



第八章 句法分析

詹卫东

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



提纲

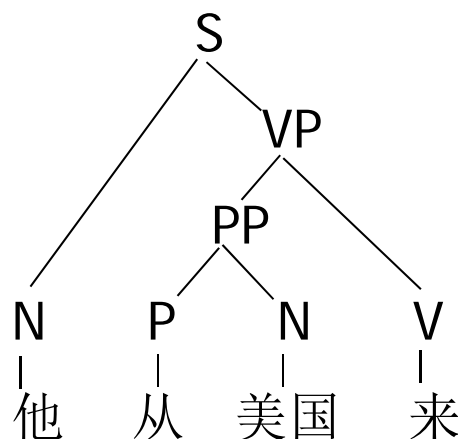
- 1 句法分析概说
- 2 自顶向下分析法
- 3 自底向上分析法
- 4 小结

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

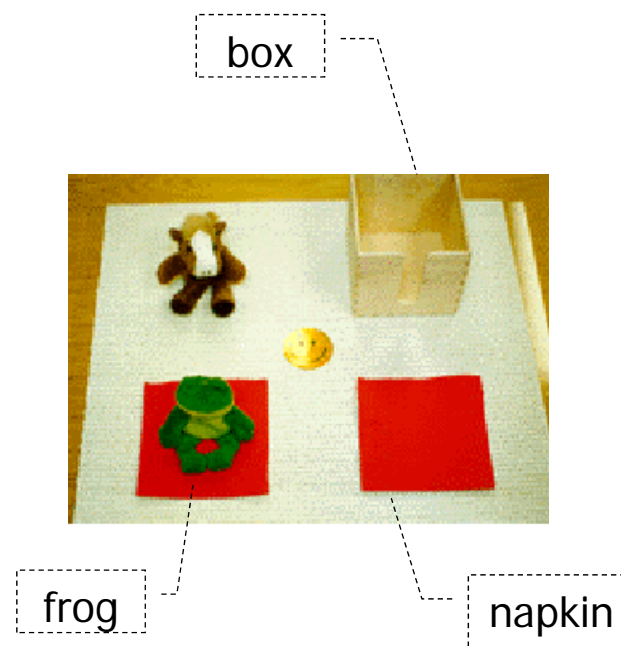
garden-path sentence/ 花园路径句

- Put the frog on the napkin in the box

- 麻子无头发黑脸大脚不大好看

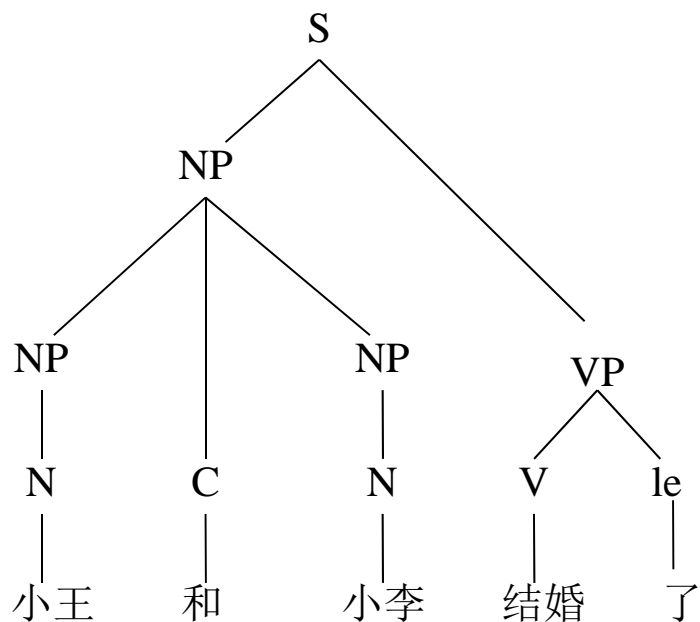
麻子无 头发黑 脸大脚不大 好看

麻子 无头发 黑脸 大脚 不大好看

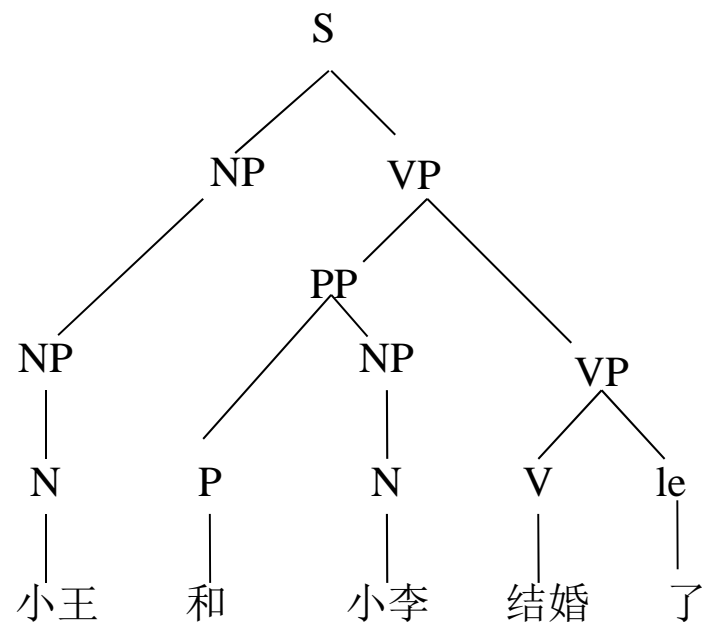


自然语言 —— 歧义无处不在（续）

小王和小李结婚了



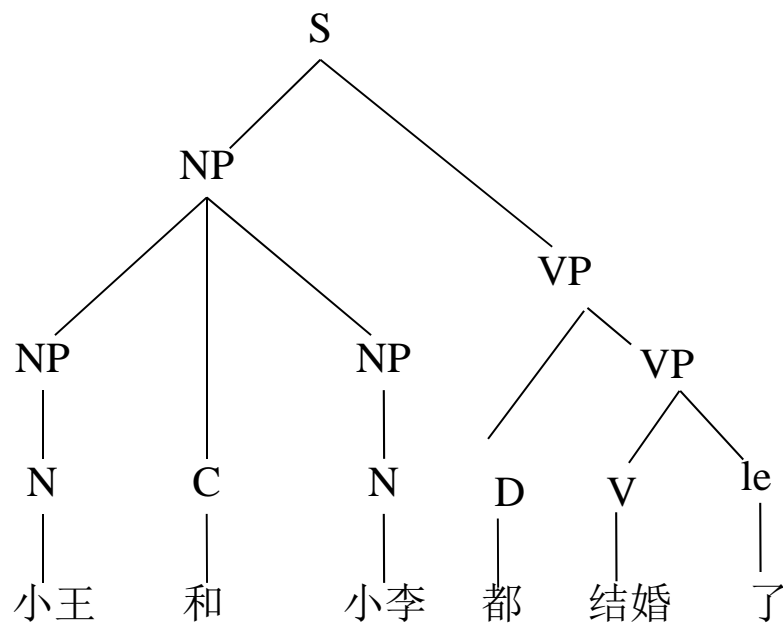
小王和小李都结婚了



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



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。
如何在一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 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

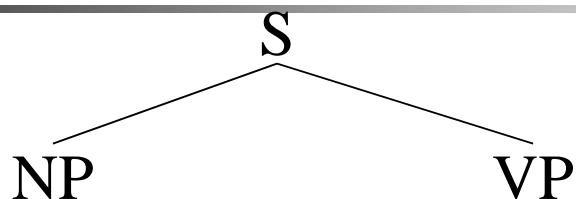
- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$

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) $V' \rightarrow V V$



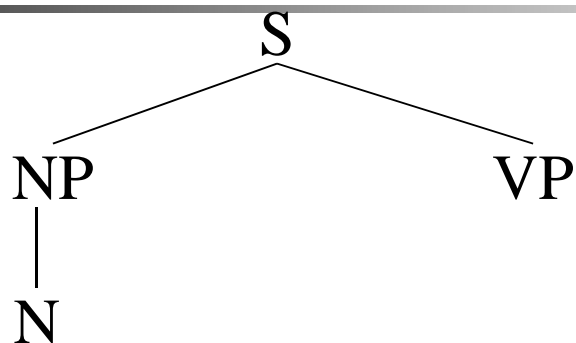
使用规则:

$S \rightarrow NP VP$

N	V	N	V	V	de
张三	是	县长	派	来	的

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) $V' \rightarrow V V$



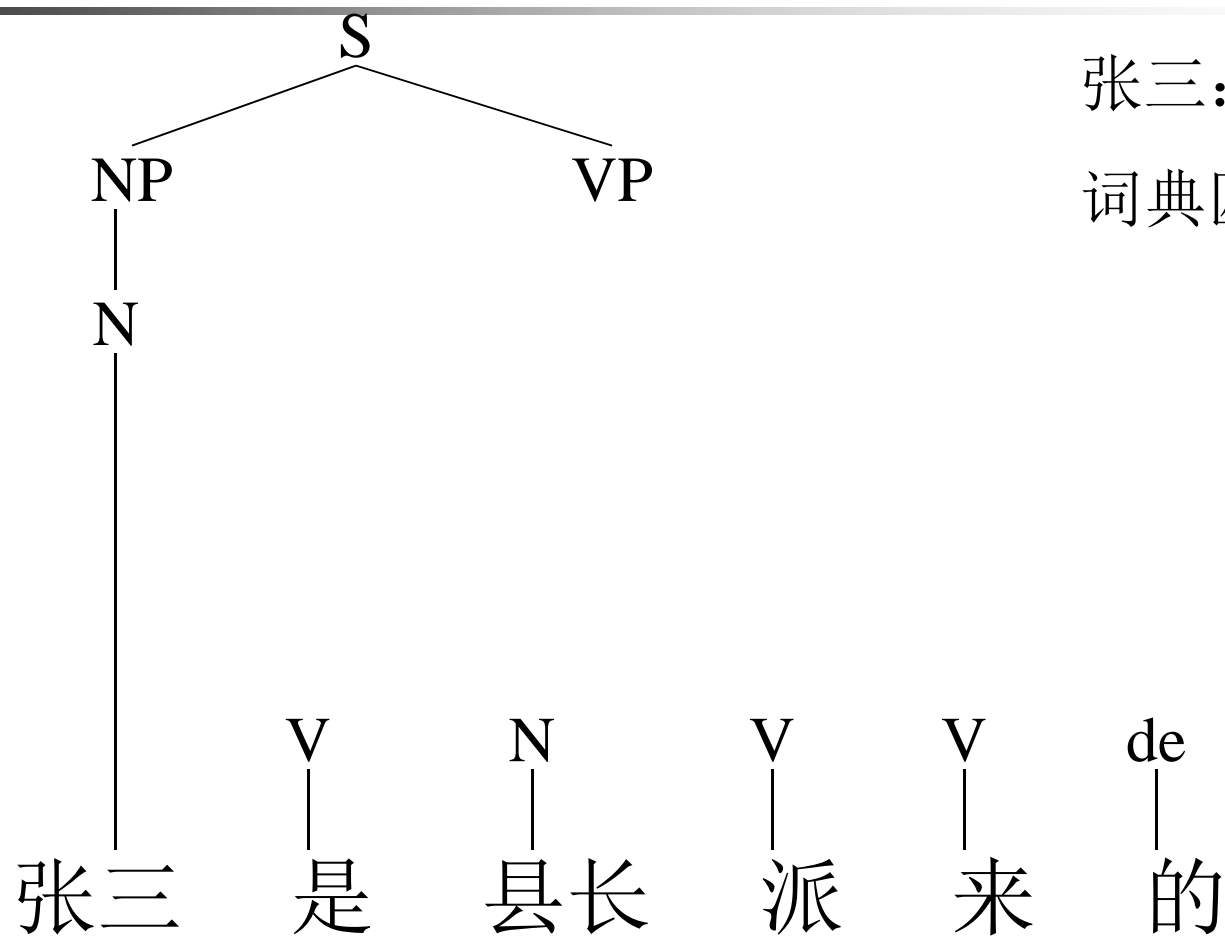
使用规则:

$NP \rightarrow N$



自顶向下分析法示例-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) $V' \rightarrow V V$

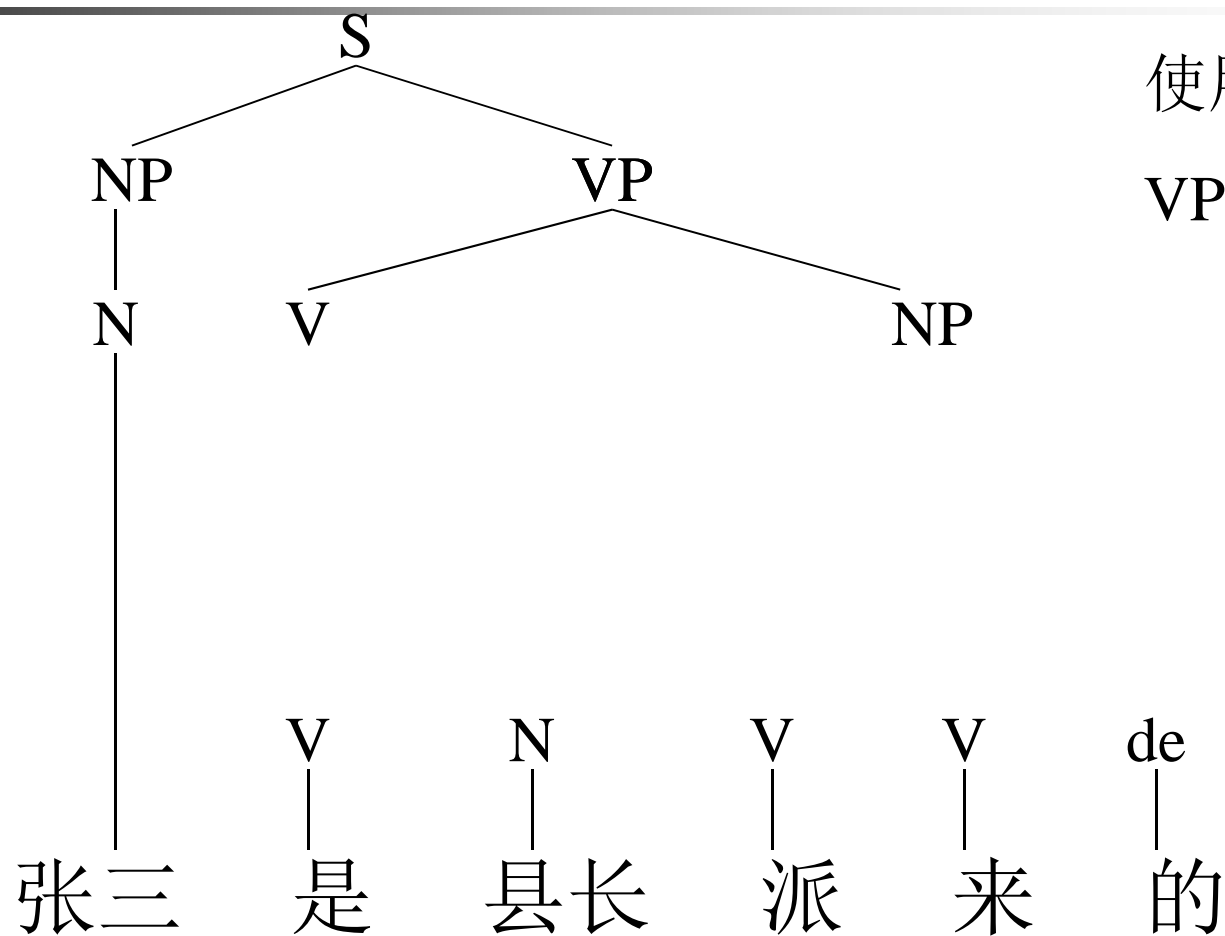


张三: N

词典匹配成功

自顶向下分析法示例-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) $V' \rightarrow V V$

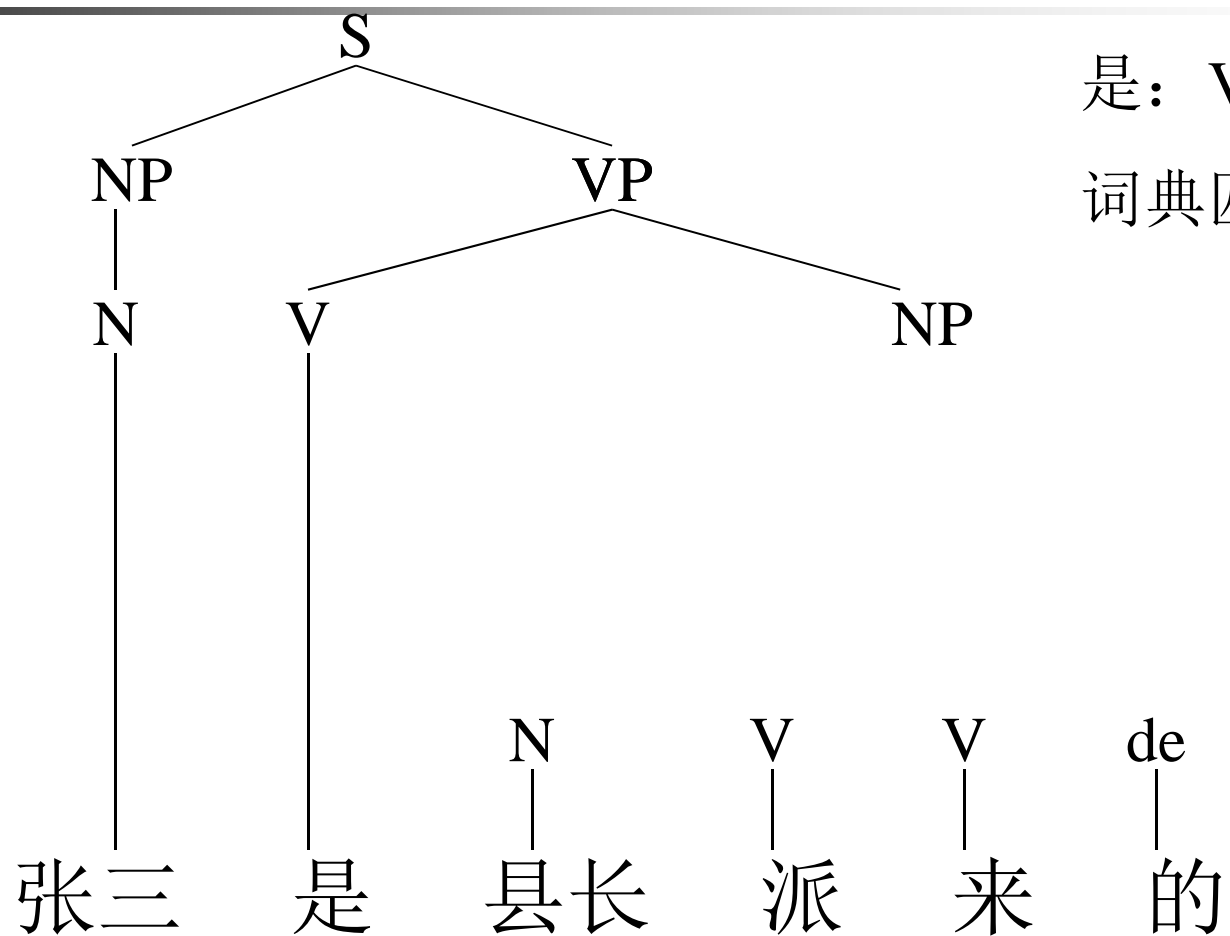


使用规则:

$VP \rightarrow V NP$

自顶向下分析法示例-6

- (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) $V' \rightarrow V V$



是: V

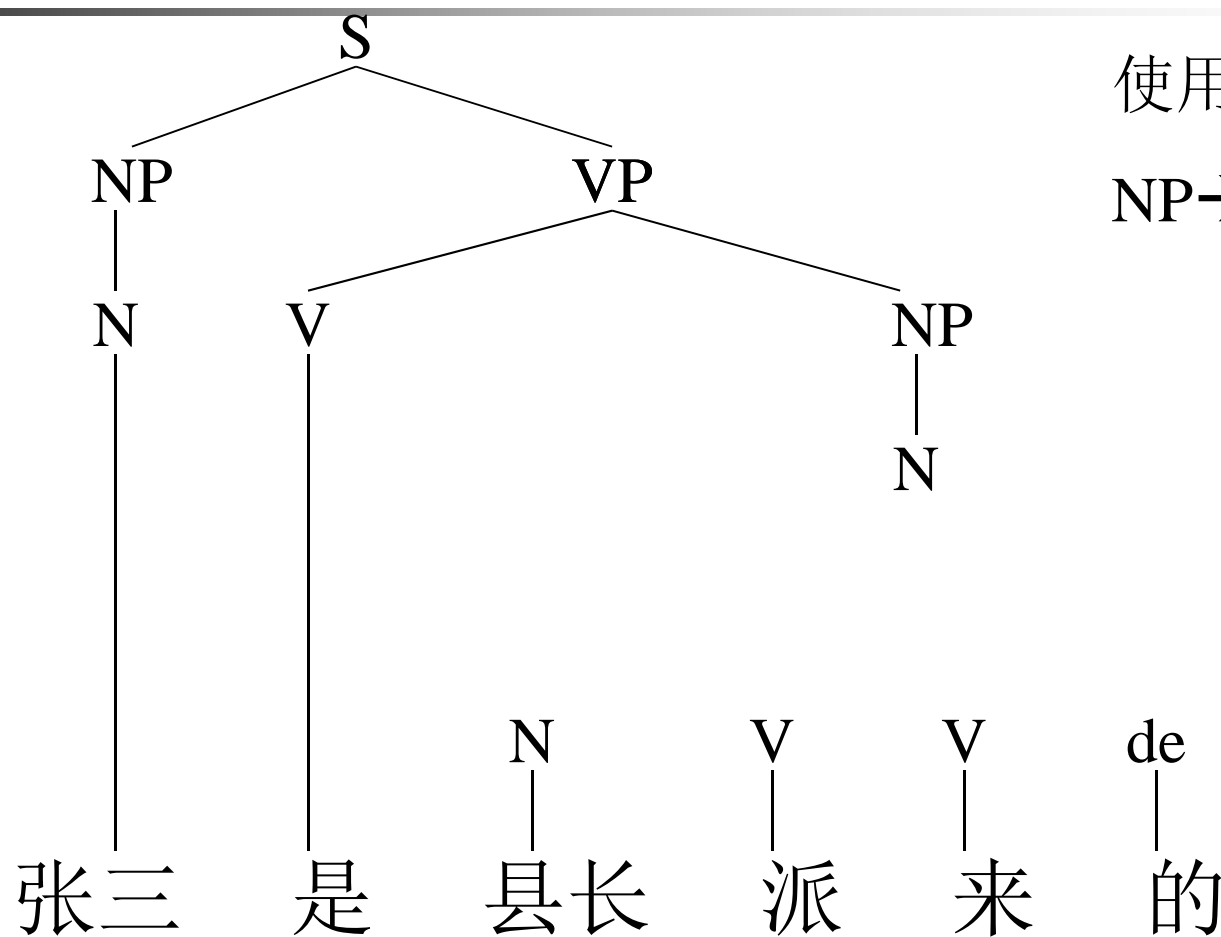
词典匹配成功

自顶向下分析法示例-7

- (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) $V' \rightarrow V V$

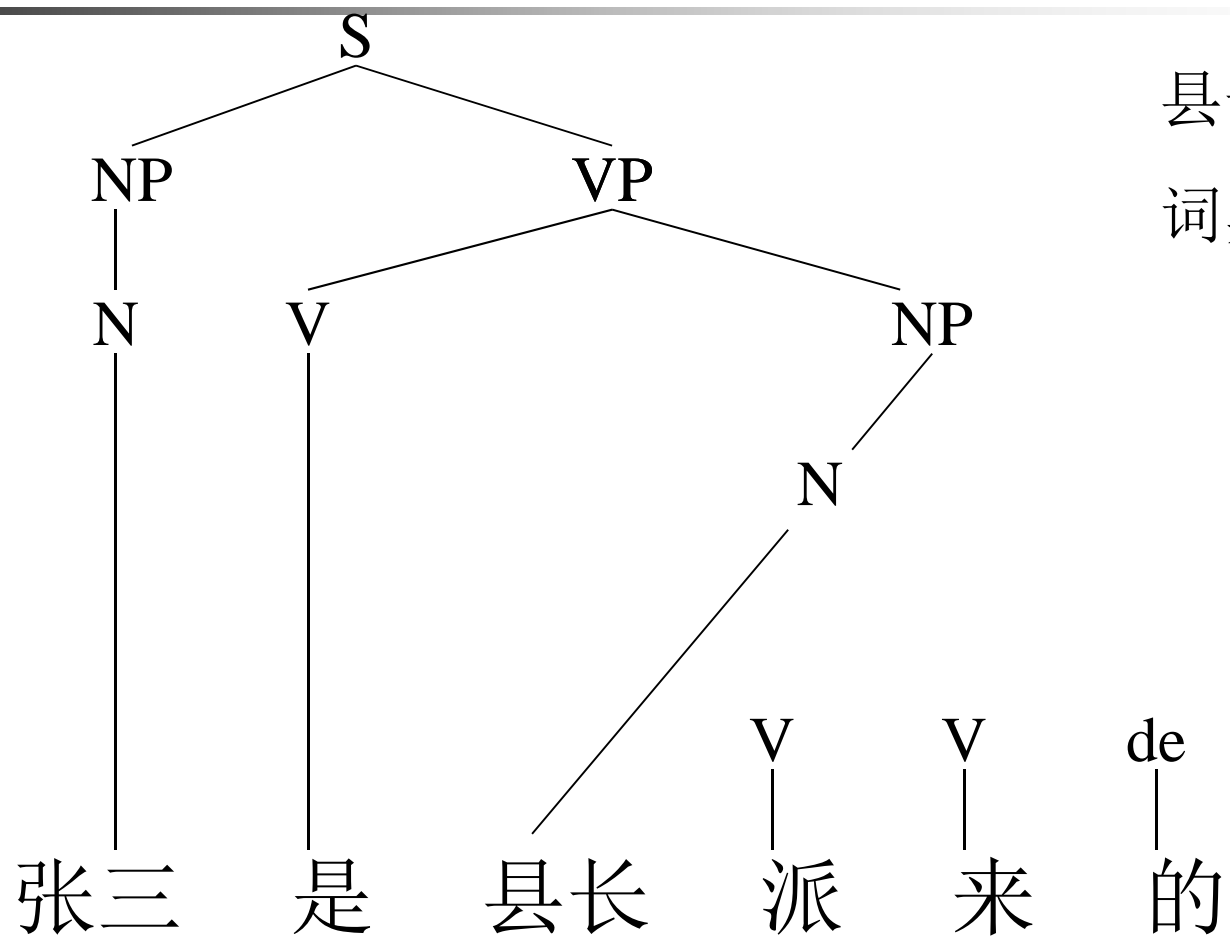
使用规则:

$NP \rightarrow N$



自顶向下分析法示例-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) $V' \rightarrow V V$

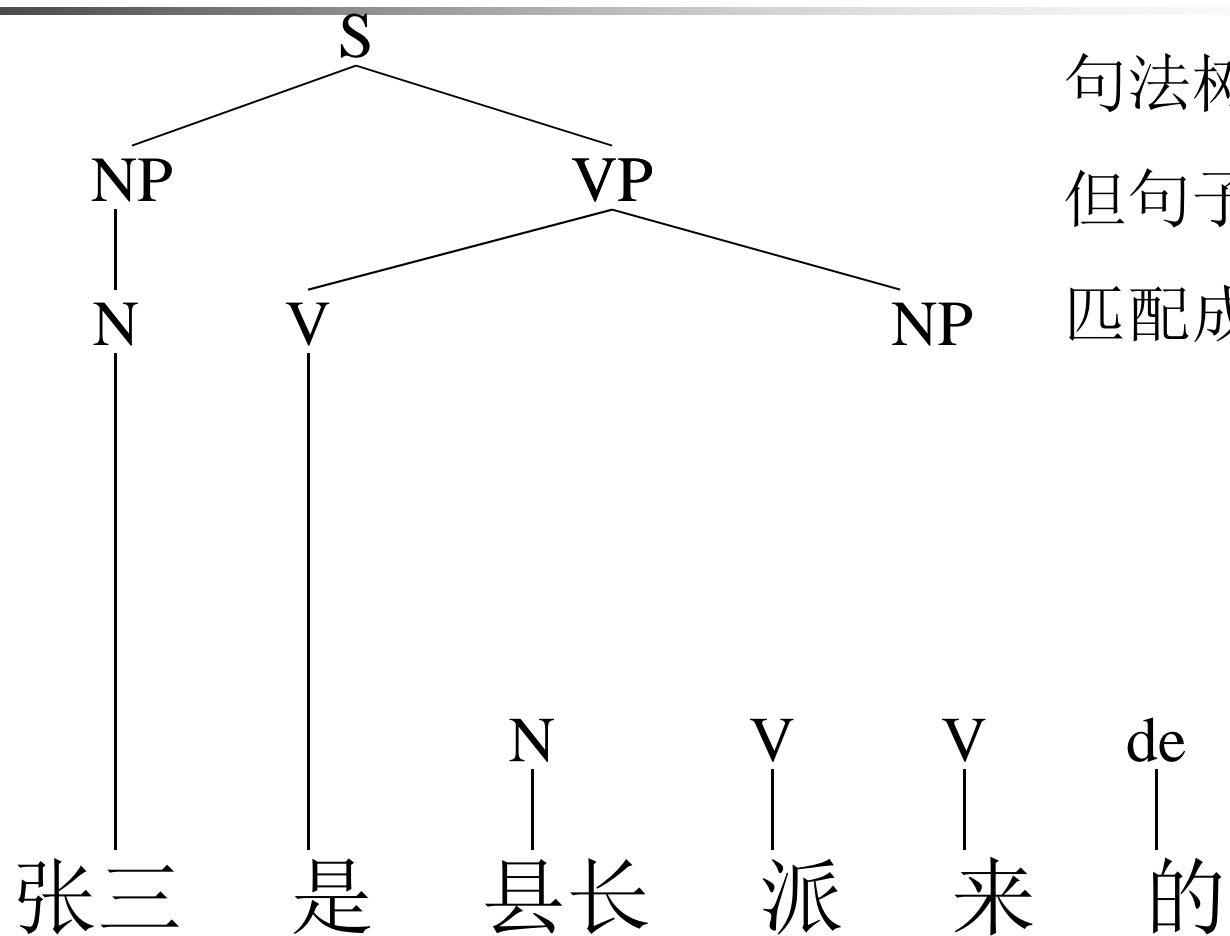


县长: N

词典匹配成功

自顶向下分析法示例-9

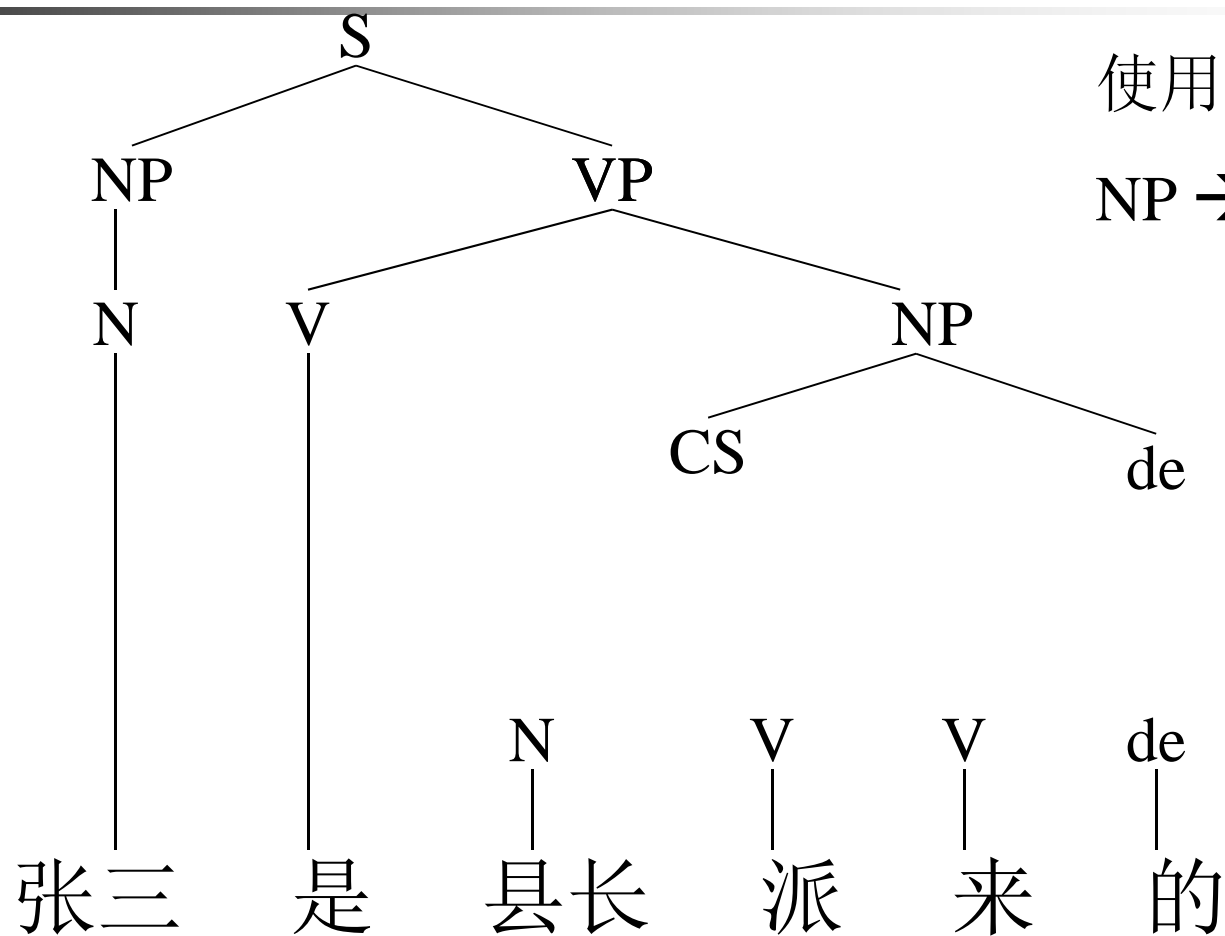
- (1) $S \rightarrow NP VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS$ 的
- (4) $CS \rightarrow NP V'$
- (5) $VP \rightarrow V NP$
- (6) $V' \rightarrow V V$



句法树扩展完毕，
但句子没有完全
匹配成功，回溯

自顶向下分析法示例-10

- (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) $V' \rightarrow V V$

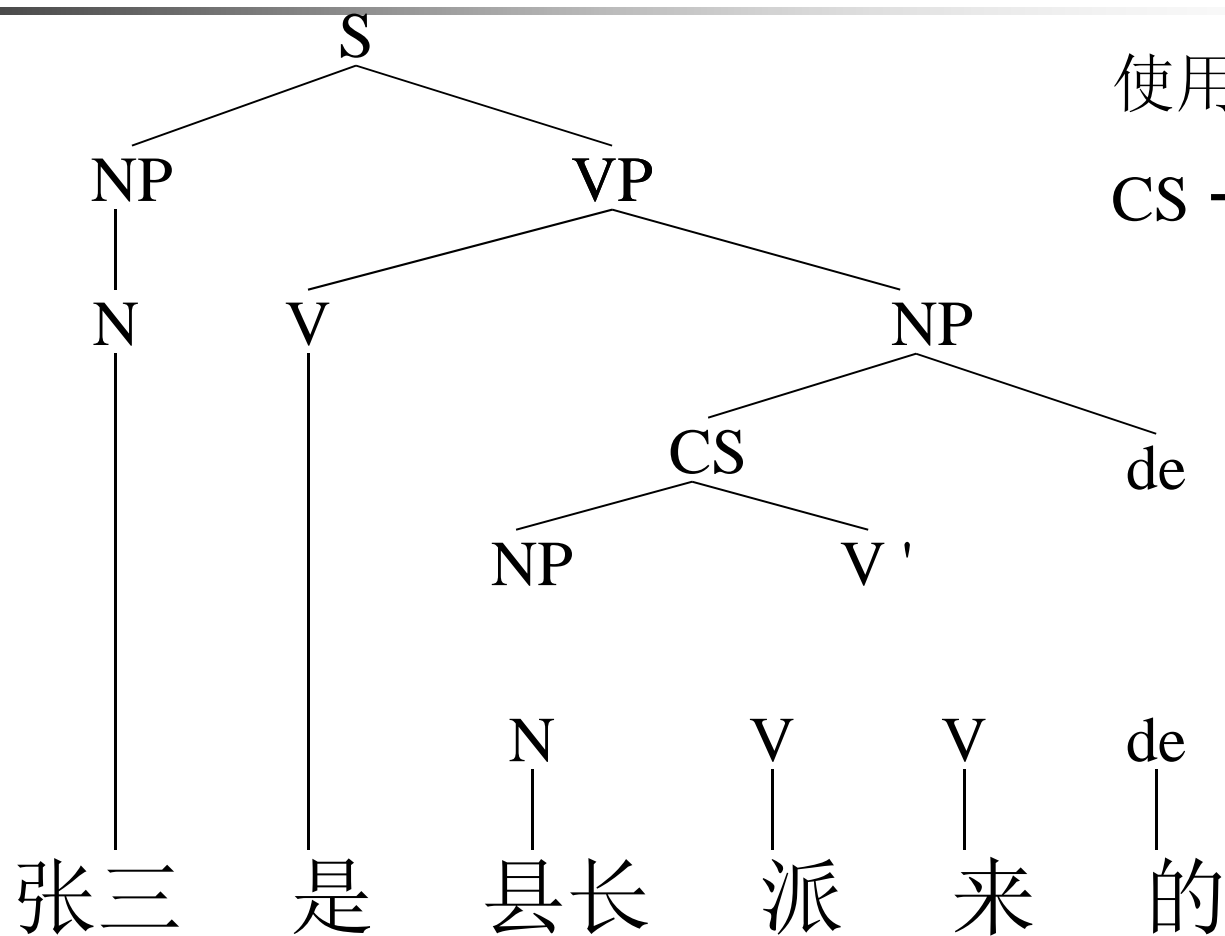


使用规则:

$NP \rightarrow CS$ de

自顶向下分析法示例-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:

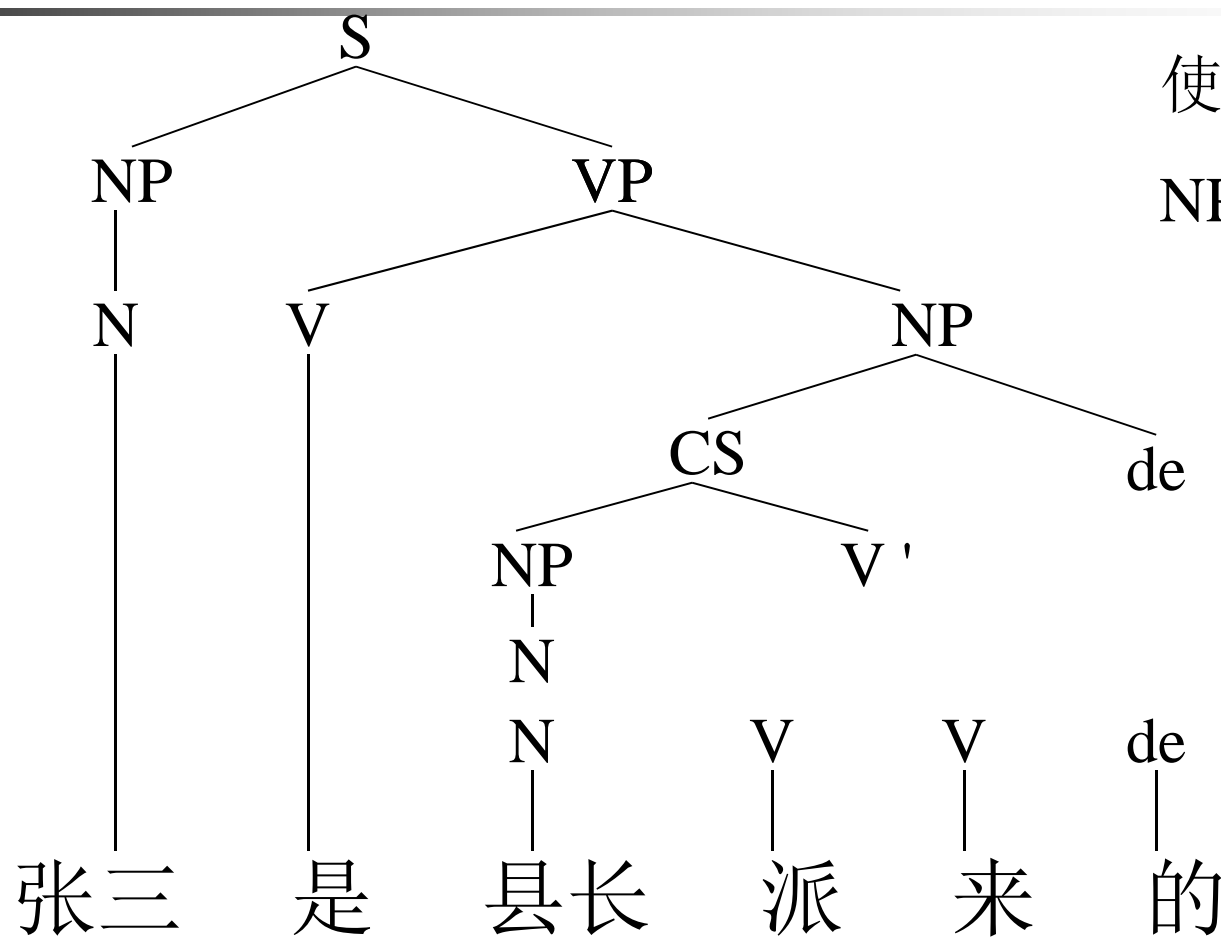
$CS \rightarrow NP V'$

自顶向下分析法示例-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) $V' \rightarrow V V$

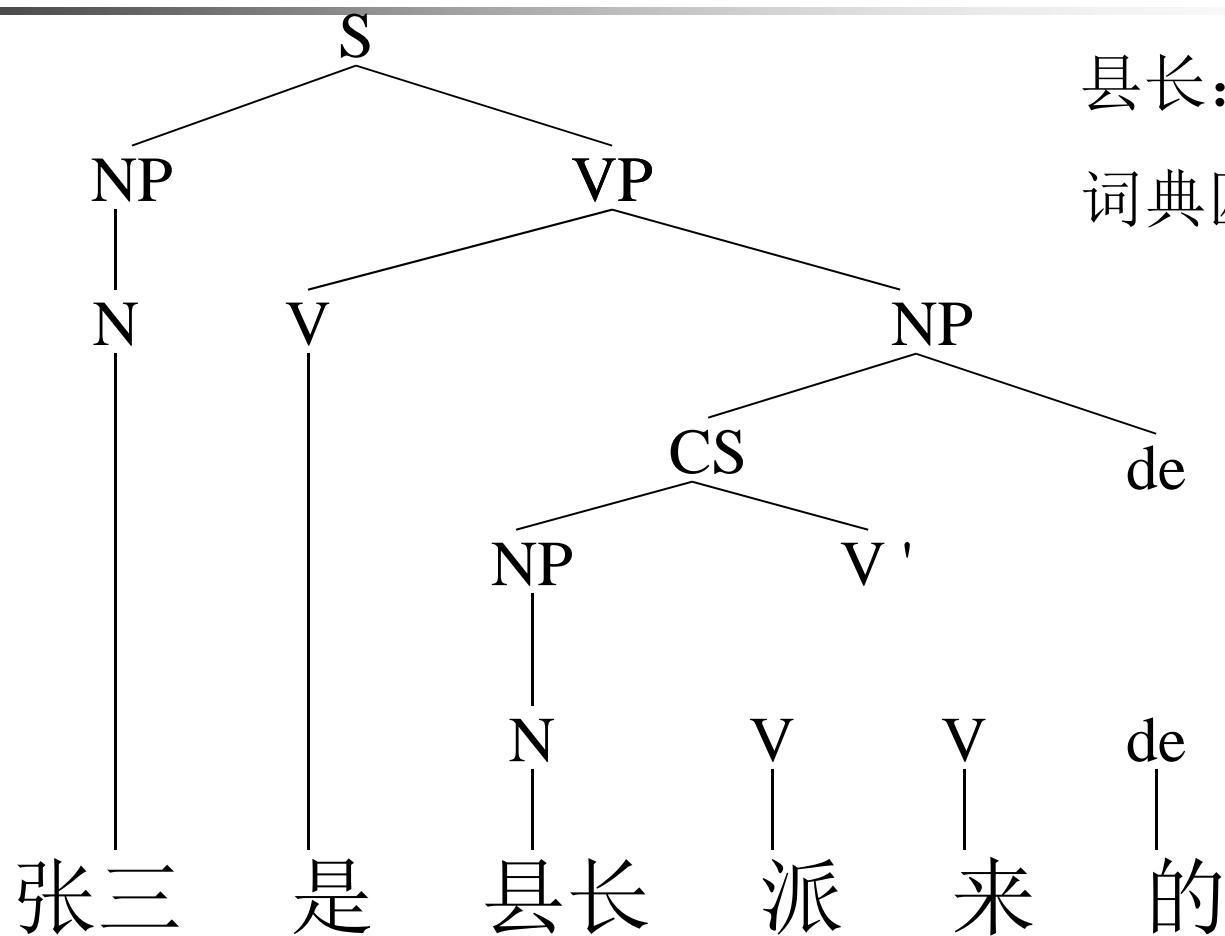
使用规则:

$NP \rightarrow N$



自顶向下分析法示例-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) $V' \rightarrow V V$

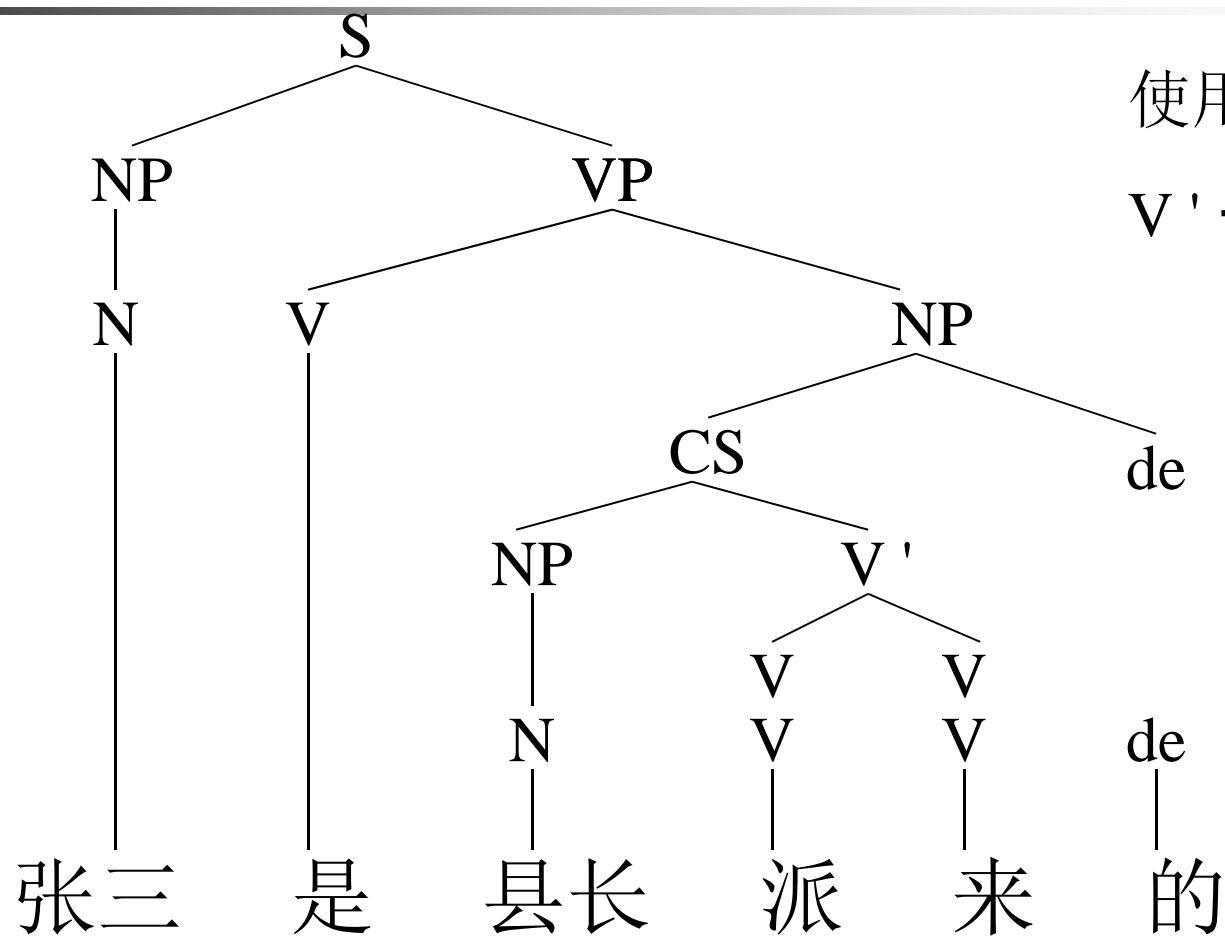


县长: N

词典匹配成功

自顶向下分析法示例-14

- (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) $V' \rightarrow V V$

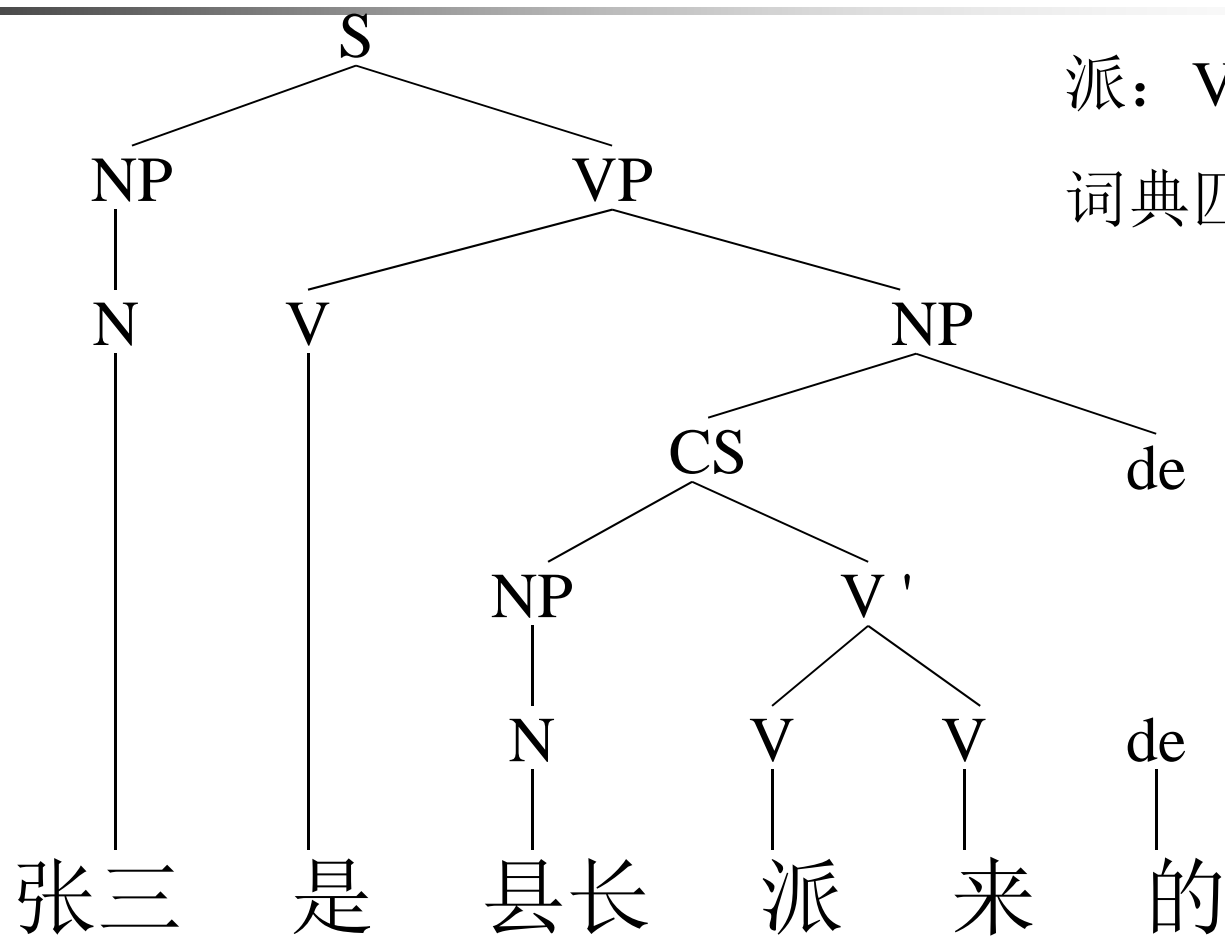


使用规则:

$V' \rightarrow V V$

自顶向下分析法示例-15

- (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) $V' \rightarrow V V$

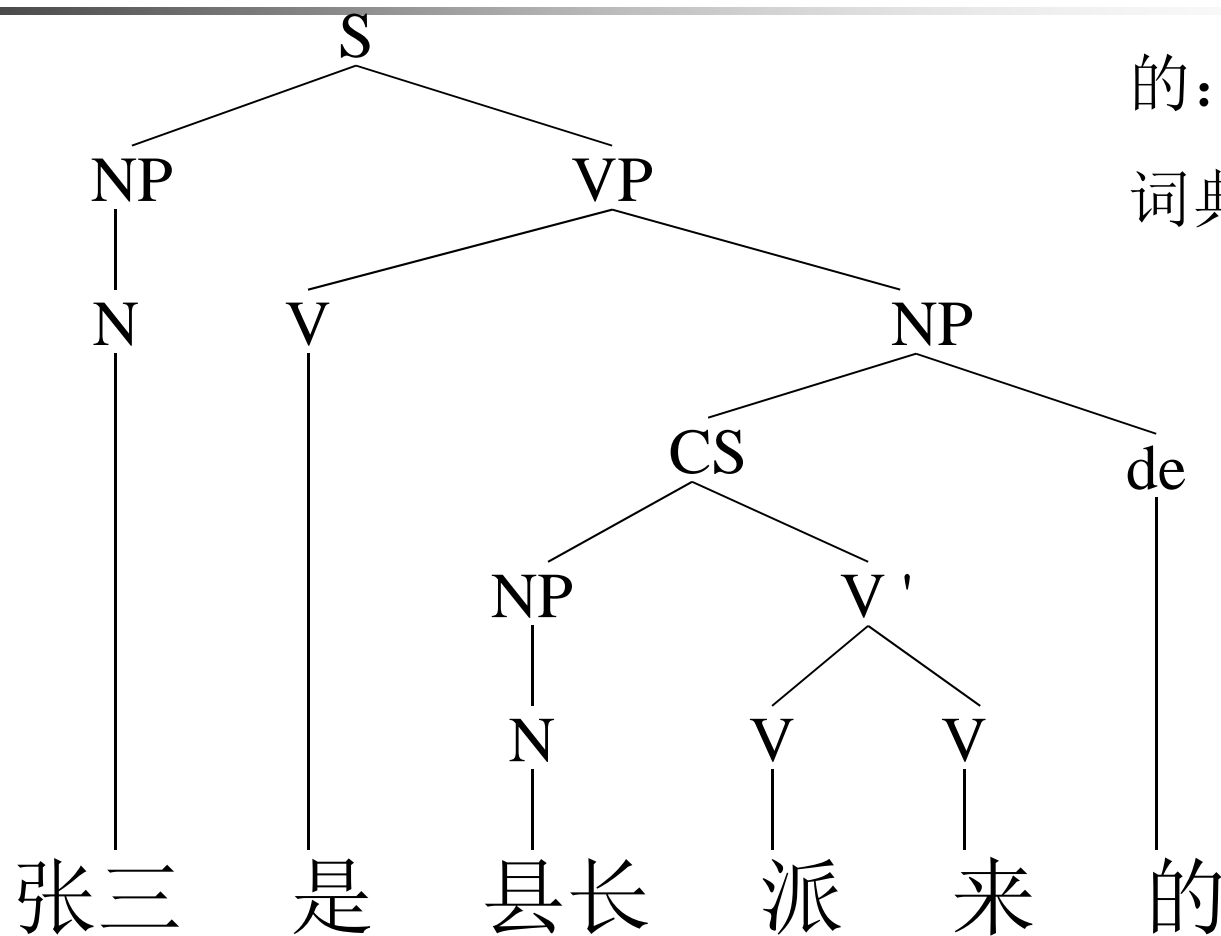


派: V 来: V

词典匹配成功

自顶向下分析法示例-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) $V' \rightarrow V V$



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。



自底向上分析法示例-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

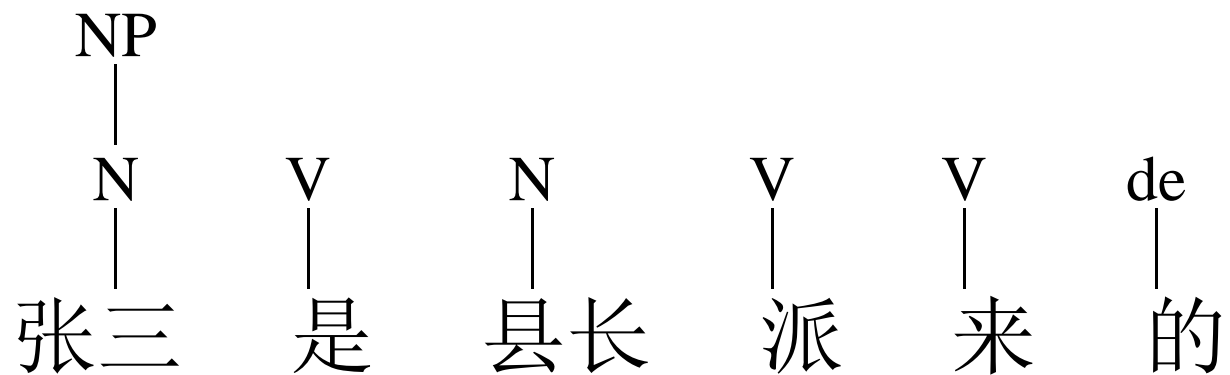
N V N V V de
| | | | | |
张三 是 县长 派 来 的

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:

$NP \rightarrow N$

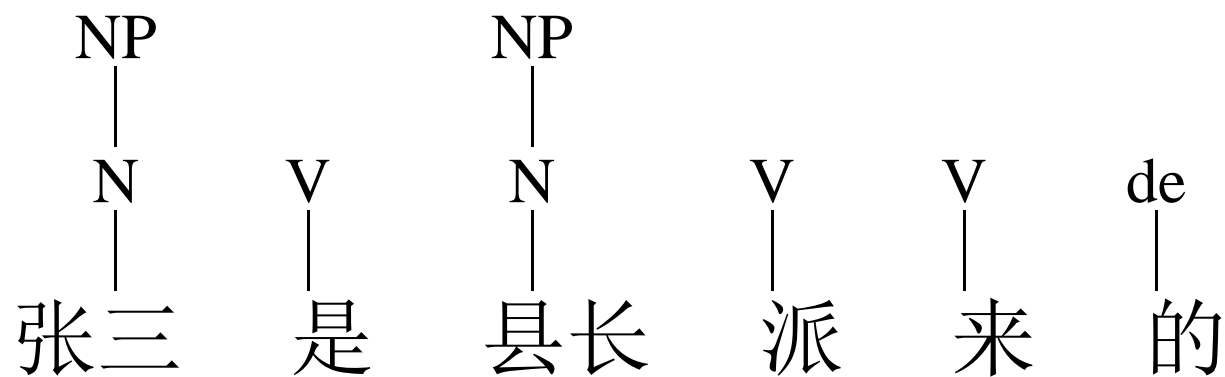


自底向上分析法示例-3

- (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:

$NP \rightarrow N$

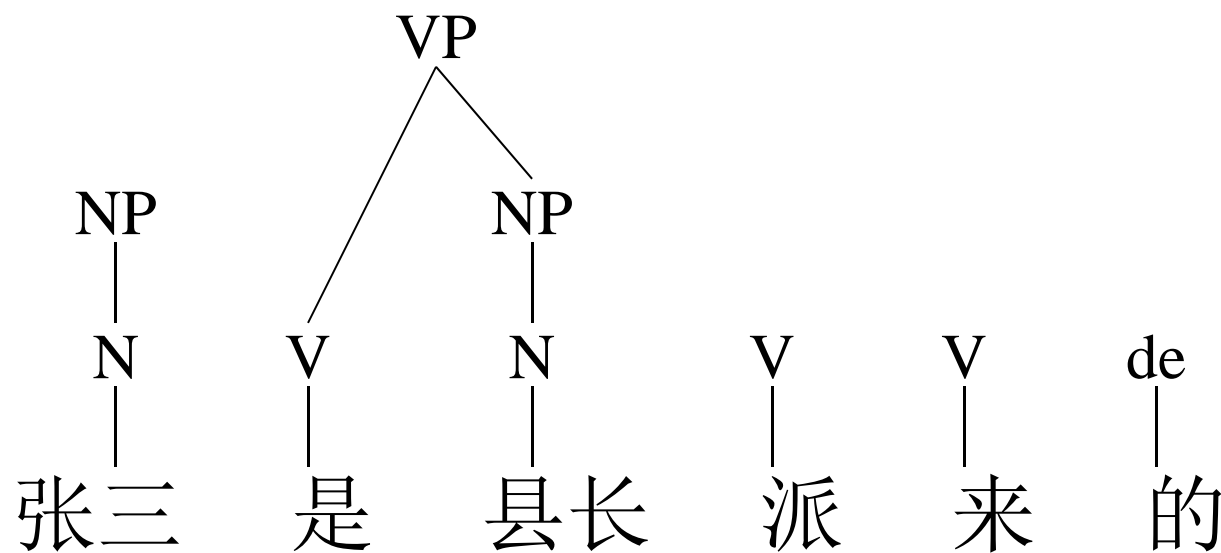


自底向上分析法示例-4

- (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:

$VP \rightarrow V NP$

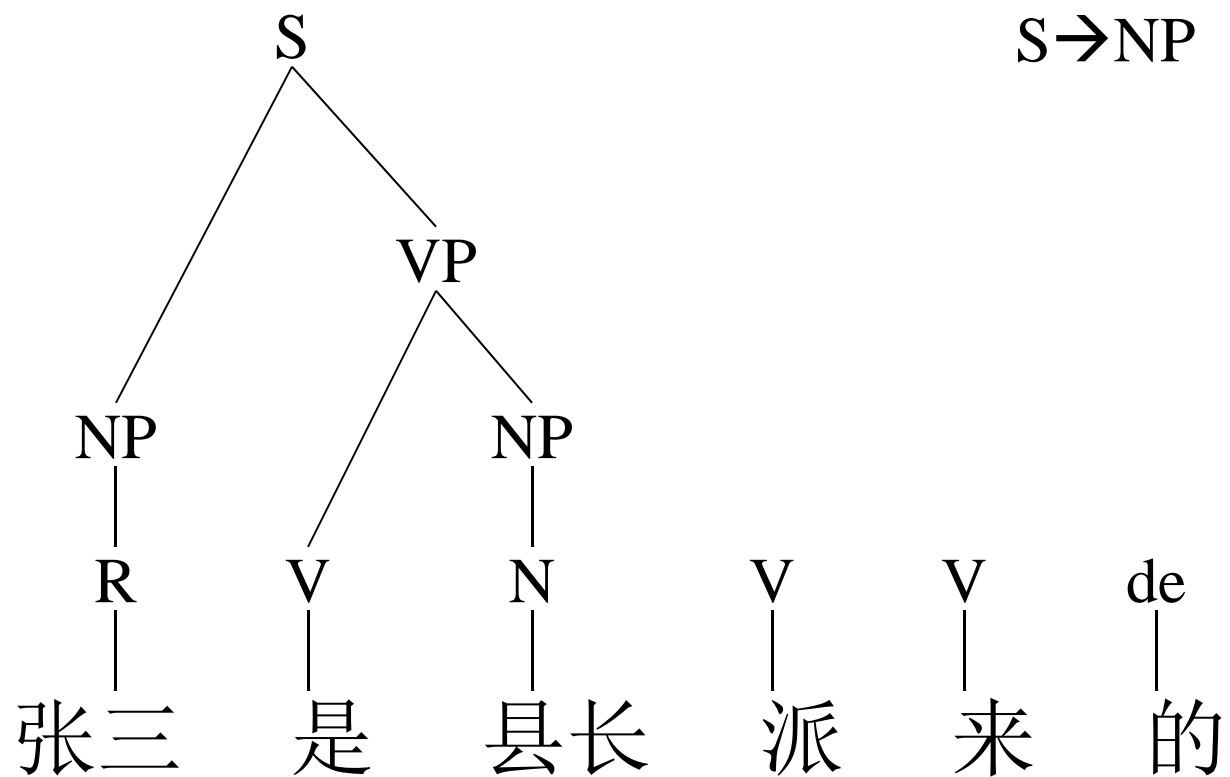


自底向上分析法示例-5

- (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:

$S \rightarrow NP VP$

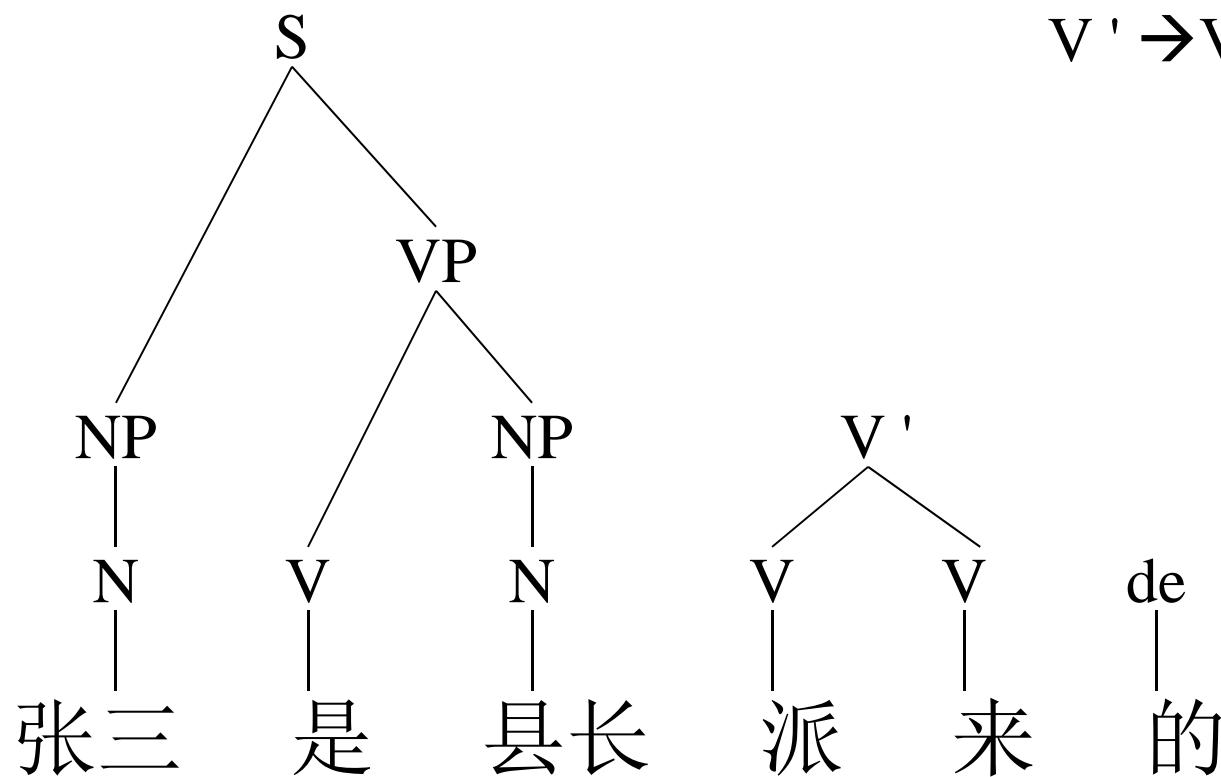


自底向上分析法示例-6

- (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:

$V' \rightarrow V V$

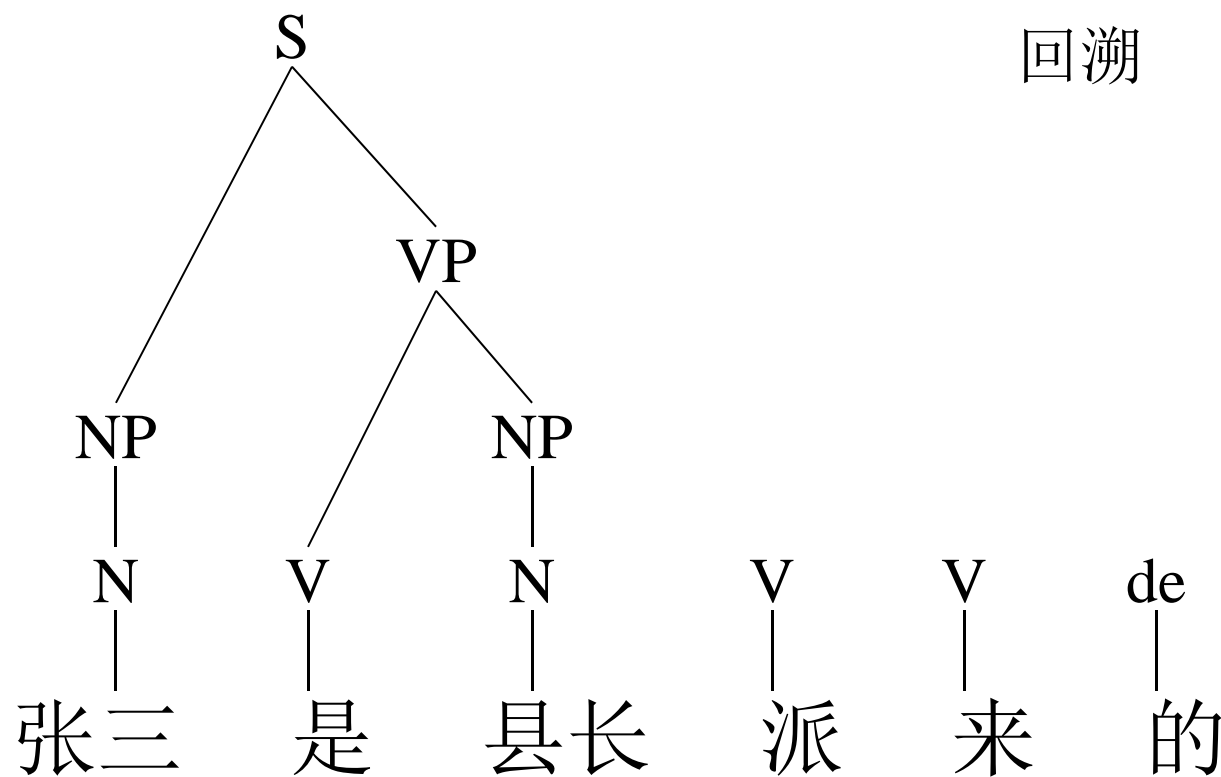


自底向上分析法示例-7

- (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

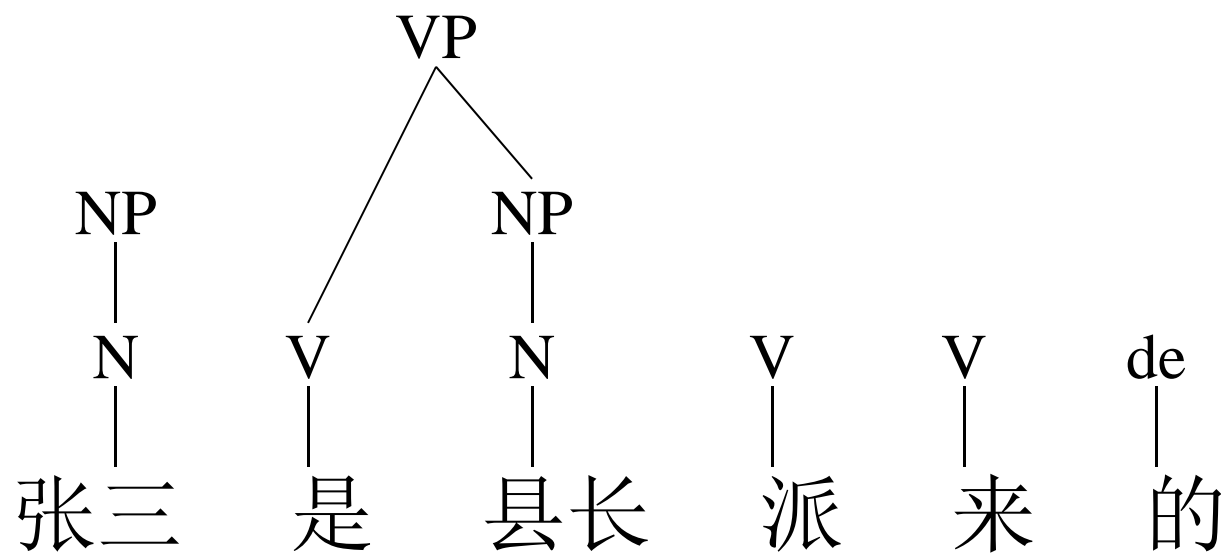


自底向上分析法示例-8

- (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

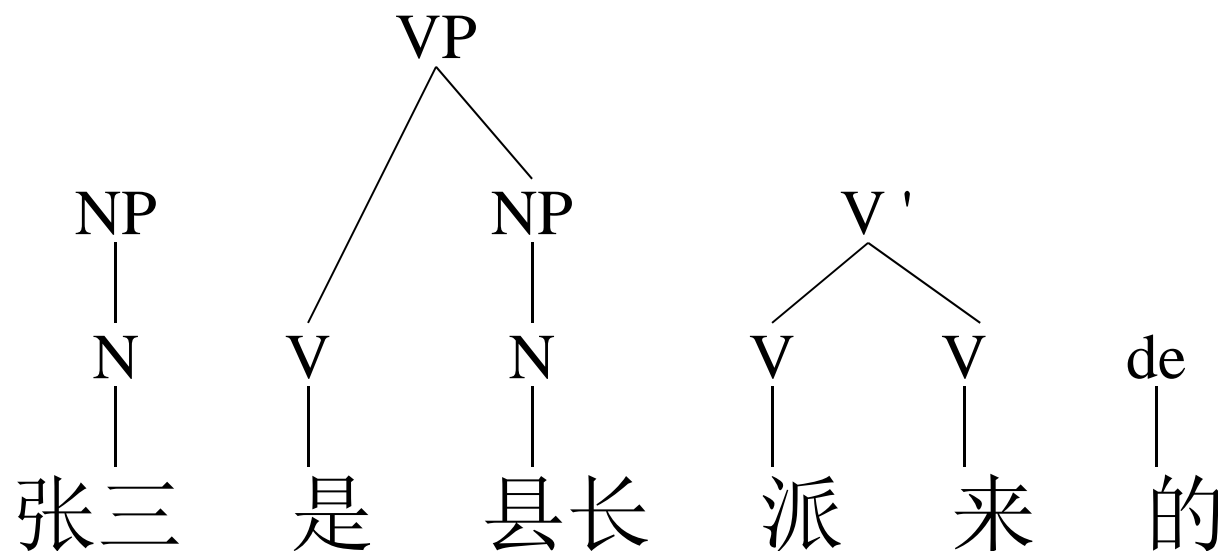


自底向上分析法示例-9

- (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:

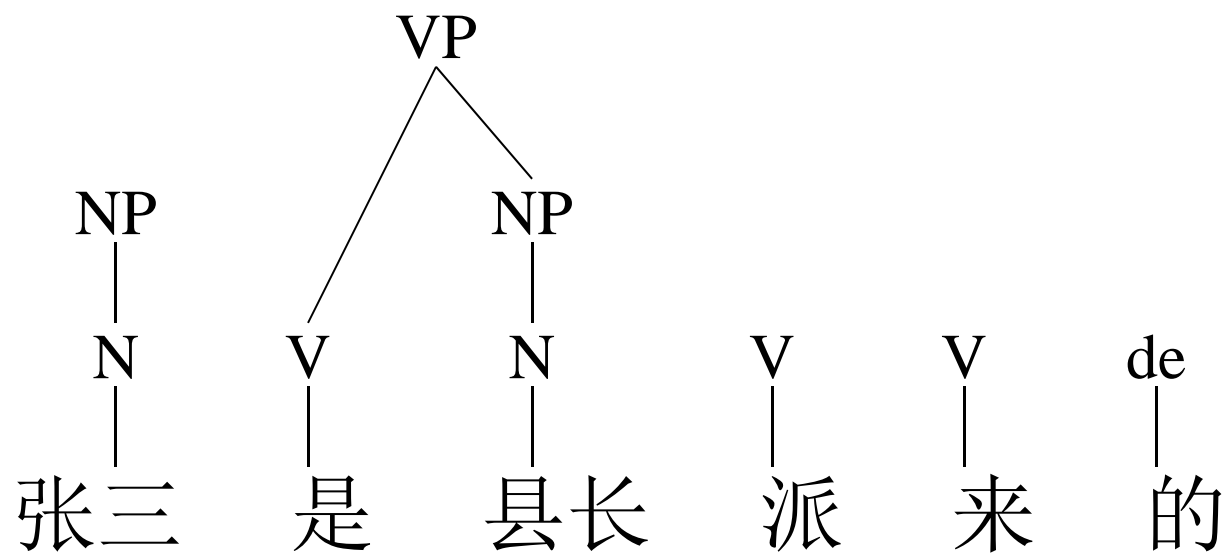
$V' \rightarrow V V$



自底向上分析法示例-10

- (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) $V' \rightarrow V V$

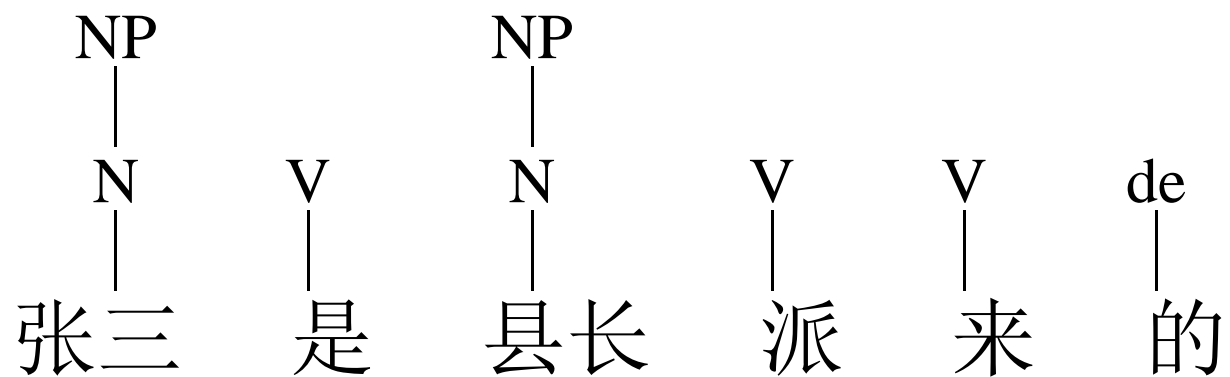
无规则可用，
回溯



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

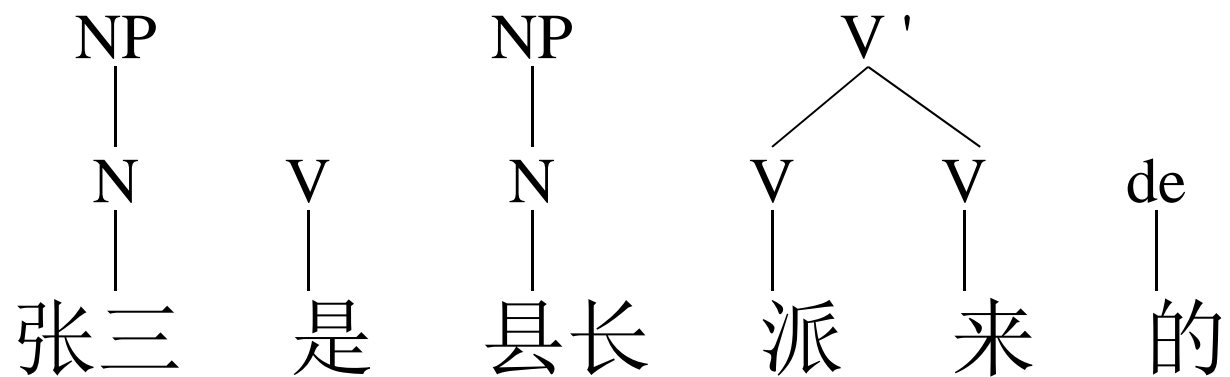


自底向上分析法示例-12

- (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) $V' \rightarrow V V$

使用规则:

$V' \rightarrow V V$

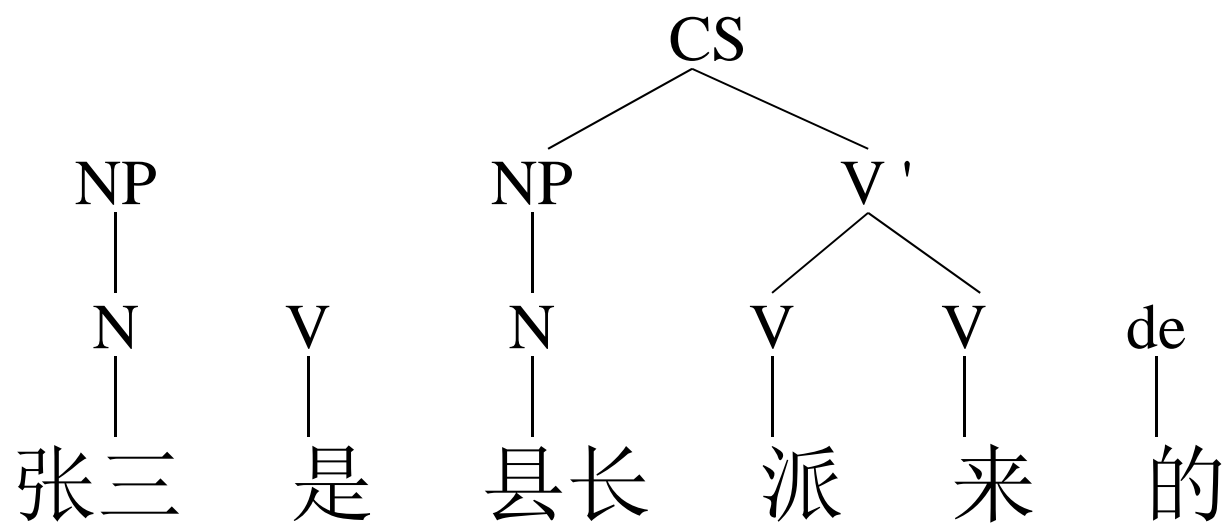


自底向上分析法示例-13

- (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) $V' \rightarrow V V$

使用规则:

$CS \rightarrow NP V'$

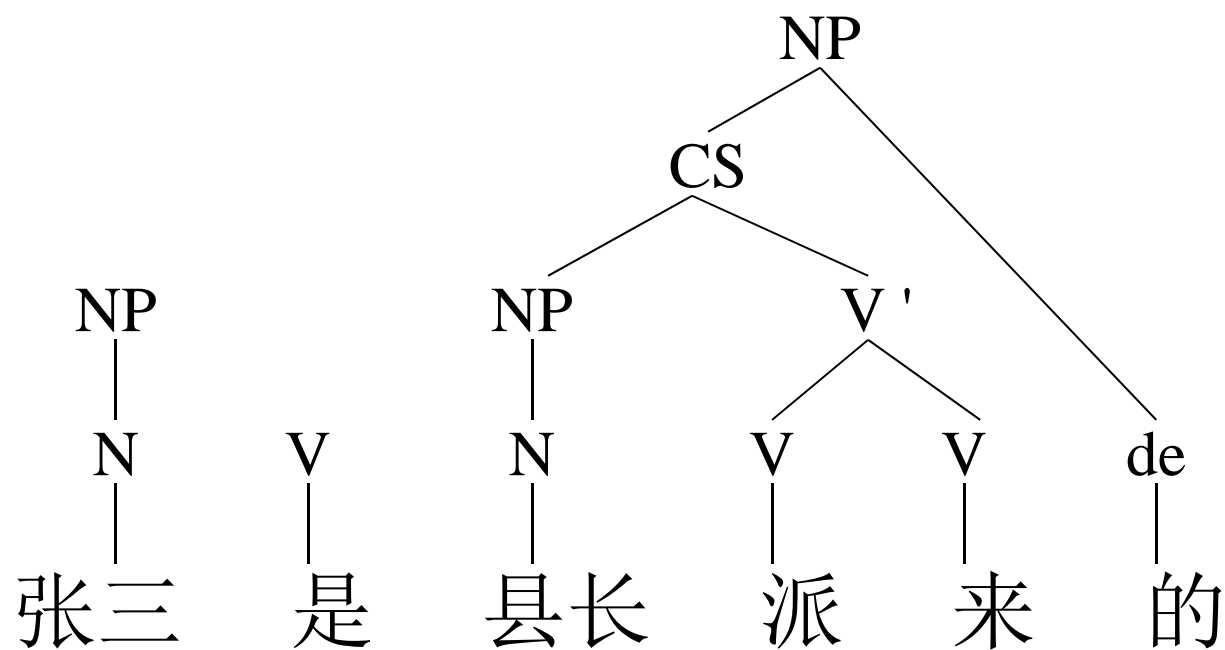


自底向上分析法示例-14

- (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:

$NP \rightarrow CS$ de

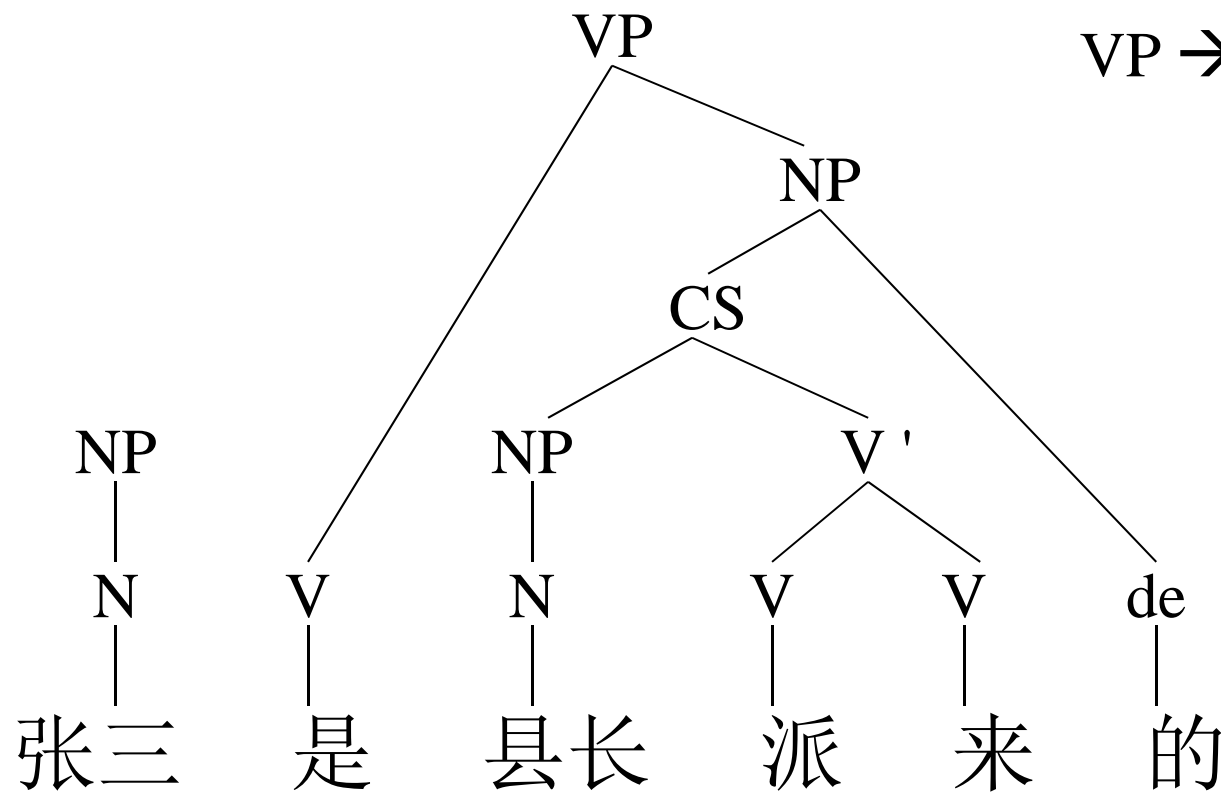


自底向上分析法示例-15

- (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) $V' \rightarrow V V$

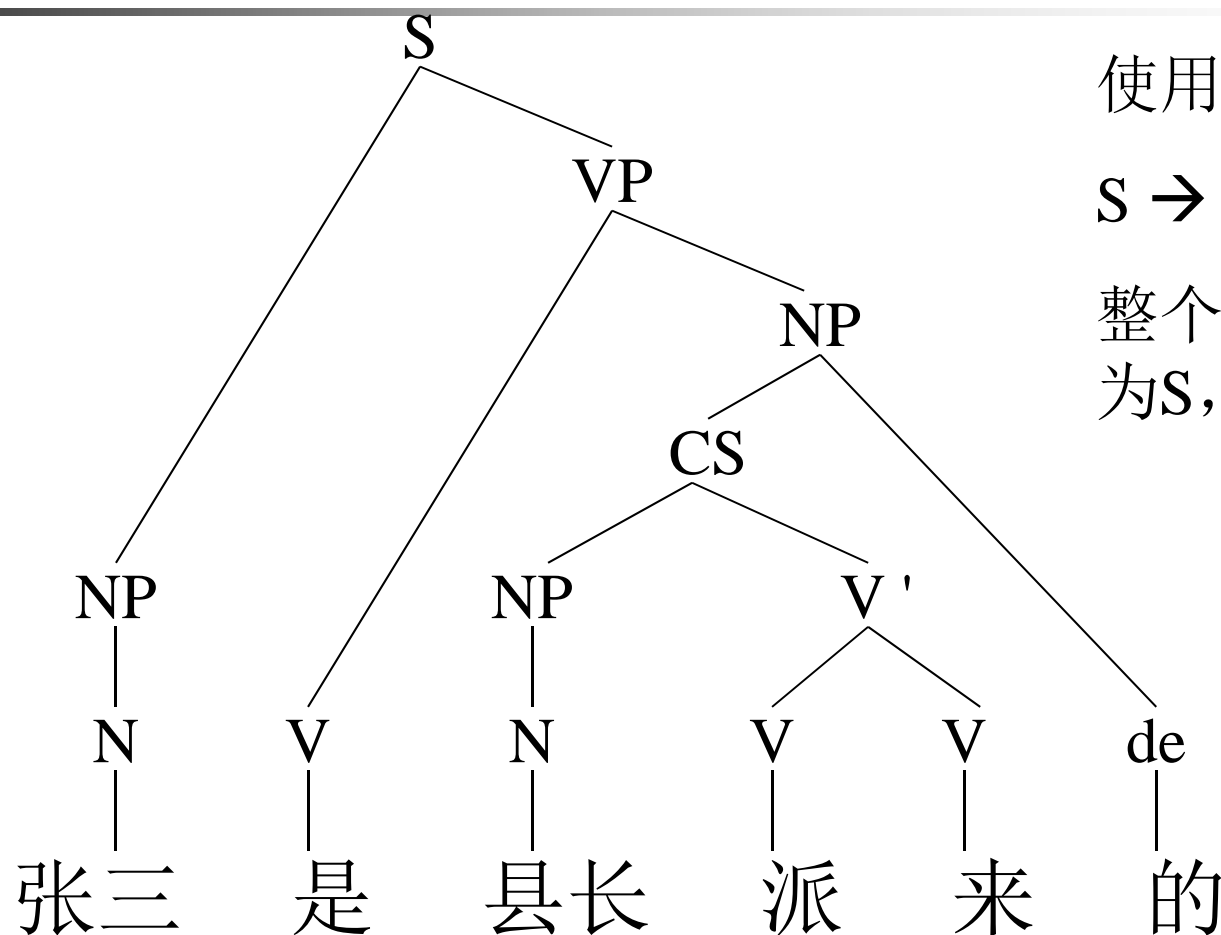
使用规则:

$VP \rightarrow V NP$



自底向上分析法示例-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) $V' \rightarrow V V$



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。

4 小结

- 语言模型：保证句法结构分析的准确
- 分析算法：提升句法结构分析的效率

把事情做对
把事情做好

LFG（词汇功能语法）

语言模型

FUG（功能合一语法）

HPSG（中心驱动的短语结构语法）

TAG（树邻接语法）

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

Earley算法

GLR算法

线图分析算法

链语法分析算法

依存句法分析算法

.....



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
- Dick Grune & Cerial Jacobs, 1990, Parsing Techniques: A Practical Guide, First Published in 1990 by ELLIS HORWOOD LIMITED, Reprinted in 1997, 1998



复习思考题

- 1 构造一个适用于部分汉语句子的汉语CFG文法，用CYK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。