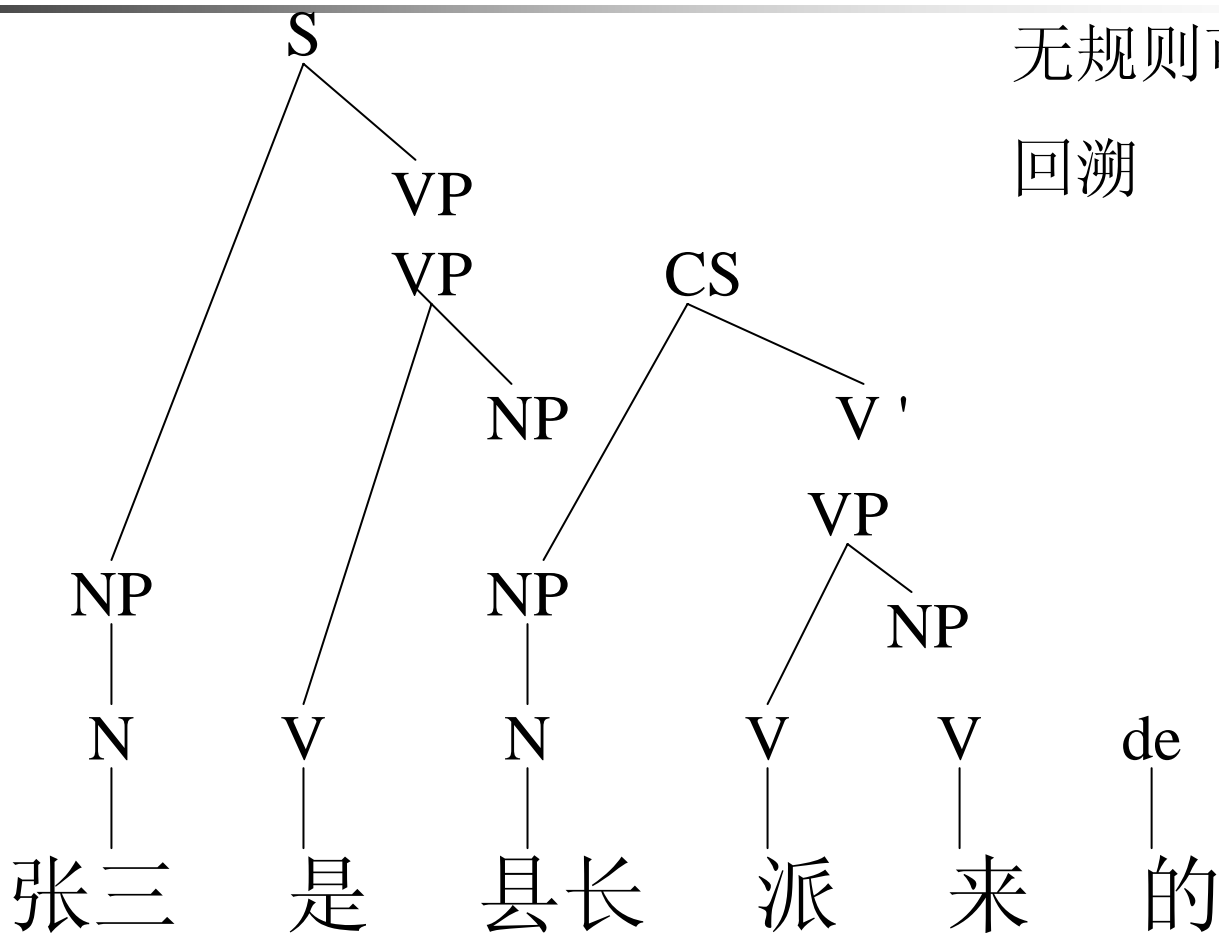


使用规则:

$VP \rightarrow V NP$

- (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) $V' \rightarrow V V$

左角分析法示例-51

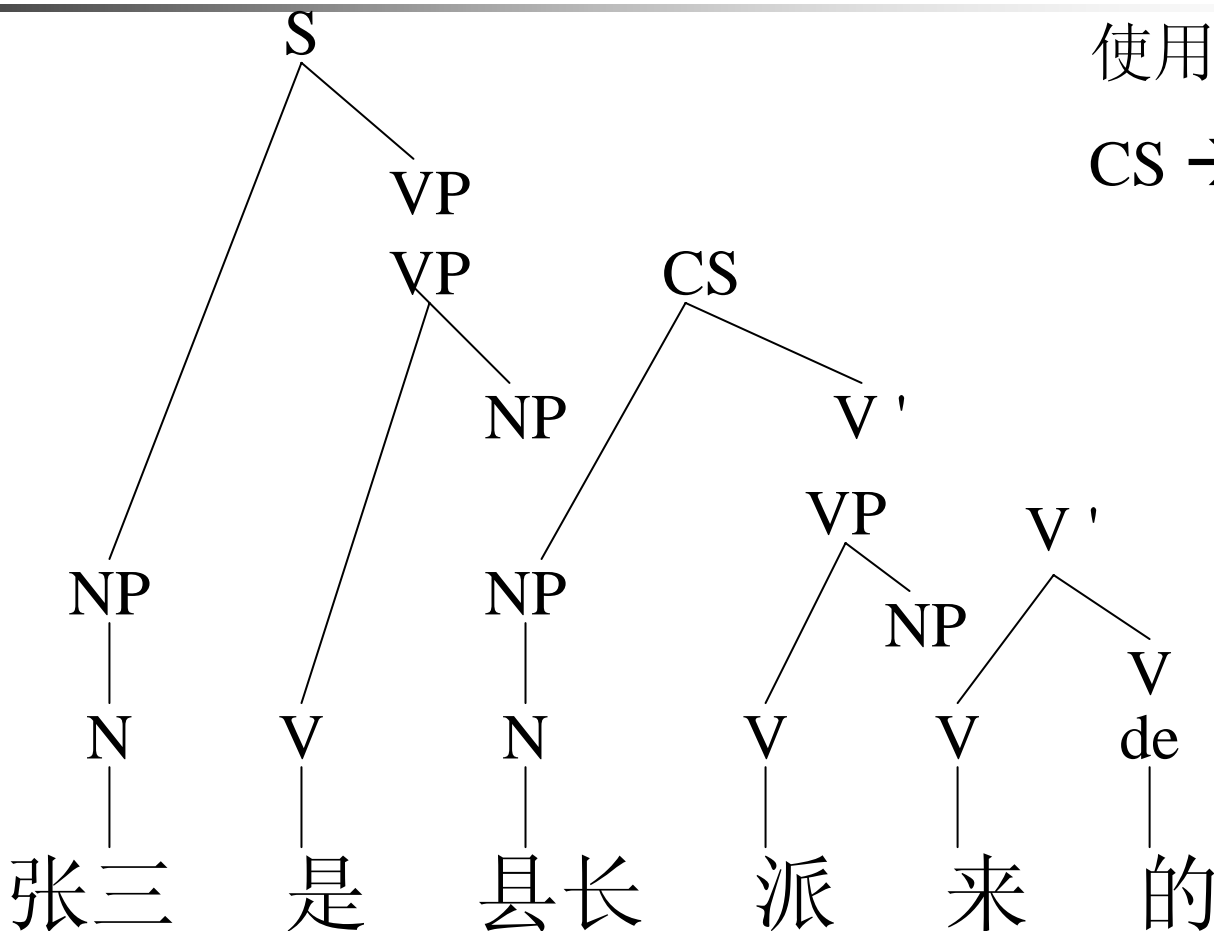


- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

左角分析法示例-52

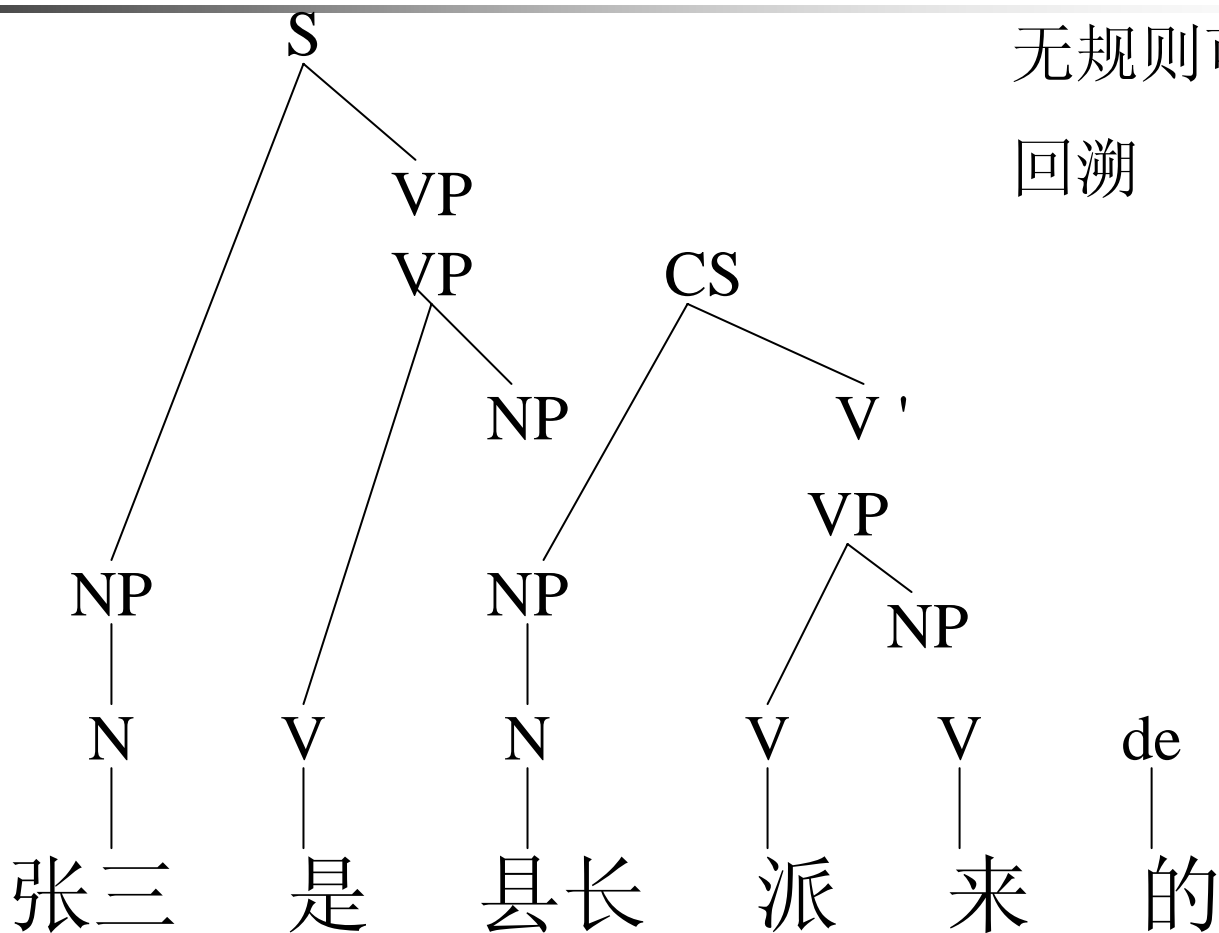
使用规则:

$CS \rightarrow NP V'$



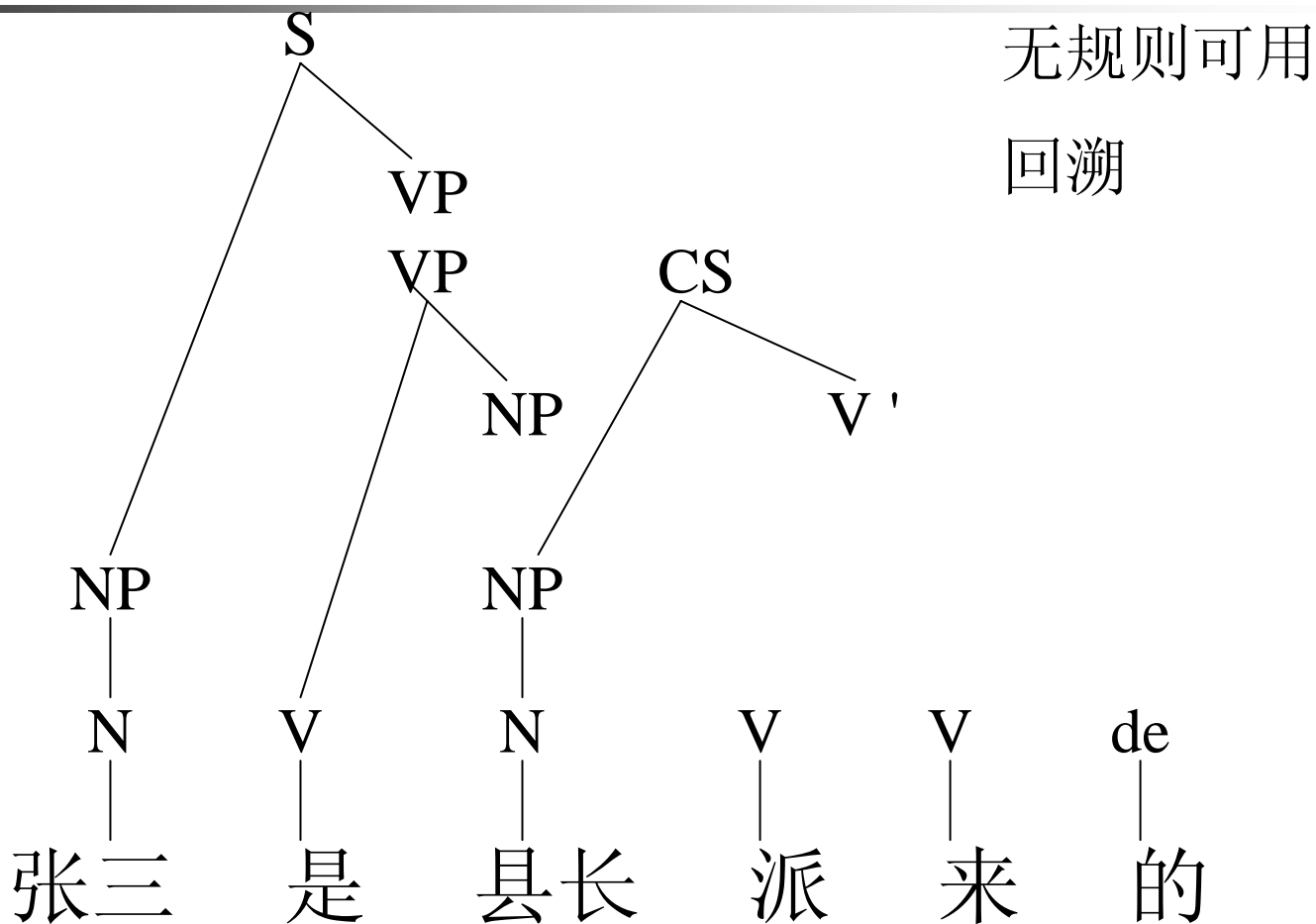
- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

左角分析法示例-53



- (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) $V' \rightarrow V V$

左角分析法示例-54

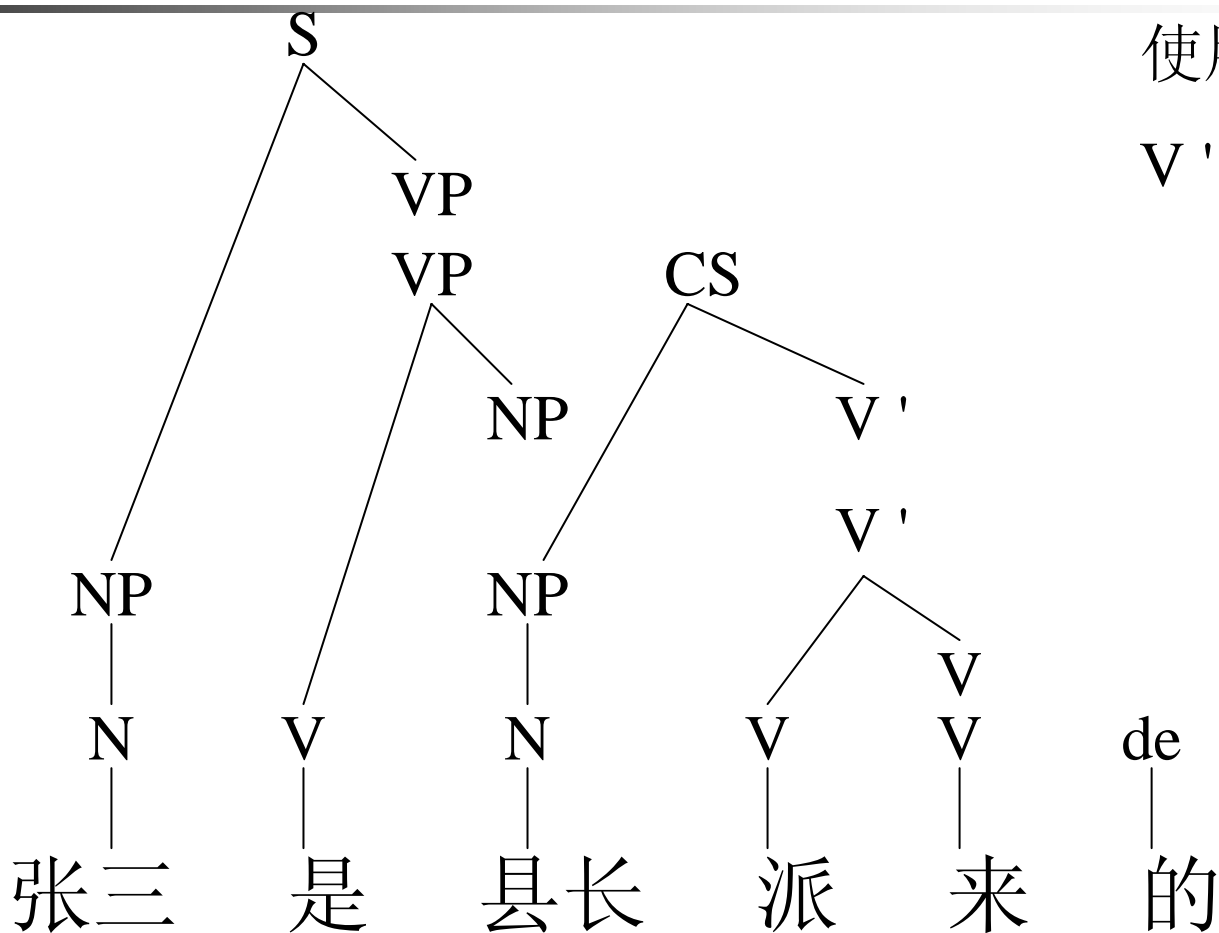


左角分析法示例-55

- (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) $V' \rightarrow V V$

使用规则:

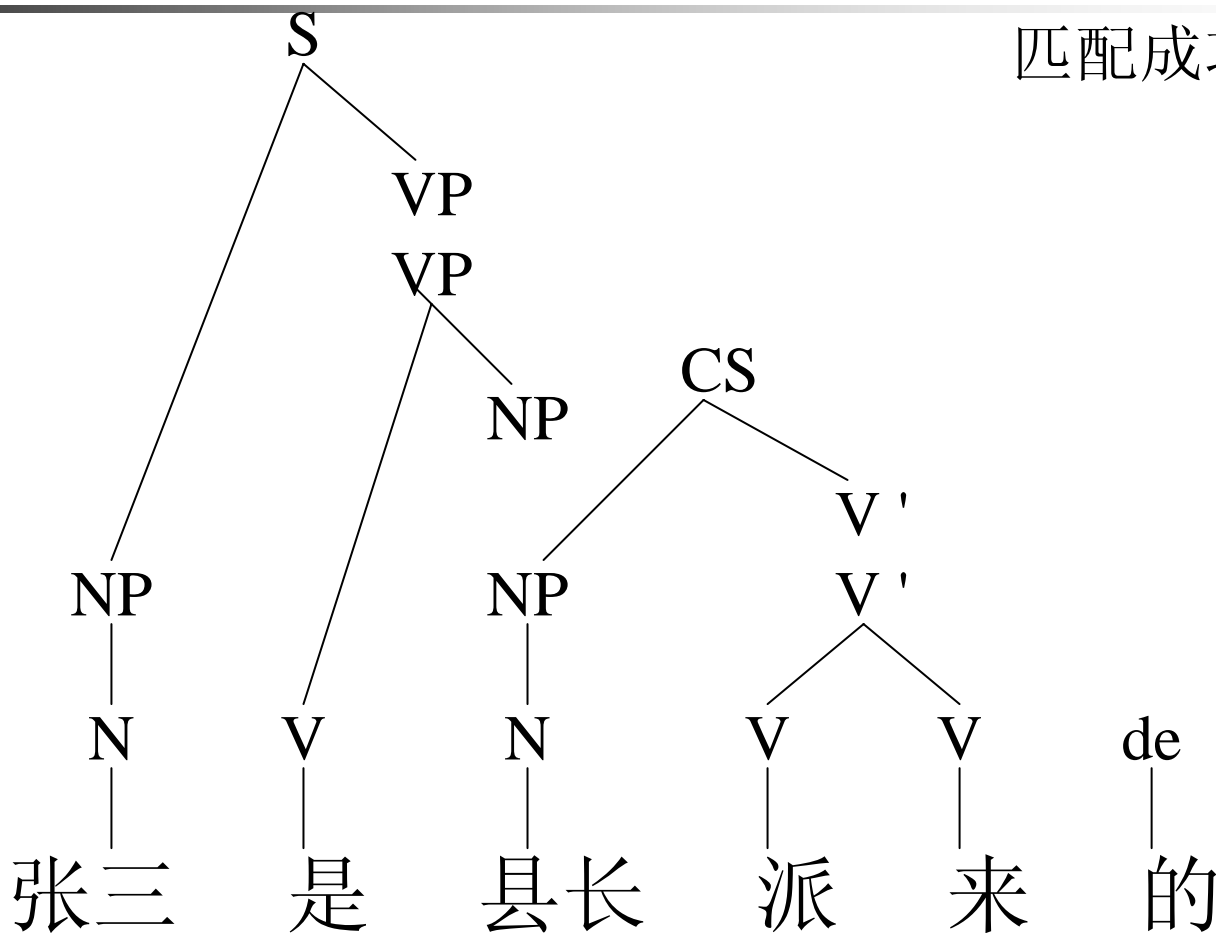
$V' \rightarrow V V$



- (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) $V' \rightarrow V V$

左角分析法示例-56

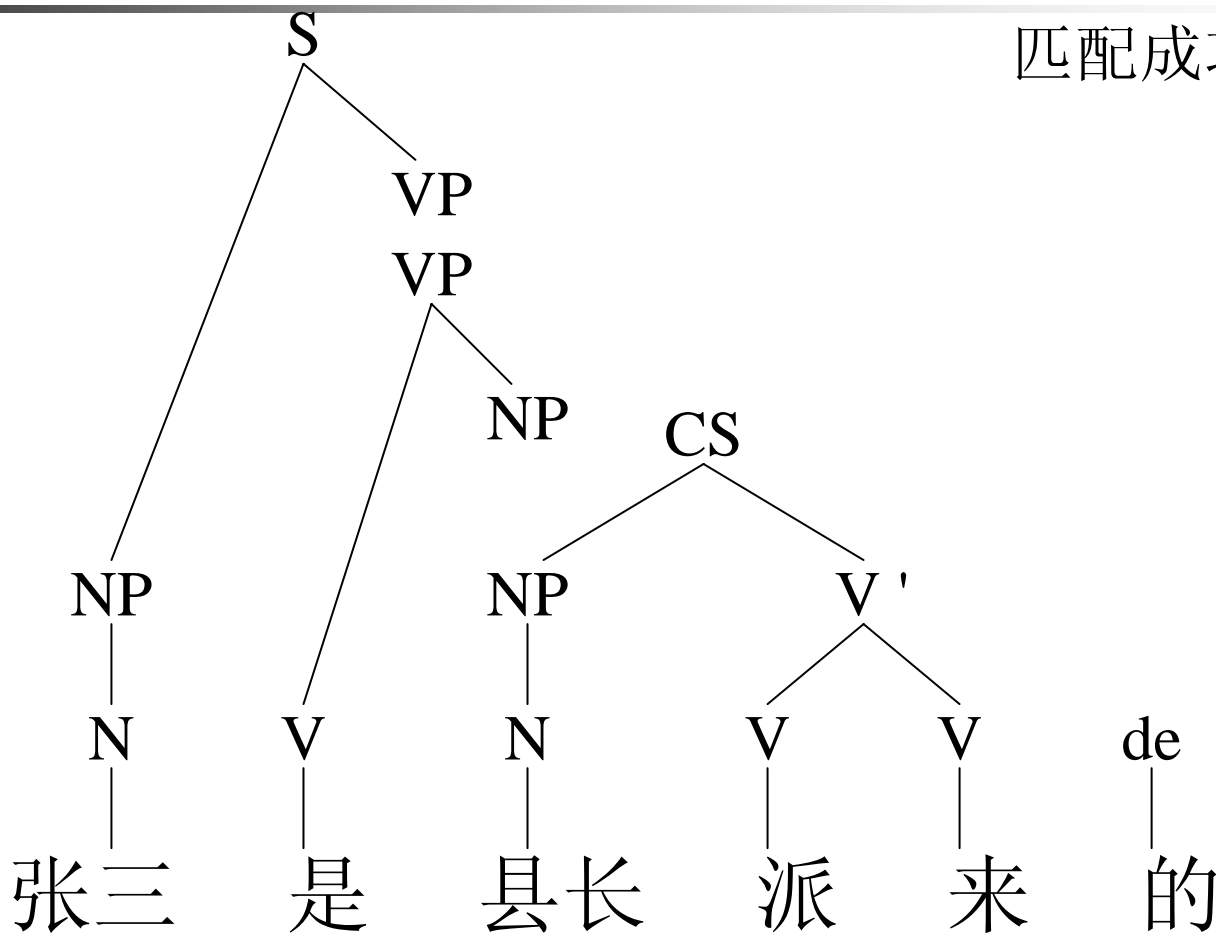
匹配成功



- (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) $V' \rightarrow V V$

左角分析法示例-57

匹配成功

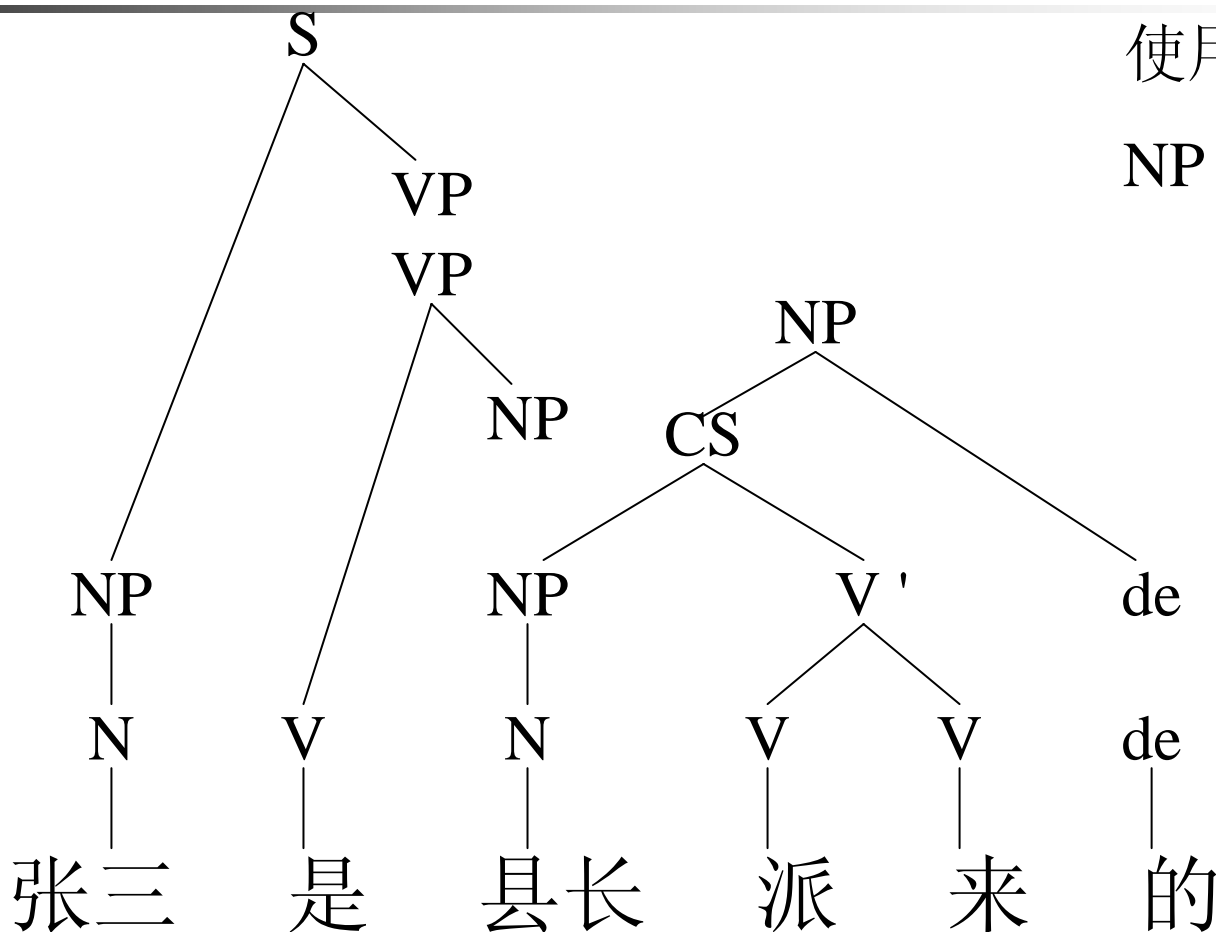


- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

左角分析法示例-58

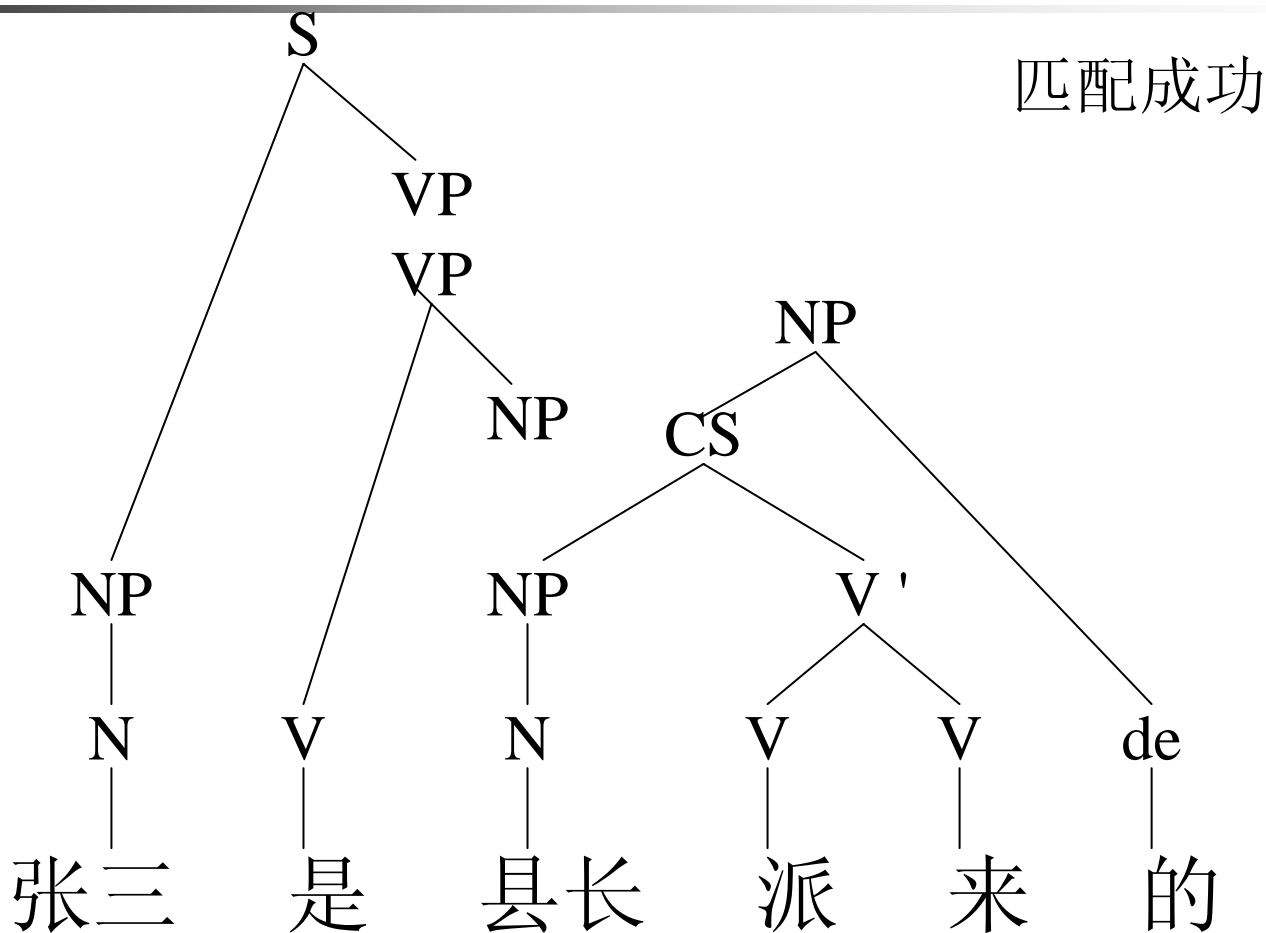
使用规则:

$NP \rightarrow CS$ de



- (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) $V' \rightarrow V V$

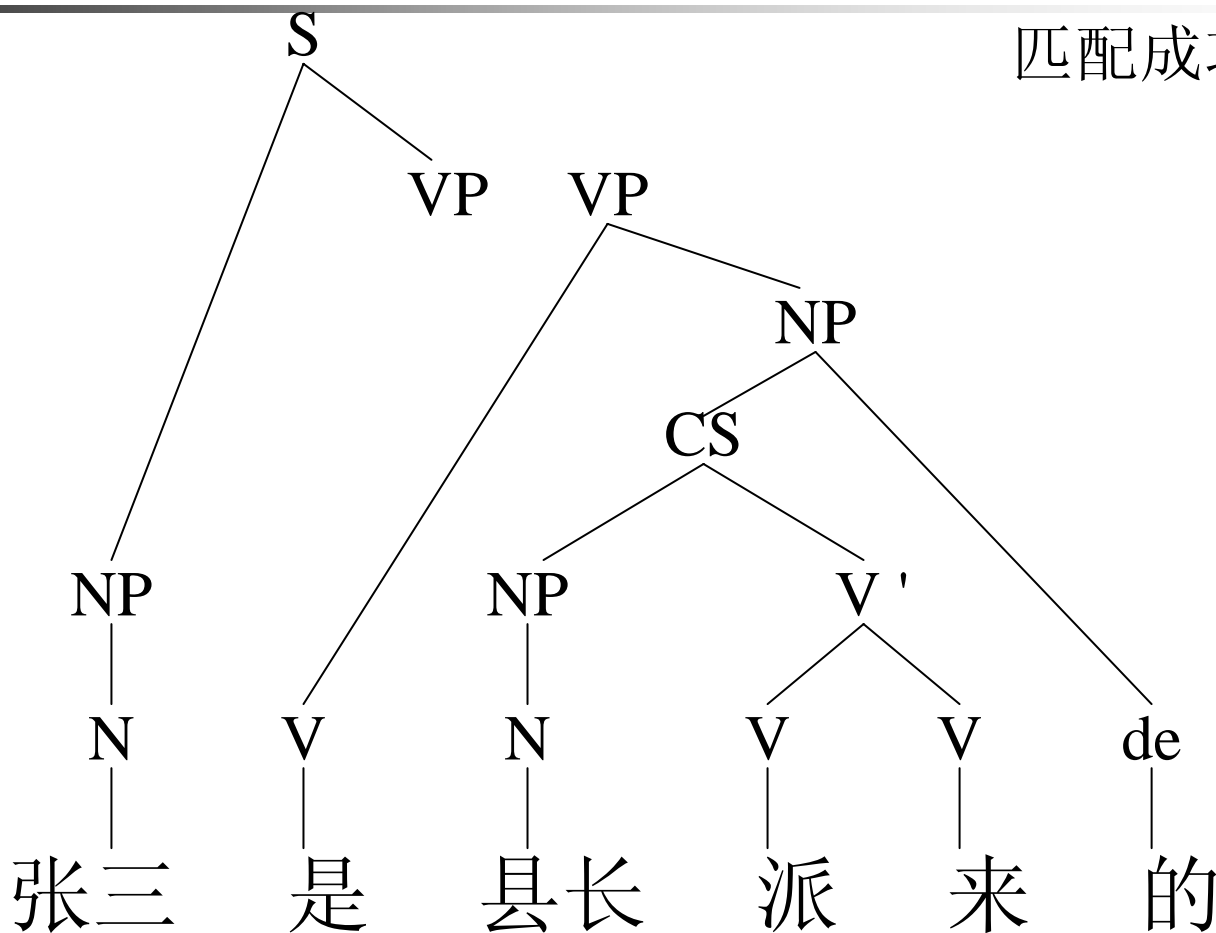
左角分析法示例-59



- (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) $V' \rightarrow V V$

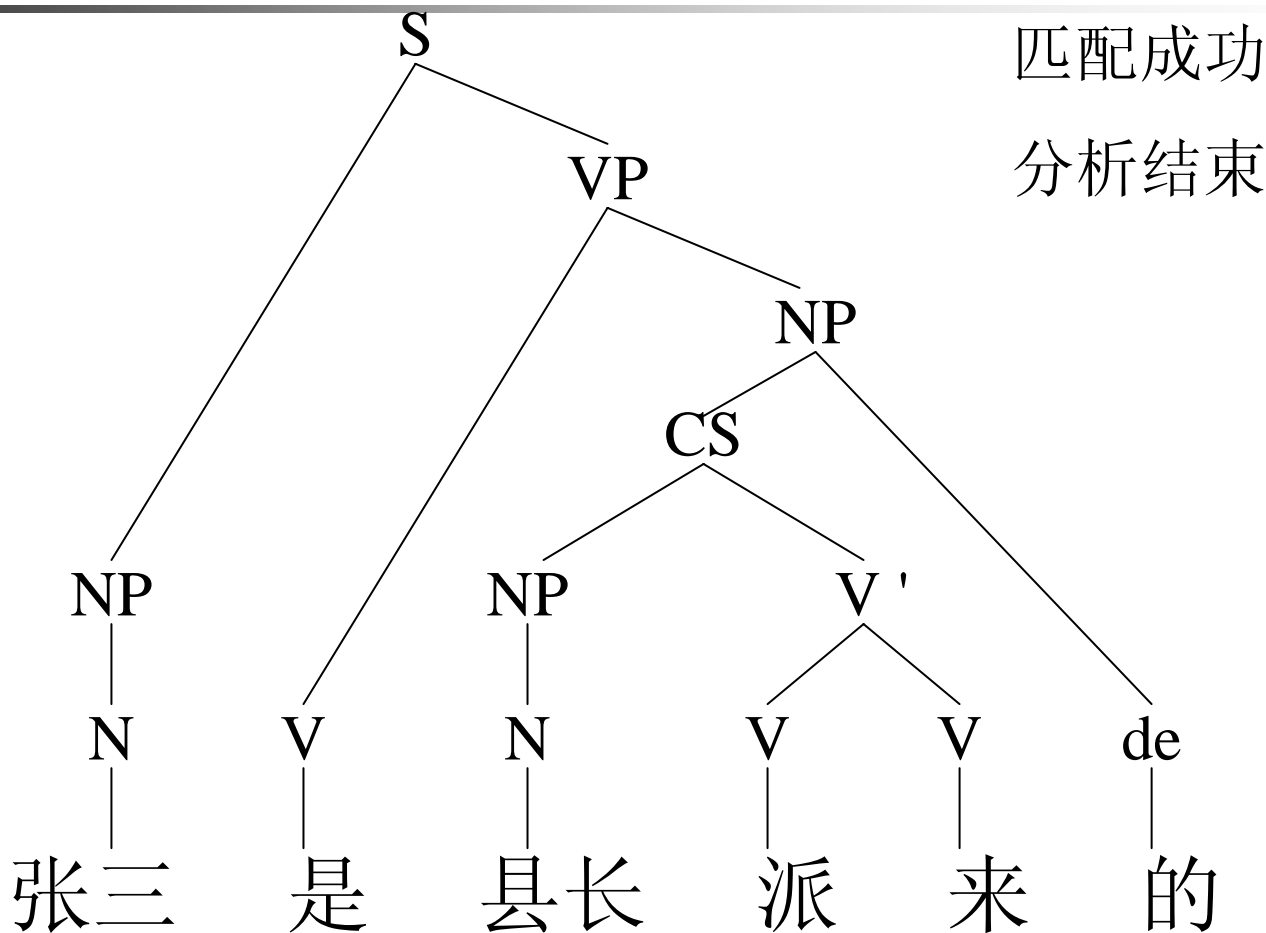
左角分析法示例-60

匹配成功



- (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) $V' \rightarrow V V$

左角分析法示例-61





关于左角分析法的思考

回溯过多，对这个例子来说，是由于深度优先，一开始走错了，以后步步皆错，要回溯到一开始的地方，才能重新走回到正确的道路上，分析效率很低。



4 CYK算法

- CYK算法: Cocke-Younger-Kasami算法
- CYK算法是一种并行算法, 不需要回溯;
- CYK算法建立在Chomsky范式的基础上
 - Chomsky范式的规则只有两种形式: $A \rightarrow BC$ $A \rightarrow x$
这里A,B,C是非终结符, x是终结符
 - 由于后一种形式实际上就是词典信息, 在句法分析之前已经进行了替换, 所以在分析中我们只考虑形如 $A \rightarrow BC$ 形式的规则
 - 由于任何一个上下文无关语法都可以转化成符合Chomsky范式的语法, 因此CYK算法可以应用于任何一个上下文无关语法

CYK算法分析结果示意

跨 度	6	S					
	5		VP				
	4			NP			
	3	S		CS			
	2		VP		V'		
	1	NP,N	V	NP,N	V	V	de
	1	2	3	4	5	6	位置
	张三	是	县长	派	来	的	



CYK算法—数据结构

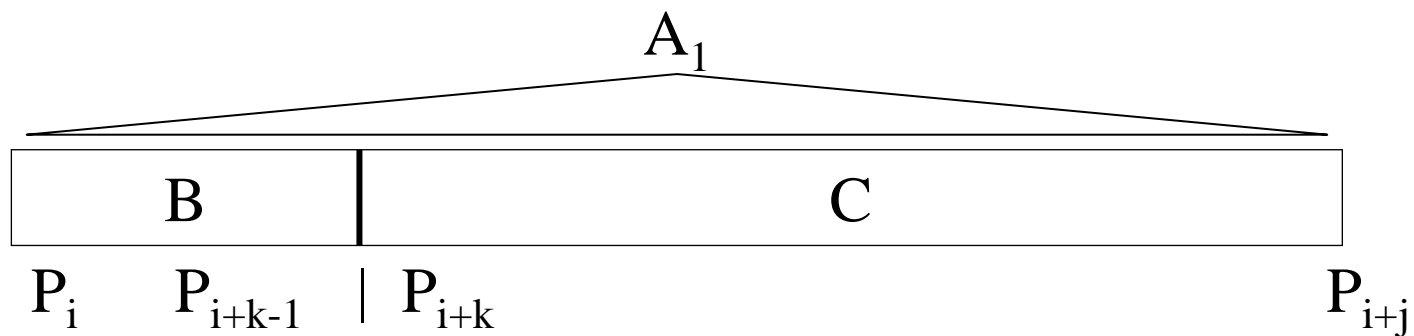
- 一个二维矩阵：{ $P(i, j)$ }
 - 每一个元素 $P(i, j)$ 对应于输入句子中某一个跨度（Span）上所有可能形成的短语的非终结符的集合
 - 横坐标 i ：该跨度左侧第一个词的位置
 - 纵坐标 j ：该跨度包含的词数
- $P(3,1)=\{NP,N\}$ 表示“县长”可以归约成N和NP
3表示从第3个词位置开始，1表示NP,N覆盖1个词。
- $P(3,3)=\{CS\}$ 表示“县长+派+来”可以规约成CS
3表示从第3个词位置开始，3表示CS覆盖3个词

CYK算法：算法描述

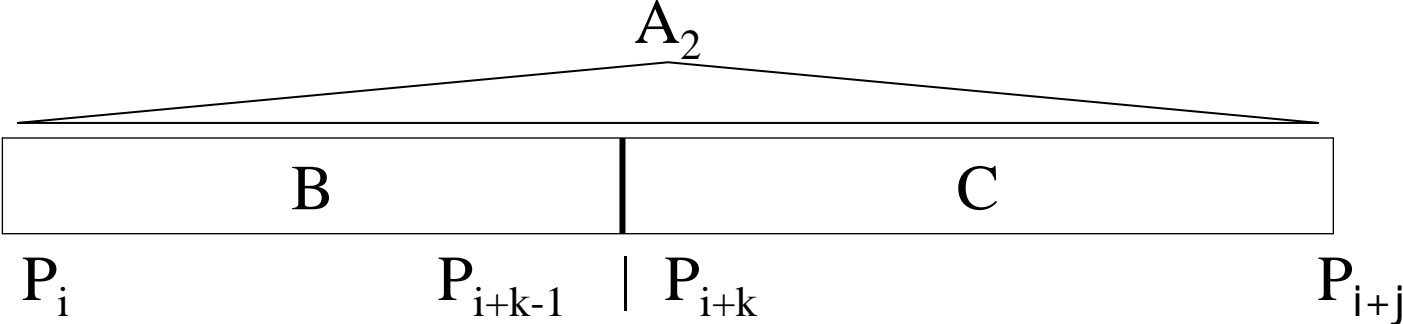
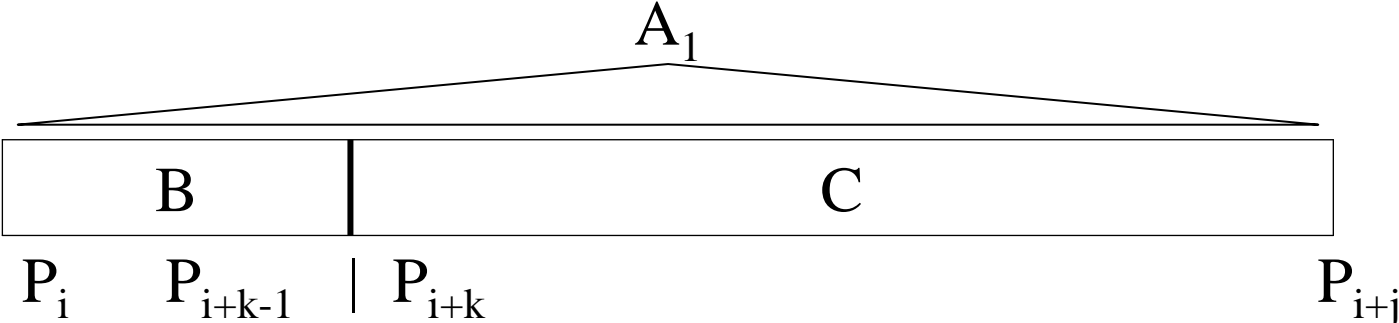
i 从1到 n 的循环

1. 对 $i = 1 \dots n$, $j = 1$ (填写第一行, 跨度为1)
对于每一条规则 $A \rightarrow W_i$,
将非终结符 A 加入集合 $P(i, j)$;
2. 对 $j = 2 \dots n$ (填写第2到 n 行, 跨度为 j)
对 $i = 1 \dots n - j + 1$ (对于所有起点 i)
对 $k = 1 \dots j - 1$ (对于一个跨度内所有分割点 k)
对每一条规则 $A \rightarrow BC$,
如果 $B \in P(i, k)$ 且 $C \in P(i + k, j - k)$
那么将非终结符 A 加入集合 $P(i, j)$
3. 如果 $S \in P(1, n)$, 那么分析成功, 否则分析失败

CYK算法成分组合示意



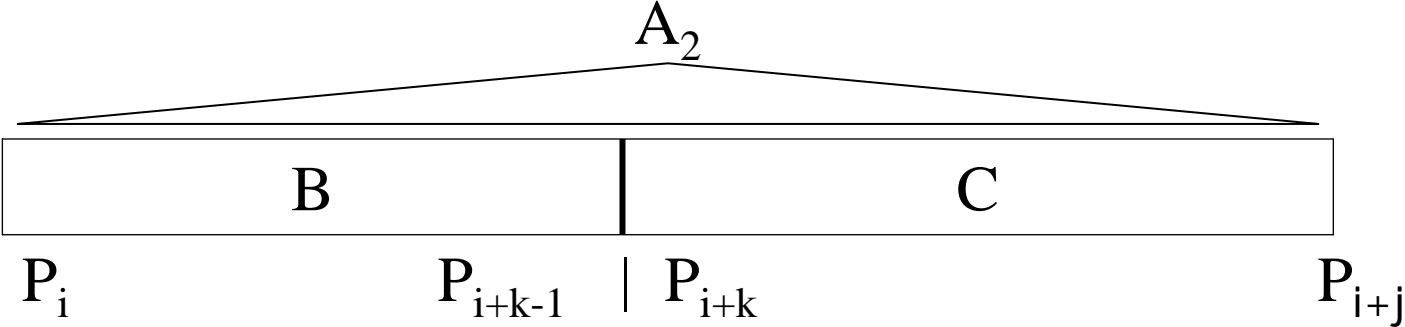
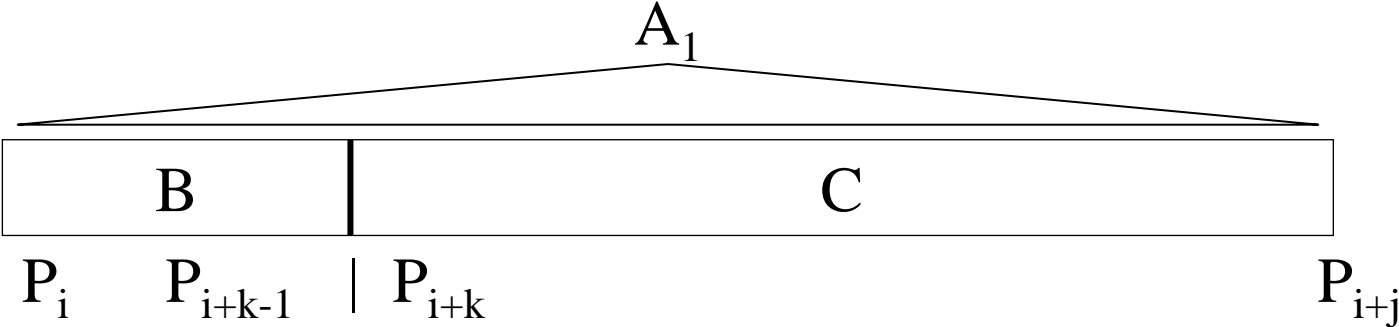
CYK算法成分组合示意



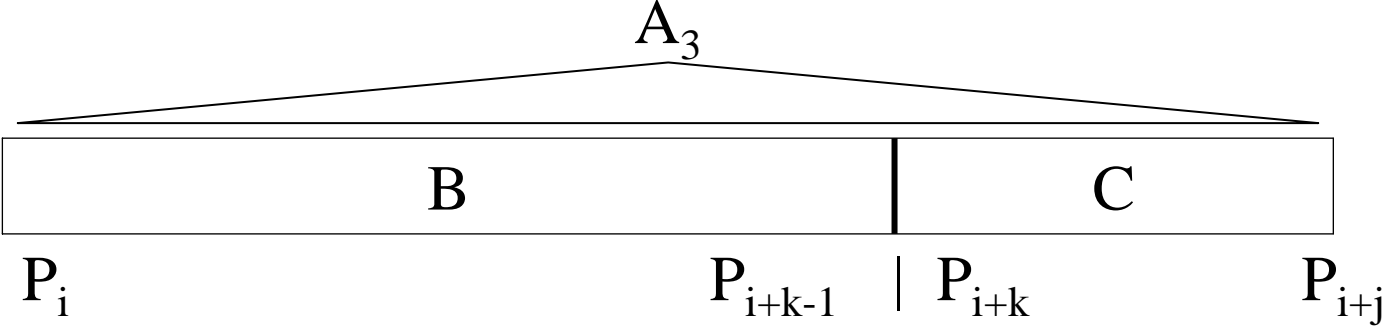
.....

CYK算法成分组合示意

k是在以i为起点，j跨度内进行短语组成成分分割的一个刻度位置

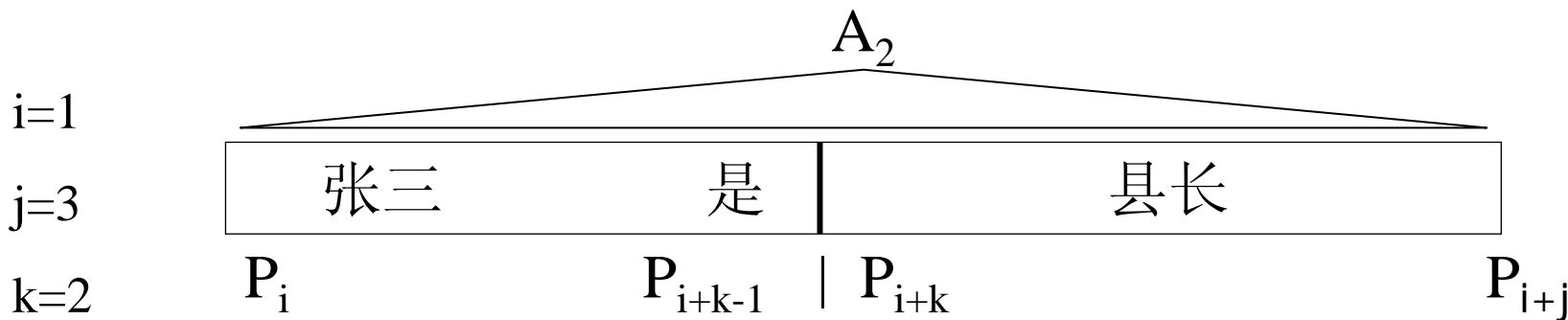
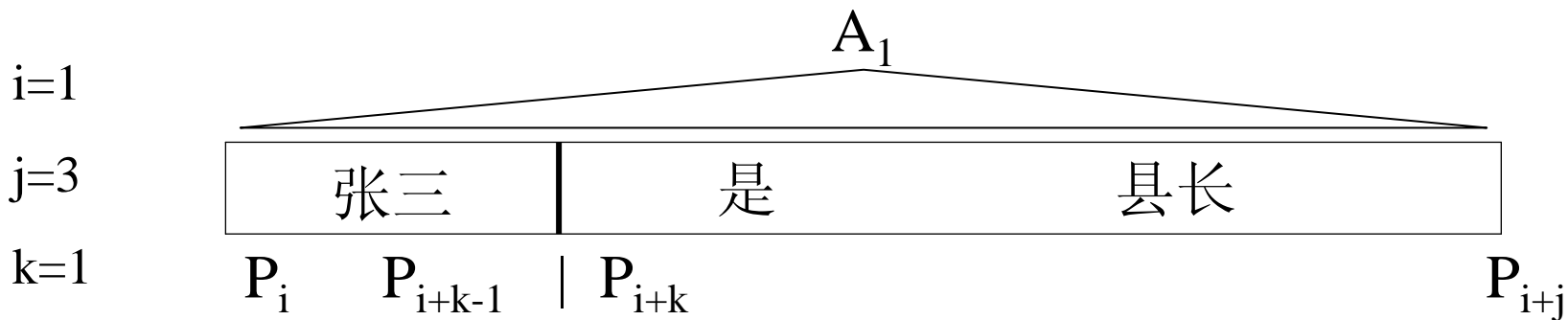


.....



CYK算法成分组合示意

k是在以i为起点，j跨度内进行短语组成成分分割的一个刻度位置



- (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) $V' \rightarrow V V$

CYK算法示例-填写第1行

跨 度	6						
	5						
	4						
	3						
	2						
	1	N	V	N	V	V	de
	1	2	3	4	5	6	位置
	张三	是	县长	派	来	的	

CYK算法示例-填写第1行 (续)

- (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) $V' \rightarrow V V$

跨度

6						
5						
4						
3						
2						
1	NP,N	V	N	V	V	de

$j=1, i=1, k=1$

有规则 $NP \rightarrow N$

因此将NP加入P(1,1)

1

2

3

4

5

6

位置

张三

是

县长

派

来

的

CYK算法示例-填写第1行 (续)

- (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) $V' \rightarrow V V$

跨 度	6						
	5						
	4						
	3						
	2						
	1	NP,N	V	NP,N	V	V	de
	1	2	3	4	5	6	位置
	张三	是	县长	派	来	的	

$j=1, i=3, k=1$
 有规则 $NP \rightarrow N$
 因此将NP加入P(3,1)

- (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) $V' \rightarrow V V$

CYK算法示例-填写第2行

跨
度

6						
5						
4						
3						
2						
1	NP,N	V	NP,N	V	V	de

$j = 2, i = 1, k = 1$
 $P(i,k) = P(1,1) = NP, N$
 $P(i+k, j-k) = P(2,1) = V$
 $\Rightarrow P(i,j) = P(1,2) = \text{“”}$

1 2 3 4 5 6 位置

张三 是 县长 派 来 的

CYK算法示例-填写第2行 (续)

- (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) $V' \rightarrow V V$

跨度

6						
5						
4						
3						
2		VP				
1	NP,N	V	NP,N	V	V	de

$j = 2, i = 2, k = 1$
 $P(i,k) = P(2,1) = V$
 $P(i+k,j-k) = P(3,1) = NP$
 $\Rightarrow P(i,j) = P(2,2) = VP$

1 2 3 4 5 6 位置

张三 是 县长 派 来 的

CYK算法示例-填写第2行 (续)

- (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) $V' \rightarrow V V$

跨度

6						
5						
4						
3						
2		VP		V'		
1	NP,N	V	NP,N	V	V	de

$$j = 2, i = 4, k = 1$$

$$P(i,k) = P(4,1) = V$$

$$P(i+k, j-k) = P(5,1) = V$$

$$\Rightarrow P(i,j) = P(4,2) = V'$$

1

2

3

4

5

6

位置

张三

是

县长

派

来

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) $V' \rightarrow V V$

CYK算法示例-填写第3行

跨度

6						
5						
4						
3	S					
2		VP		V'		
1	NP,N	V	NP,N	V	V	de

$$j = 3, i = 1, k = 1$$

$$P(i,k) = P(1,1) = NP$$

$$P(i+k, j-k) = P(2,2) = VP$$

$$\Rightarrow P(i,j) = P(1,3) = S$$

1

2

3

4

5

6

位置

张三

是

县长

派

来

的

CYK算法示例-填写第3行 (续)

- (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) $V' \rightarrow V V$

跨度

6						
5						
4						
3	S		CS			
2		VP		V'		
1	NP,N	V	NP,N	V	V	de

$$j = 3, i = 3, k = 1$$

$$P(i,k) = P(3,1) = NP$$

$$P(i+k, j-k) = P(4,2) = V'$$

$$\Rightarrow P(i,j) = P(3,3) = CS$$

1

2

3

4

5

6

位置

张三

是

县长

派

来

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) $V' \rightarrow V V$

CYK算法示例-填写第4行

跨 度	6						
	5						
	4			NP			
	3	S		CS			
	2		VP		V'		
	1	NP,N	V	NP,N	V	V	de
	1	2	3	4	5	6	位置
	张三	是	县长	派	来	的	

$j = 4, i = 3, k = 3$
 $P(i,k) = P(3,3) = CS$
 $P(i+k, j-k) = P(6,1) = de$
 $\Rightarrow P(i,j) = P(3,4) = NP$

- (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) $V' \rightarrow V V$

CYK算法示例-填写第5行

跨度	6					
5		VP				
4			NP			
3	S		CS			
2		VP		V'		
1	NP,N	V	NP,N	V	V	de
	1	2	3	4	5	6
	张三	是	县长	派	来	的

$j = 5, i = 2, k = 1$
 $P(i,k) = P(2,1) = V$
 $P(i+k, j-k) = P(3,4) = NP$
 $\Rightarrow P(i,j) = P(2,5) = VP$

位置

- (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) $V' \rightarrow V V$

CYK算法示例-填写第6行

跨 度	6	S						
	5		VP					
	4			NP				
	3	S		CS				
	2		VP		V'			
	1	NP,N	V	NP,N	V	V	de	
		1	2	3	4	5	6	位置
		张三	是	县长	派	来	的	

$j = 6, i = 1, k = 1$
 $P(i,k) = P(1,1) = NP$
 $P(i+k, j-k) = P(2,5) = VP$
 $\Rightarrow P(i,j) = P(1,6) = S$



CYK算法：特点

- 本质上是一种自底向上分析法；
- 采用广度优先的搜索策略；
- 采用并行算法，不需要回溯，没有冗余的操作；
- 时间复杂度 $O(n^3)$ ；
- 由于采用广度优先搜索，在歧义较多时，必须分析到最后才知道结果，无法采用启发式策略进行改进。



5 ATN分析方法

张三是县长派来的
苍蝇是瞎子打死的
主意是董永想出来的

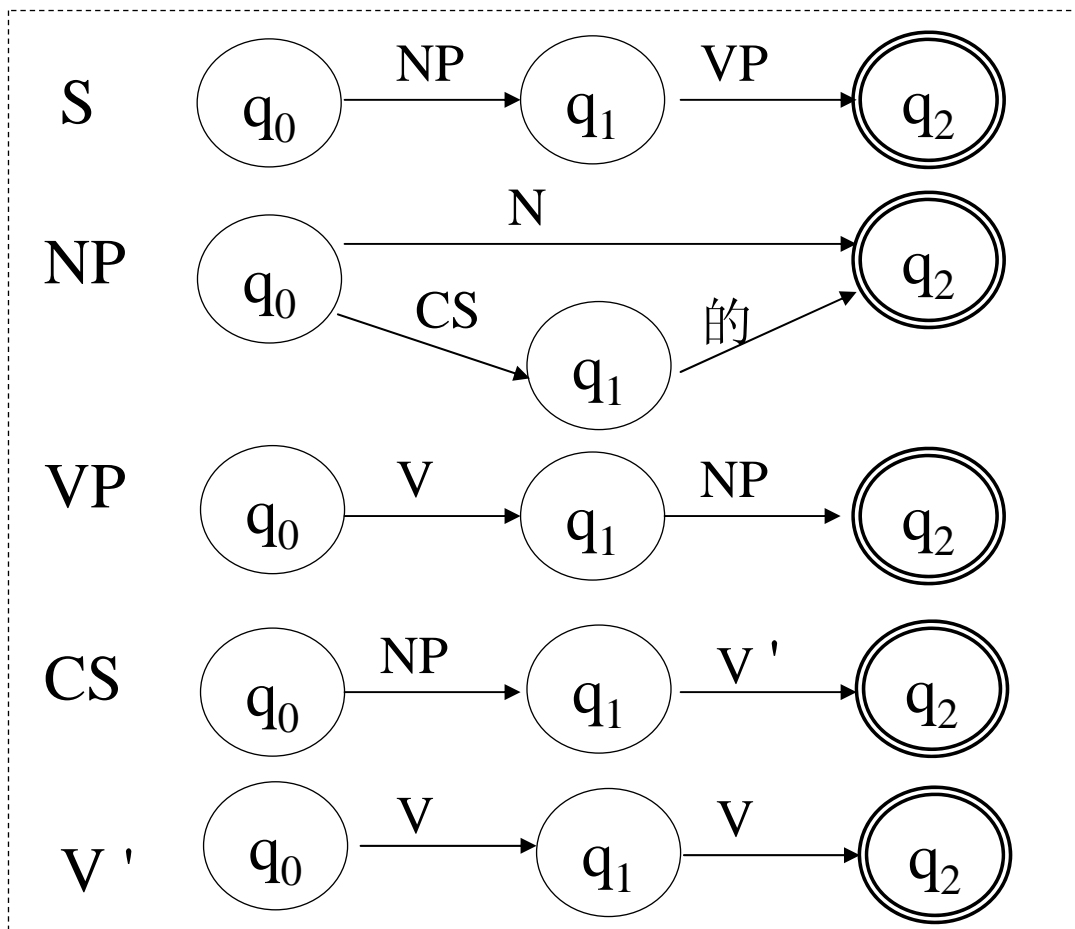
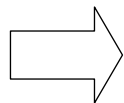
.....

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) $V' \rightarrow V V$

从CFG到递归转移网络 (RTN)

- (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) $V' \rightarrow V V$



RTN算法描述 —— 基本概念

子网名称	S, NP, VP, CS, V' (S是个特殊的子网)
状态节点	q_0, q_1, q_2 (其中 q_0 是开始状态, q_2 是结束状态)
出边	从当前状态向下一个状态转移的弧 (NP, VP, V, ...)
待分析字符串	$w_1w_2w_3\dots$ (主意是董永想出来的, 张三是县长派来的)
递归栈	记录来自哪个子网, 以及回到上层子网时应处的状态
当前状态	<当前子网, 当前状态节点> <当前出边, 后续状态节点>

对过去加以保留

对将来进行预测

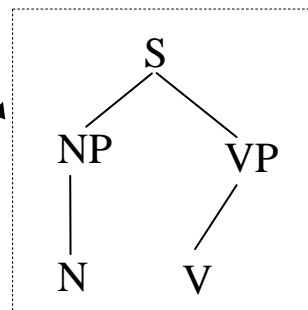
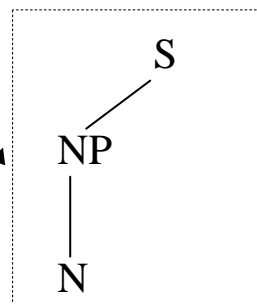
RTN算法描述 (top-down)

1. 初始化：当前状态为 $\langle S, q_0 \rangle$ ，字符串指针指向待分析字符串第一个字符，递归栈清空。
2. 2.1 如果当前状态节点不是终止状态：
把当前状态节点出边指针指向第一个出边，
 - 2.1.1 如果当前出边的标记为终结符，
比较该标记与当前字符串移动指针所指字符
 - 2.1.1.1 如果相等，预测得到验证，构造子树，将当前状态节点设为当前出边的后续状态节点，同时修改当前出边；
 - 2.1.1.2 如果不相等，出边指针指向下一个出边。如果不存在下一个出边并且存在回溯点，则回溯，否则分析失败。
 - 2.1.2 如果当前出边的标记为非终结符，把当前子网名称及当前节点出边的后续状态压栈，并把当前状态节点设为当前子网的开始状态，同时修改当前出边及后续状态。
 - 2.1.3 如果当前出边有多重选择，需要设置恢复断点，保存递归栈、待分析字符串、当前ATN状态以及出边列表状态的全部内容，以便回溯时使用。
- 2.2. 如果当前状态节点是终止状态但不是子网S的终止状态：
将递归栈退栈，并将当前状态设置为当前栈顶的状态，转入步骤2继续执行。
- 2.3. 如果当前状态节点是终止状态并且是子网S的终止状态：
 - 2.3.1 若递归栈已空且待分析字符串已空，则分析成功，结束；
 - 2.3.2 否则，若存在回溯点，回溯。
 - 2.3.3 若不存在回溯点，分析失败。

RTN算法过程示例-1

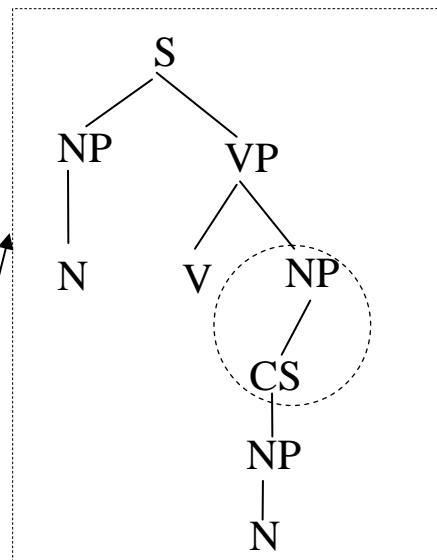
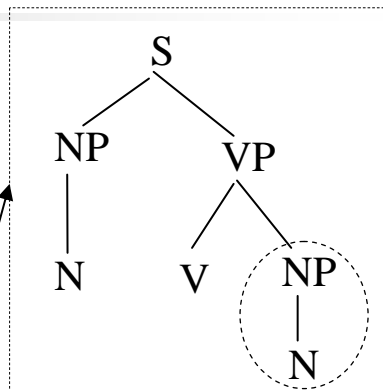
记录回溯点,即当前选择的出边序号,回溯时+1

递归栈内容	待分析字符串	<子网,当前状态>	<出边,后续状态>	回溯
空	N V N V V 的	<S, q ₀ >	<NP, q ₁ >	
<S, q ₁ >	N V N V V 的	<NP, q ₀ >	<N, q ₂ >, <CS, q ₁ >	1
<S, q ₁ ><NP, q ₂ >	N V N V V 的		N 预测验证成功	
<S, q ₁ ><NP, q ₂ >	V N V V 的	<NP, q ₂ >	NP 预测验证成功	
<S, q ₁ >	V N V V 的	<S, q ₁ >	<VP, q ₂ >	
<S, q ₂ >	V N V V 的	<VP, q ₀ >	<V, q ₁ >	
<S, q ₂ ><VP, q ₁ >	V N V V 的		V 预测验证成功	



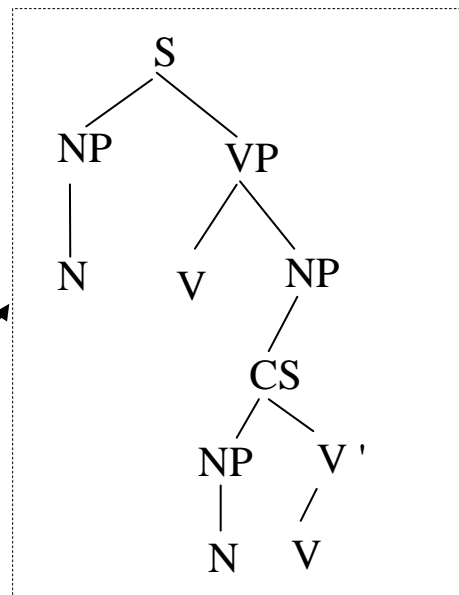
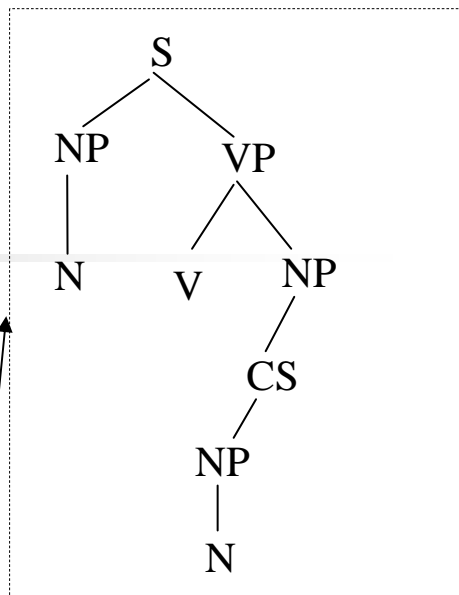
RTN算法过程示例-2

递归栈内容	待分析字符串	<子网,当前状态>	<出边,后续状态>	回溯
<S, q ₂ ><VP, q ₂ >	N V V 的	<NP, q ₀ >	<N, q ₂ >, <CS, q ₁ >	1
<S, q ₂ ><VP, q ₂ > <NP, q ₂ >	N V V 的		N 预测验证成功	
<S, q ₂ ><VP, q ₂ > <NP, q ₂ >	V V 的	<NP, q ₂ >	NP 预测验证成功	
<S, q ₂ ><VP, q ₂ >	V V 的	<VP, q ₂ >	无出边	
<S, q ₂ ><VP, q ₂ >	V V 的	<VP, q ₂ >	回溯 (选择另一条出边)	
<S, q ₂ ><VP, q ₂ >	N V V 的	<NP, q ₀ >	<CS, q ₁ >	2
<S, q ₂ ><VP, q ₂ > <NP, q ₁ >	N V V 的	<CS, q ₀ >	<NP, q ₁ >	
<S, q ₂ ><VP, q ₂ > <NP, q ₁ ><CS, q ₁ >	N V V 的	<NP, q ₀ >	<N, q ₂ >, <CS, q ₁ >	1



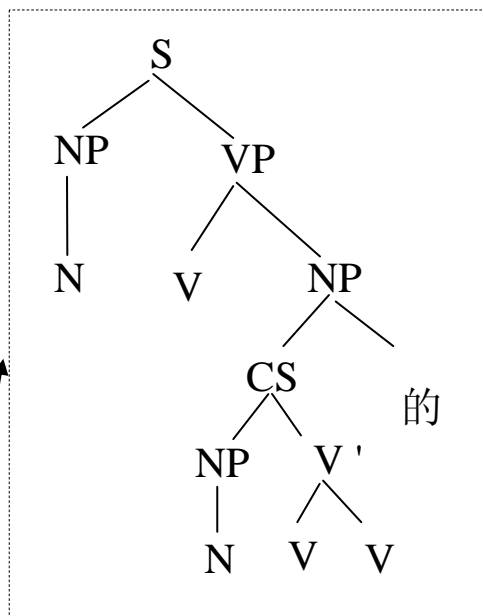
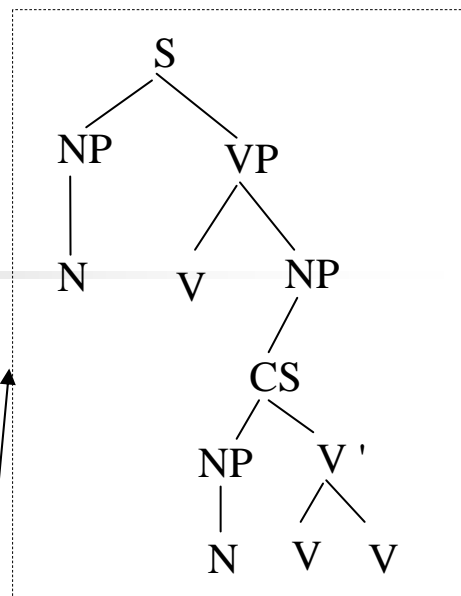
RTN算法过程示例-3

递归栈内容	待分析字符串	<子网,当前状态>	<出边,后续状态>	回溯
$\langle S, q_2 \rangle \langle VP, q_2 \rangle$ $\langle NP, q_2 \rangle \langle CS, q_1 \rangle$ $\langle NP, q_1 \rangle$	N V V 的		N 预测验证成功	
$\langle S, q_2 \rangle \langle VP, q_2 \rangle$ $\langle NP, q_2 \rangle \langle CS, q_1 \rangle$ $\langle NP, q_1 \rangle$	V V 的	$\langle NP, q_0 \rangle$	NP 预测验证成功	
$\langle S, q_2 \rangle \langle VP, q_2 \rangle$ $\langle NP, q_1 \rangle \langle CS, q_1 \rangle$	V V 的	$\langle CS, q_1 \rangle$	$\langle V', q_2 \rangle$	
$\langle S, q_2 \rangle \langle VP, q_2 \rangle$ $\langle NP, q_1 \rangle \langle CS, q_2 \rangle$	V V 的	$\langle V', q_0 \rangle$	$\langle V, q_1 \rangle$	
$\langle S, q_2 \rangle \langle VP, q_2 \rangle$ $\langle NP, q_1 \rangle \langle CS, q_2 \rangle$ $\langle V', q_1 \rangle$	V V 的		V 预测验证成功	



RTN算法过程示例-4

递归栈内容	待分析字符串	<子网,当前状态>	<出边,后续状态>	回溯
<S, q ₂ ><VP, q ₂ > <NP, q ₁ ><CS, q ₂ > <V', q ₂ >	V 的	<V', q ₁ >	<V, q ₂ >	
<S, q ₂ ><VP, q ₂ > <NP, q ₁ ><CS, q ₂ > <V', q ₂ >	V 的		V 预测验证成功	
<S, q ₂ ><VP, q ₂ > <NP, q ₁ ><CS, q ₂ >	的	<CS, q ₂ >	V'预测验证成功	
<S, q ₂ ><VP, q ₂ > <NP, q ₁ >	的	<NP, q ₁ >	<的, q ₂ > CS 预测验证成功	
<S, q ₂ ><VP, q ₂ > <NP, q ₁ >	的		"的"预测验证成功	
<S, q ₂ ><VP, q ₂ >	\$	<NP, q ₂ >	NP 预测验证成功	
<S, q ₂ >	\$	<VP, q ₂ >	VP 预测验证成功	
空	\$	<S, q ₂ >	S 预测验证成功 分析成功结束	





从RTN到ATN

- 寄存器 (register)
- 条件测试 (condition)
- 动作 (action)

一件衣服 —— * 一件小孩

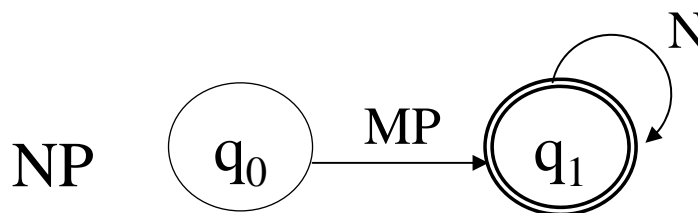
一件小孩的衣服

张三打李四 —— 李四被张三打了

施事：张三

受事：李四

ATN示例



NP: $q_0 - MP - q_1$

动作:

- (1) 进入MP网络
- (2) 将“量词”寄存器的值置为“MP*”

NP: $q_1 - N - q_1$

条件:

- (1) 当前词是名词
- (2) 名词寄存器表中的“量词”值为“MP*”

一件衣服

“一件” → “量词”寄存器

“MP*” = “件”

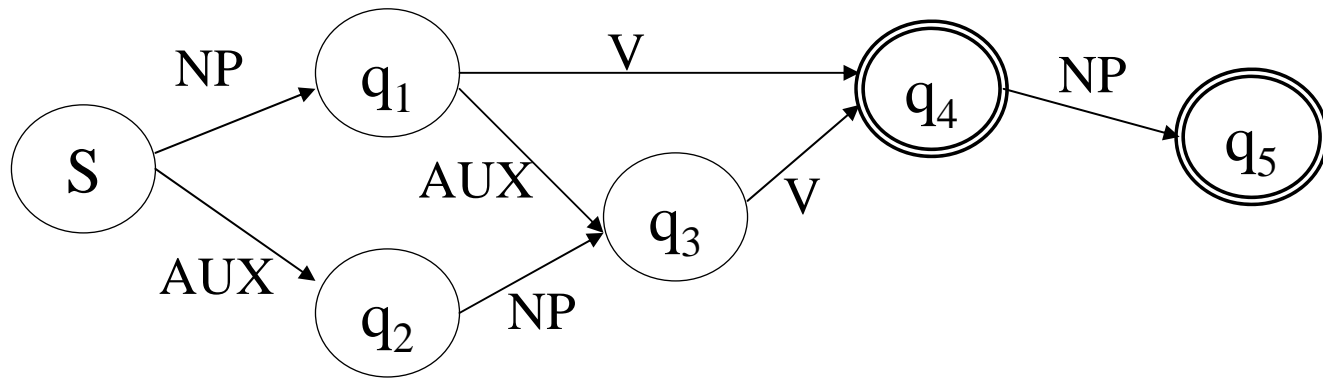
“衣服”的“量词”值为“件”

可以匹配，NP分析成功

一件小孩

分析失败

ATN示例 (续)



You can help them

NP-AUX-V-NP

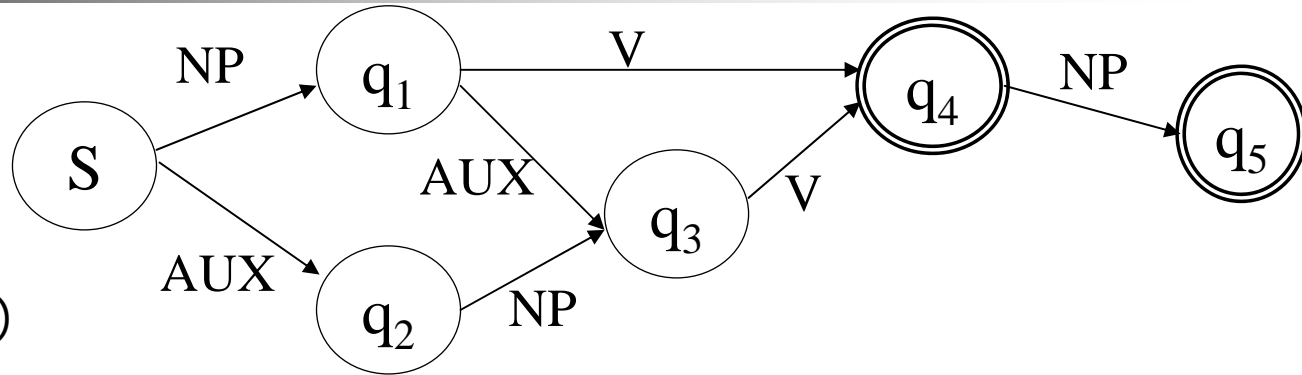
Can you help them

AUX-NP-V-NP

引自刘开瑛、郭炳炎，1991，P29

ATN的Lisp表示法

```
1 ((S/ (PUSH NP / T
2 (SETR SUBJ *)
3 (SETR TYPE "DCL")
4 (TO q1)
5 (CAT AUX T
6 (SETR AUX *)
7 (SETR TYPE "QUESTION")
8 (TO q2))
9 (q1 (CAT V T
10 (SETR AUX NIL)
11 .....
```

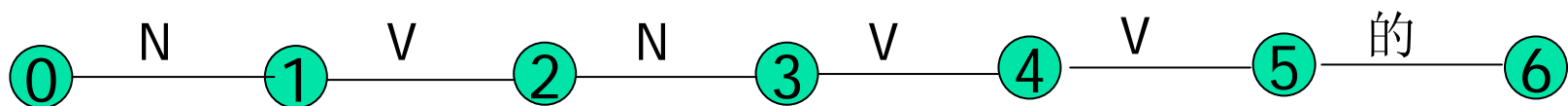




6 Earley算法

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)

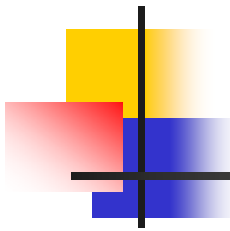
一个状态由四部分组成：

- (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) 整数 i ：状态起点（已分析子串的起点）
- (4) 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 (Predicator)**：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 $Z \rightarrow \alpha \cdot X \beta \quad j, k$ 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: $X \rightarrow \cdot \gamma \quad k, k$

Scanner: 对于状态 $Z \rightarrow \alpha \cdot X \beta \quad j, k$ 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: $Z \rightarrow \alpha X \cdot \beta \quad j, k+1$

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

说明: 以上 α, β, γ 是终结符或非终结符串, 其中 α, β 均为空字符 $i \leq j \leq k$



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) 将语法规则中形如 $S \rightarrow \alpha$ 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**Predictor**;

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

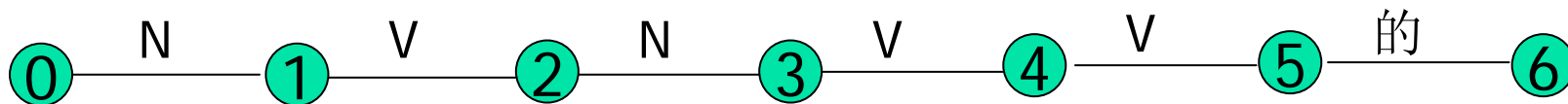
Earley算法过程示例

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) $V' \rightarrow V V$

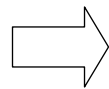


Earley算法过程示例-1

当前间隔 ①

预测

$\langle S \rightarrow \cdot NP VP \ 0,0 \rangle$
 $\langle NP \rightarrow \cdot N \ 0,0 \rangle$
 $\langle NP \rightarrow \cdot CS \text{ 的} \ 0,0 \rangle$
 $\langle CS \rightarrow \cdot NP V' \ 0,0 \rangle$



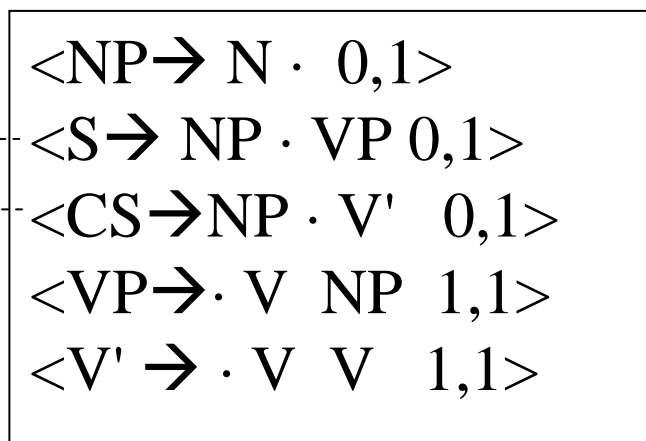
当前间隔 ②

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

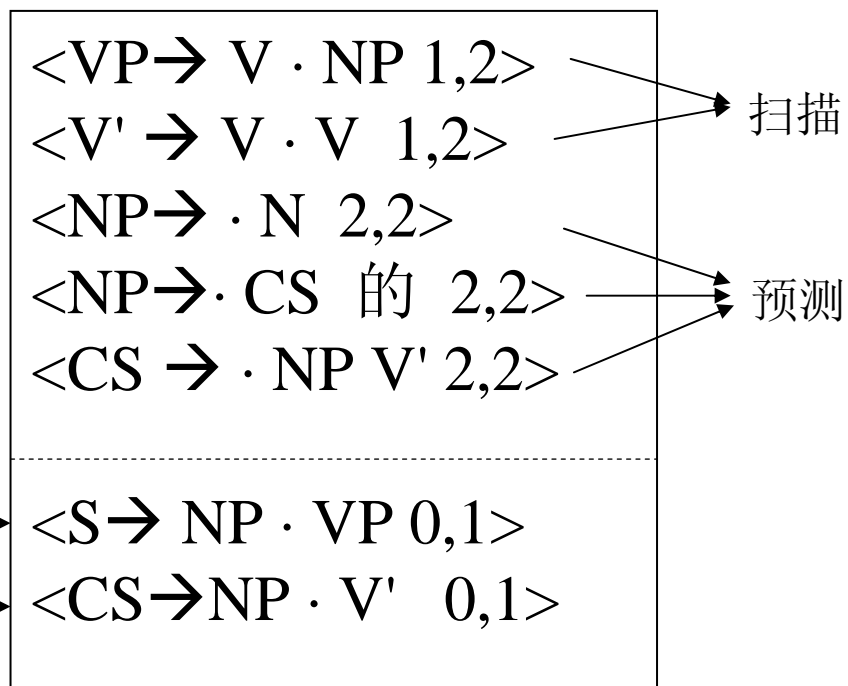
预测 → 扫描 → 归约 → 预测 → 扫描 → 归约 → 预测

Earley算法过程示例-2

当前间隔 ①



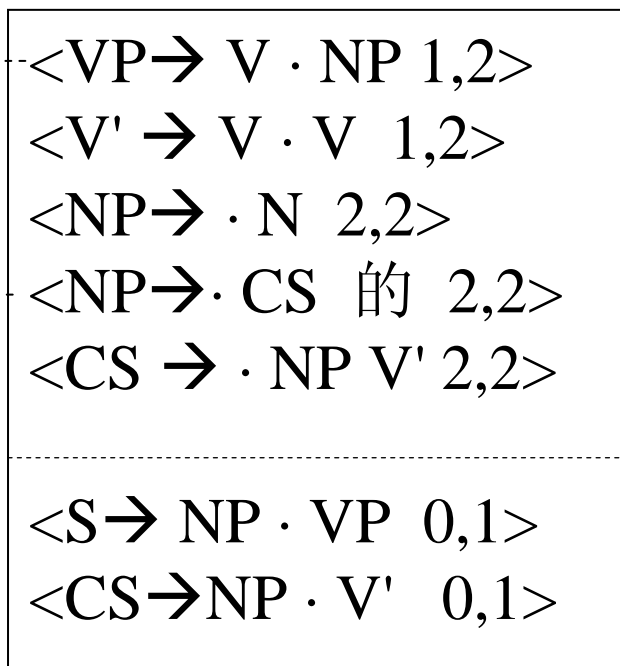
当前间隔 ②



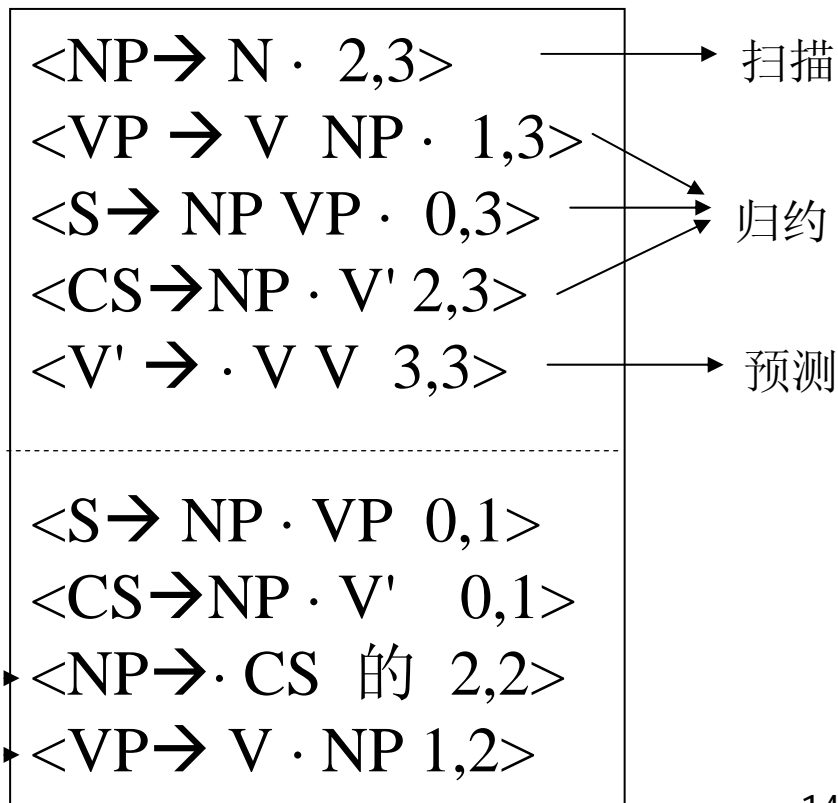
-----> 保留状态

Earley算法过程示例-3

当前间隔 ②

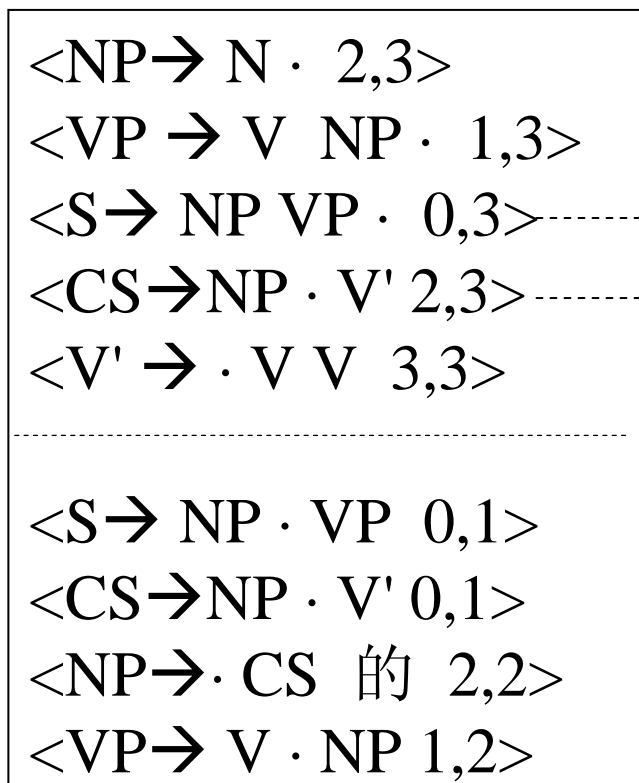


当前间隔 ③

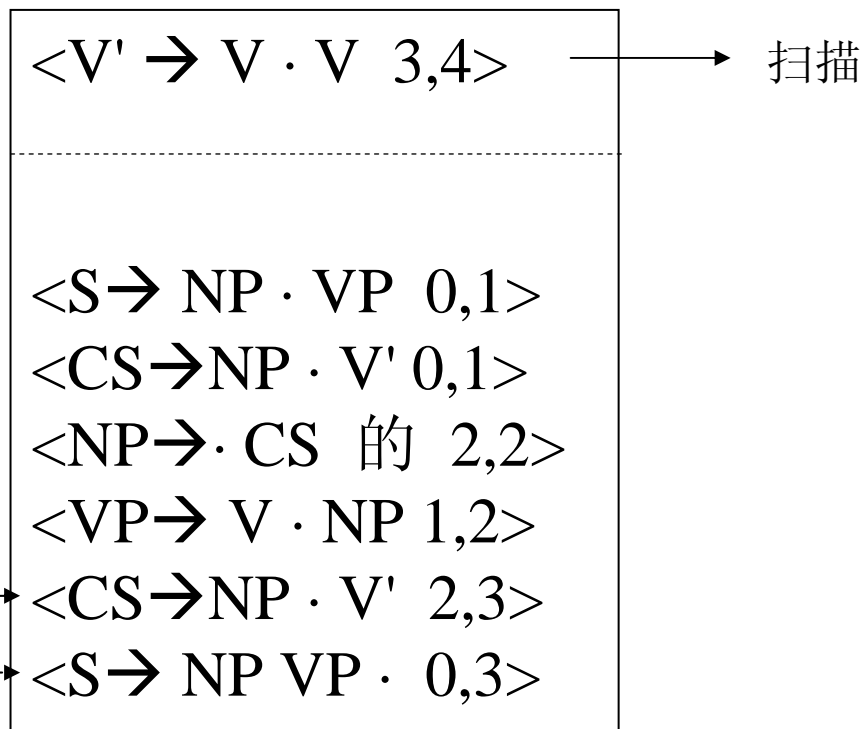


Earley算法过程示例-4

当前间隔 ③

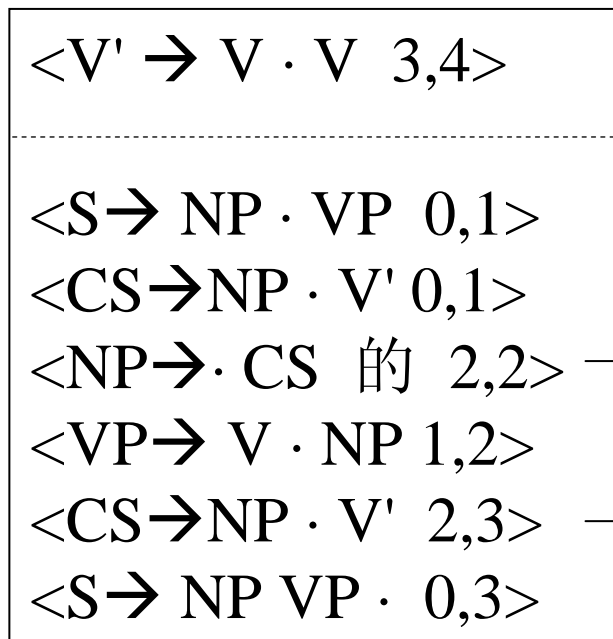


当前间隔 ④

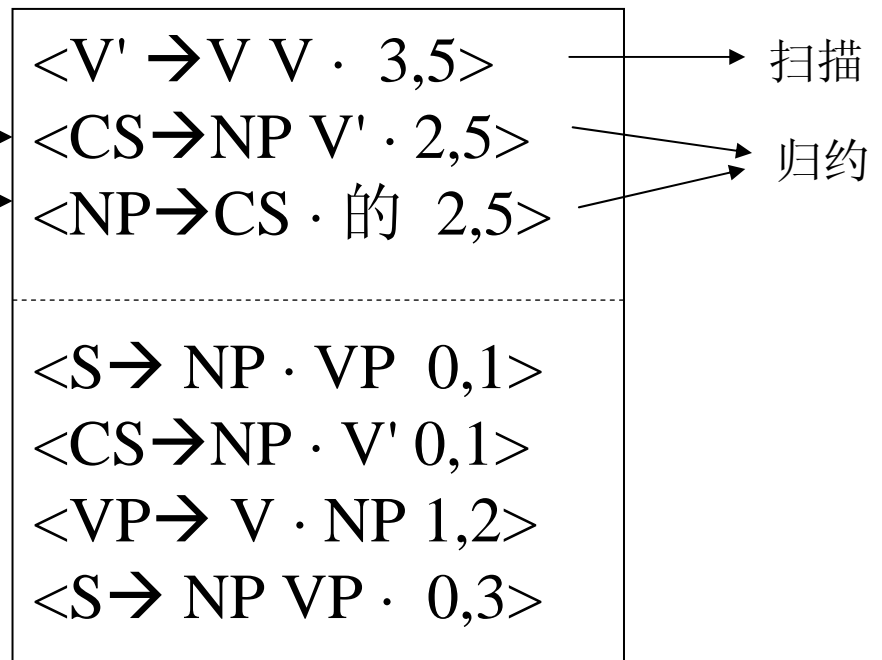


Earley算法过程示例-5

当前间隔 ④

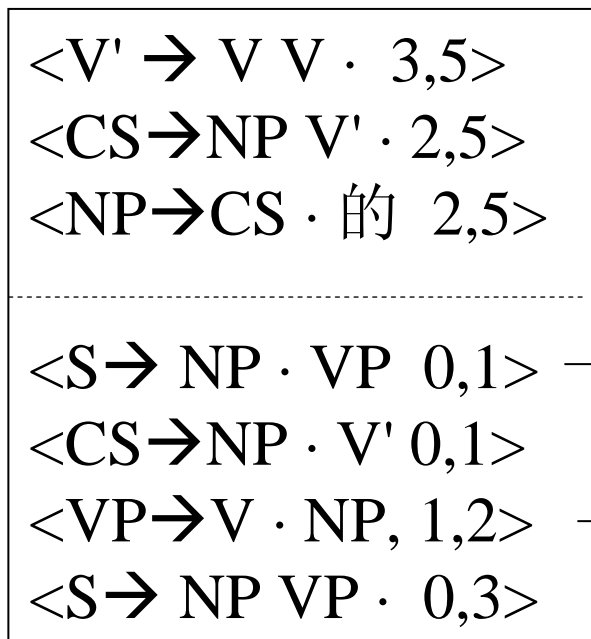


当前间隔 ⑤

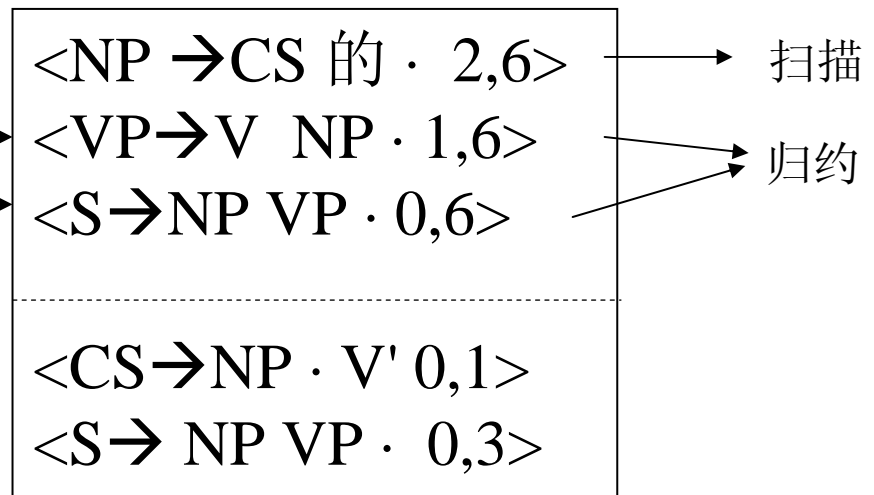


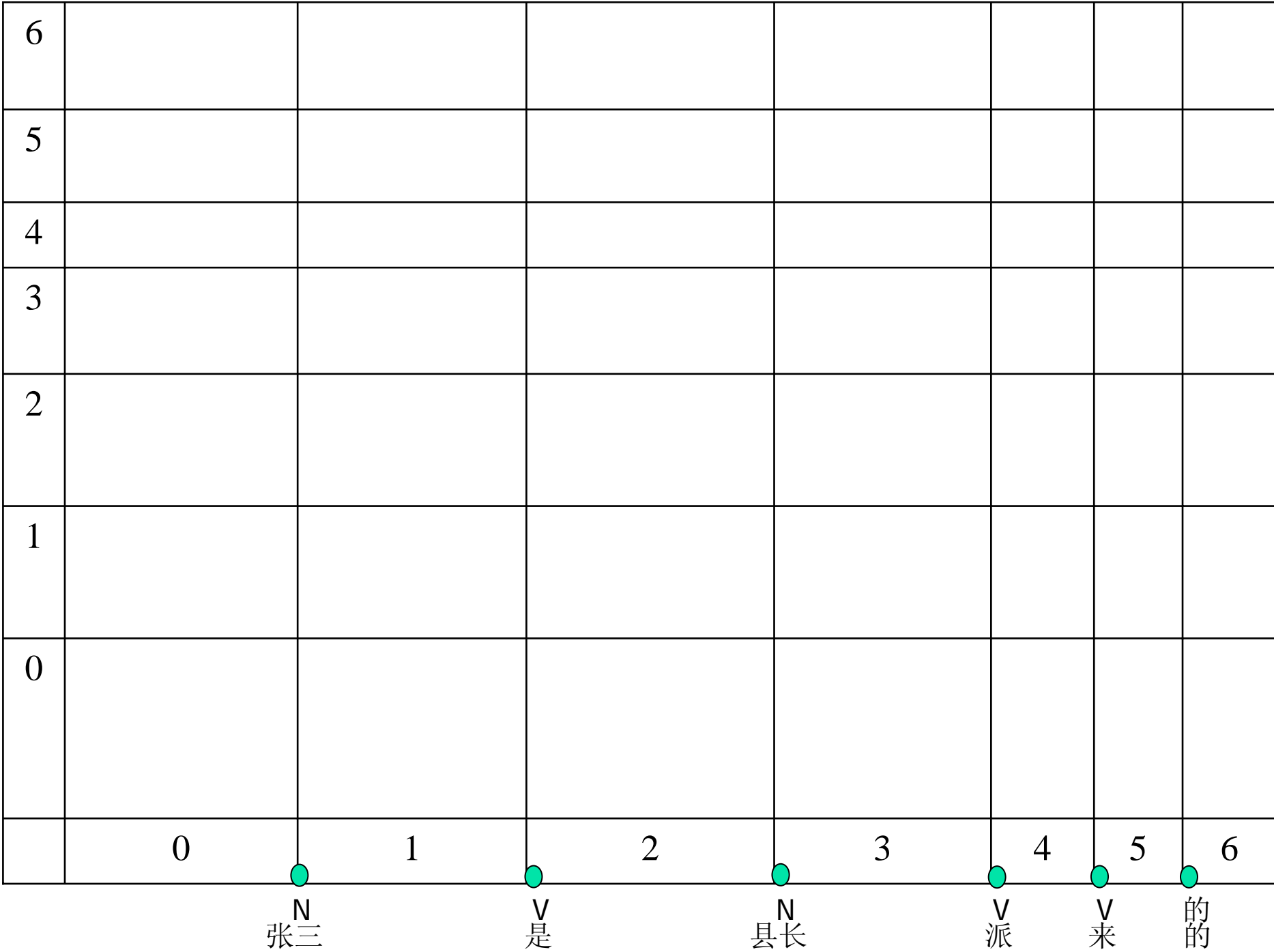
Earley算法过程示例-6

当前间隔 ⑤



当前间隔 ⑥





6									
5									
4									
3									
2									
1									
0	S → · NP VP								
	0	1	2	3	4	5	6		

N
张三

V
是

N
县长

V
派

V
来

的
的

种子

6									
5									
4									
3									
2									
1									
0	CS → · NP V'								
	NP → · N								
	NP → · CS 的								
	S → · NP VP								
	0	1	2	3	4	5	6		

预测
预测
预测
种子

N 张三 V 是 N 县长 V 派 V 来 的

6									
5									
4									
3									
2									
1	NP → N .								扫描
0	CS → . NP V ' NP → . N NP → . CS 的 S → . NP VP								预测 预测 预测 种子
	0	1	2	3	4	5	6		

N
V
N
V
V
的

张三
是
县长
派
来
的

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

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

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

6									
5									
4									
3									
2		$VP \rightarrow V \cdot NP$ $V' \rightarrow V \cdot V$	$NP \rightarrow \cdot N$ $NP \rightarrow \cdot CS$ 的						预测扫描
1	$S \rightarrow NP \cdot VP$ $CS \rightarrow NP \cdot V'$ $NP \rightarrow N \cdot$	$VP \rightarrow \cdot V NP$ $V' \rightarrow \cdot V V$							预测归约扫描
0	$CS \rightarrow \cdot NP V'$ $NP \rightarrow \cdot N$ $NP \rightarrow \cdot CS$ 的 $S \rightarrow \cdot NP VP$								预测 预测 预测 种子
	0	1	2	3	4	5	6		
		N 张三	V 是	N 县长	V 派	V 来	的		

6									
5									
4									
3									
2		VP → V · NP V' → V · V	CS → · NP V' NP → · N NP → · CS 的						预测 预测 扫描
1	S → NP · VP CS → NP · V' NP → N ·	VP → · V NP V' → · V V							预测 归约 扫描
0	CS → · NP V' NP → · N NP → · CS 的 S → · NP VP								预测 预测 预测 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									
1									
0									
	0	1	2	3	4	5	6		

NP → N .

扫描

CS → . NP V'

预测

VP → V . NP

预测

NP → . N

扫描

V' → V . V

NP → . CS 的

1

S → NP . VP

预测

CS → NP . V'

归约

NP → N .

扫描

VP → . V NP

V' → . V V

0

CS → . NP V'

预测

NP → . N

预测

NP → . CS 的

预测

S → . NP VP

种子

N
张三

V
是

N
县长

V
派

V
来

的

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

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

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

6		$VP \rightarrow V NP \cdot$	$NP \rightarrow CS \text{ 的} \cdot$				归约扫描
5			$NP \rightarrow CS \cdot \text{ 的}$ $CS \rightarrow NP V' \cdot$	$V' \rightarrow V V \cdot$			归约扫描
4				$V' \rightarrow V \cdot V$			扫描
3	$S \rightarrow NP VP \cdot$	$VP \rightarrow V NP \cdot$	$CS \rightarrow NP \cdot V'$ $NP \rightarrow N \cdot$	$V' \rightarrow \cdot V V$			预测归约扫描
2		$VP \rightarrow V \cdot NP$ $V' \rightarrow V \cdot V$	$CS \rightarrow \cdot NP V'$ $NP \rightarrow \cdot N$ $NP \rightarrow \cdot CS \text{ 的}$				预测预测扫描
1	$S \rightarrow NP \cdot VP$ $CS \rightarrow NP \cdot V'$ $NP \rightarrow N \cdot$	$VP \rightarrow \cdot V NP$ $V' \rightarrow \cdot V V$					预测归约扫描
0	$CS \rightarrow \cdot NP V'$ $NP \rightarrow \cdot N$ $NP \rightarrow \cdot CS \text{ 的}$ $S \rightarrow \cdot NP VP$						预测预测种子
	0	1	2	3	4	5	6
		N 张三	V 是	N 县长	V 派	V 来	的

6	$S \rightarrow NP VP \cdot$	$VP \rightarrow V NP \cdot$	$NP \rightarrow CS \text{ 的} \cdot$				归约扫描
5			$NP \rightarrow CS \cdot \text{ 的}$ $CS \rightarrow NP V' \cdot$	$V' \rightarrow V V \cdot$			归约扫描
4				$V' \rightarrow V \cdot V$			扫描
3	$S \rightarrow NP VP \cdot$	$VP \rightarrow V NP \cdot$	$CS \rightarrow NP \cdot V'$ $NP \rightarrow N \cdot$	$V' \rightarrow \cdot V V$			预测归约扫描
2		$VP \rightarrow V \cdot NP$ $V' \rightarrow V \cdot V$	$CS \rightarrow \cdot NP V'$ $NP \rightarrow \cdot N$ $NP \rightarrow \cdot CS \text{ 的}$				预测预测扫描
1	$S \rightarrow NP \cdot VP$ $CS \rightarrow NP \cdot V'$ $NP \rightarrow N \cdot$	$VP \rightarrow \cdot V NP$ $V' \rightarrow \cdot V V$					预测归约扫描
0	$CS \rightarrow \cdot NP V'$ $NP \rightarrow \cdot N$ $NP \rightarrow \cdot CS \text{ 的}$ $S \rightarrow \cdot NP VP$						预测预测种子
	0	1	2	3	4	5	6

N
张三

V
是

N
县长

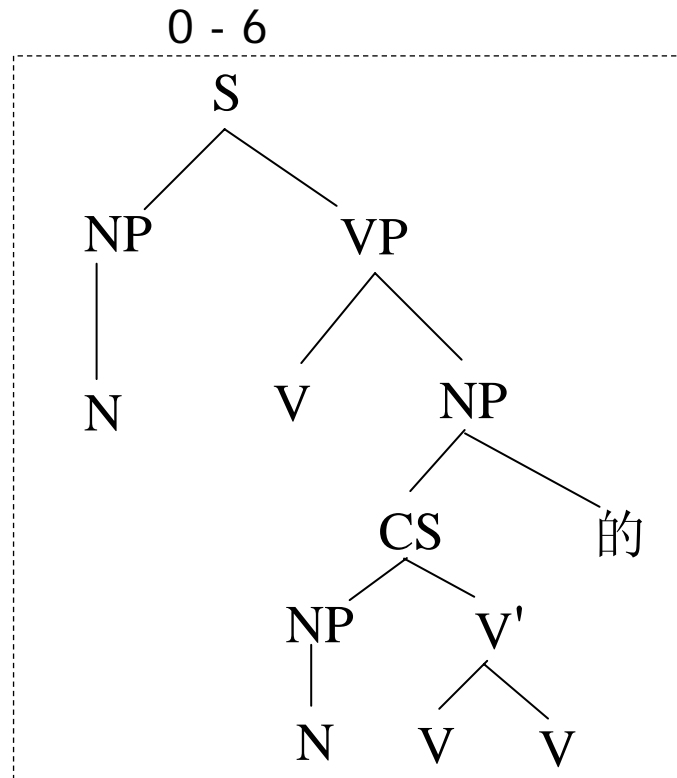
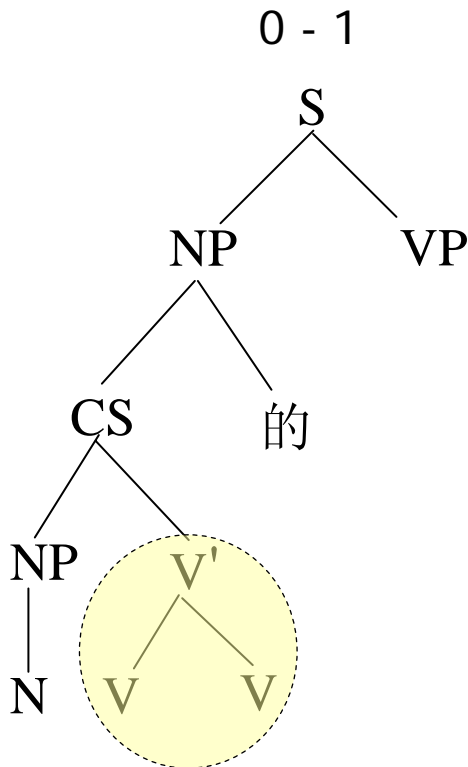
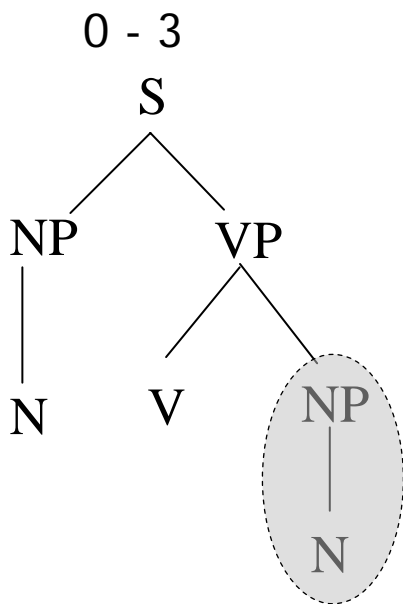
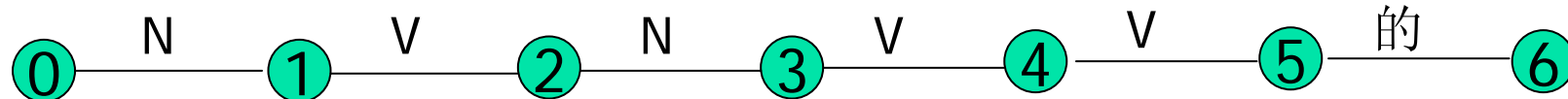
V
派

V
来

的

6	$S \rightarrow NP VP \cdot$	$VP \rightarrow V NP \cdot$	$NP \rightarrow CS \text{ 的} \cdot$				归约扫描
5			$NP \rightarrow CS \cdot \text{ 的}$ $CS \rightarrow NP V' \cdot$	$V' \rightarrow V V \cdot$			归约扫描
4				$V' \rightarrow V \cdot V$			扫描
3	$S \rightarrow NP VP \cdot$	$VP \rightarrow V NP \cdot$	$CS \rightarrow NP \cdot V'$ $NP \rightarrow N \cdot$	$V' \rightarrow \cdot V V$			预测归约扫描
2		$VP \rightarrow V \cdot NP$ $V' \rightarrow V \cdot V$	$CS \rightarrow \cdot NP V'$ $NP \rightarrow \cdot N$ $NP \rightarrow \cdot CS \text{ 的}$				预测预测扫描
1	$S \rightarrow NP \cdot VP$ $CS \rightarrow NP \cdot V'$ $NP \rightarrow N \cdot$	$VP \rightarrow \cdot V NP$ $V' \rightarrow \cdot V V$					预测归约扫描
0	$CS \rightarrow \cdot NP V'$ $NP \rightarrow \cdot N$ $NP \rightarrow \cdot CS \text{ 的}$ $S \rightarrow \cdot NP VP$						预测预测种子
	0	1	2	3	4	5	6
		N 张三	V 是	N 县长	V 派	V 来	的

Earley算法构造分析树示意图





进一步阅读文献

- 刘开瑛 郭炳炎，1991，《自然语言处理》，科学出版社，第2章
- 冯志伟，2001，《计算语言学基础》，商务印书馆，第3章
- 赵铁军 等，2000，《机器翻译原理》，哈尔滨工业大学出版社，第5.5.2节
- <http://ccl.pku.edu.cn/doubtfire/> “自然语言处理”下“parsing”板块
- Earley, J. (1970) An Efficient Context-Free-Parsing Algorithm, Communication of ACM, Vol. 6, No. 8, pp94-102.
- Dick Grune & Cerial Jacobs, 1990, Parsing Techniques: A Practical Guide, First Published in 1990 by ELLIS HORWOOD LIMITED, Reprinted in 1997, 1998

表示B是A的左角
比如S-> NP VP,
则 NP L* S

附录：左角分析法-算法描述1

- 1a. If the element on top of the stack is the nonterminal **A** and if the first symbol of the remaining input is **t**, then we may remove **t** from the input and push an item $[B \rightarrow t \cdot \alpha]$ onto the stack, provided $B L^* A$.
- 1b. If the element on top of the stack is the nonterminal **A**, then we may push an item $[B \rightarrow \cdot]$ onto the stack, provided $B L^* A$. (The item $[B \rightarrow \cdot]$ is derived from an epsilon rule $B \rightarrow \epsilon$)
2. If the element on top of the stack is the item $[A \rightarrow \alpha \cdot t \beta]$ and if the first symbol of the remaining input is **t**, then we may remove **t** from the input and replace the item by the item $[A \rightarrow \alpha t \cdot \beta]$.
3. If the top-most two elements on the stack are $B [A \rightarrow \alpha \cdot]$, then we may replace the item by an item of the form $[C \rightarrow A \cdot \beta]$, provided $C L^* B$.
4. If the top-most three elements on the stack are $[B \rightarrow \beta \cdot A \gamma] A [A \rightarrow \alpha \cdot]$, then we may replace these three elements by the item $[B \rightarrow \beta A \cdot \gamma]$.
5. If a step according to one of the previous clauses ends with an item $[A \rightarrow \alpha \cdot B \beta]$ on top of the stack, where B is a nonterminal, then we subsequently push B onto the stack.
6. If the stack consists only of the two elements $S [S \rightarrow \alpha \cdot]$ and if the input has been completely read, then we may successfully terminate the parsing process.

附录：左角分析法-算法描述2

```
Proc LC_Parse( $\alpha$ :nonterminal-category)
  begin
    read next word W;
    CAT  $\leftarrow$  CAT(W);
    if CAT =  $\alpha$  then SUCCESS
    else
      if IS-LC(CAT,  $\alpha$ ) then expand_lc(CAT,  $\alpha$ )
      else FAIL
      endif
    endif
  end LC_Parse

Proc expand_lc(CAT,goal)
  begin
    for-each rule  $\beta \rightarrow$  CAT  $\gamma$ 
      if IS-LC(CAT, goal) then
        LC_Parse( $\gamma$ );
        if end-of-input then SUCCESS
        else expand_lc( $\beta$ ,goal)
        endif
      endif
    end-for-each
  end expand_lc
```

IS-LC(A,B)函数
用来判断
A是否为B的左角



复习思考题

- 1 构造一个汉语语法（片断），用ATN方法和Earley算法描述下面两个句子的分析过程
 - (1) 一条咬死了猎人的狗
 - (2) 狮子咬死了猎人的狗
- 2 试对汉语和英语的句法结构歧义进行分析比较



复习思考题（续）

3 下面是一个CFG文法： $G = (\{S, X\}, \{a, b, c, d, e, f, g\}, S, P)$ ，其中P定义如下：

$S \rightarrow a X c$

$S \rightarrow b X c$

$S \rightarrow b X d$

$S \rightarrow b X e$

$S \rightarrow c X e$

$X \rightarrow f X$

$X \rightarrow g$

（1）对G来说，纯粹的自顶向下分析器top-down parser和纯粹的自底向上分析器bottom-up parser，哪一个效率更高？为什么？

（2）采用你所选择的分析器处理字符串“bffge”，写出整个运行流程。