D.H.Younger, 1967, Recognition of context-free language in time n^3, Information Control, vol. 10, no.2.

CYK分析算法

詹卫东
http://ccl.pku.edu.cn/doubtfire

提 纲

一 句法结构分析任务

二 分析算法数据结构

三 分析算法流程

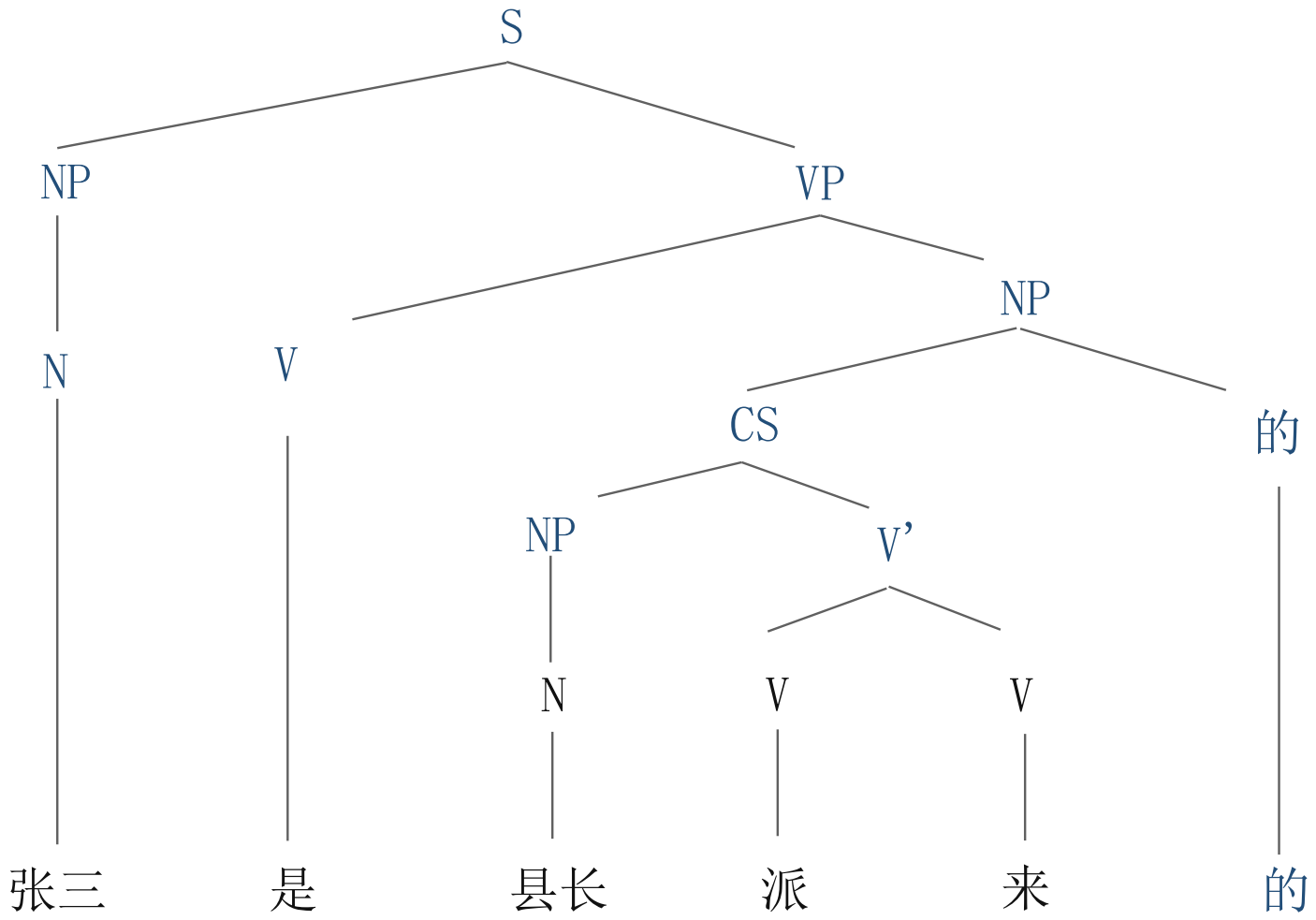
四 分析过程示例

一 句法结构分析任务

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 → NP VP
(2) NP → N
(3) NP → CS 的
(4) VP → V NP
(5) CS → NP V '
(6) V ' → V V

一 句法结构分析任务



一 句法结构分析任务

CYK算法分析结果的表格表示

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

二 CYK分析算法数据结构

- CYK算法：Cocke-Younger-Kasami算法
- CYK算法的描述基于Chomsky范式的CFG规则
  - Chomsky范式的规则只有两种形式：
    A→BC
    A→x　　这里A,B,C是非终结符，x是终结符
  - 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

二 CYK分析算法数据结构

- 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*：该跨度包含的词的个数。

P(3,1)={NP,N}表示“县长”可以归约成N、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。

三 CYK算法流程

1. 对i=1…n，j=1（填写第一行，跨度为1）

   对于每一条规则A→Wi，将非终结符A加入集合P(i,j);

2. 对j=2…n（填写第2到n行，跨度为j）

   对i=1…n-j+1（对于所有起点i）

   对k=1…j-1（对于一个跨度内所有分割点k）

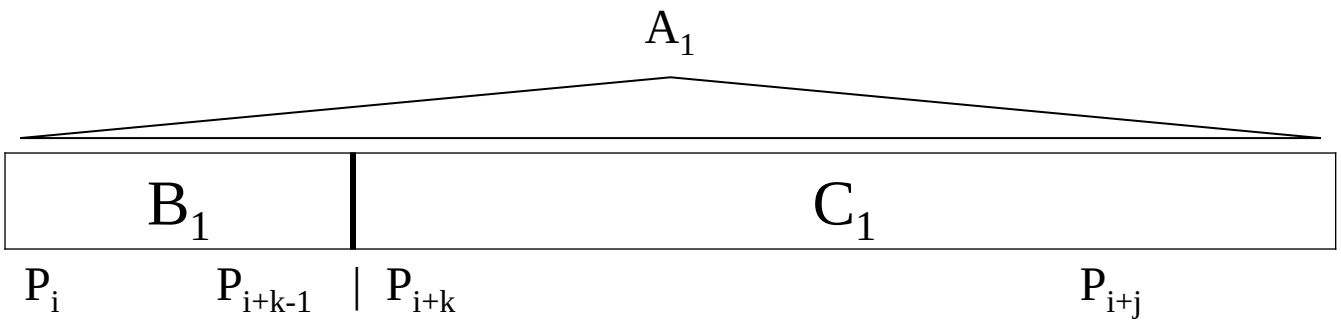
   对每一条规则 A→B C，

   如果B ∈ P(i,k) 且 C∈P(i+k, j-k)

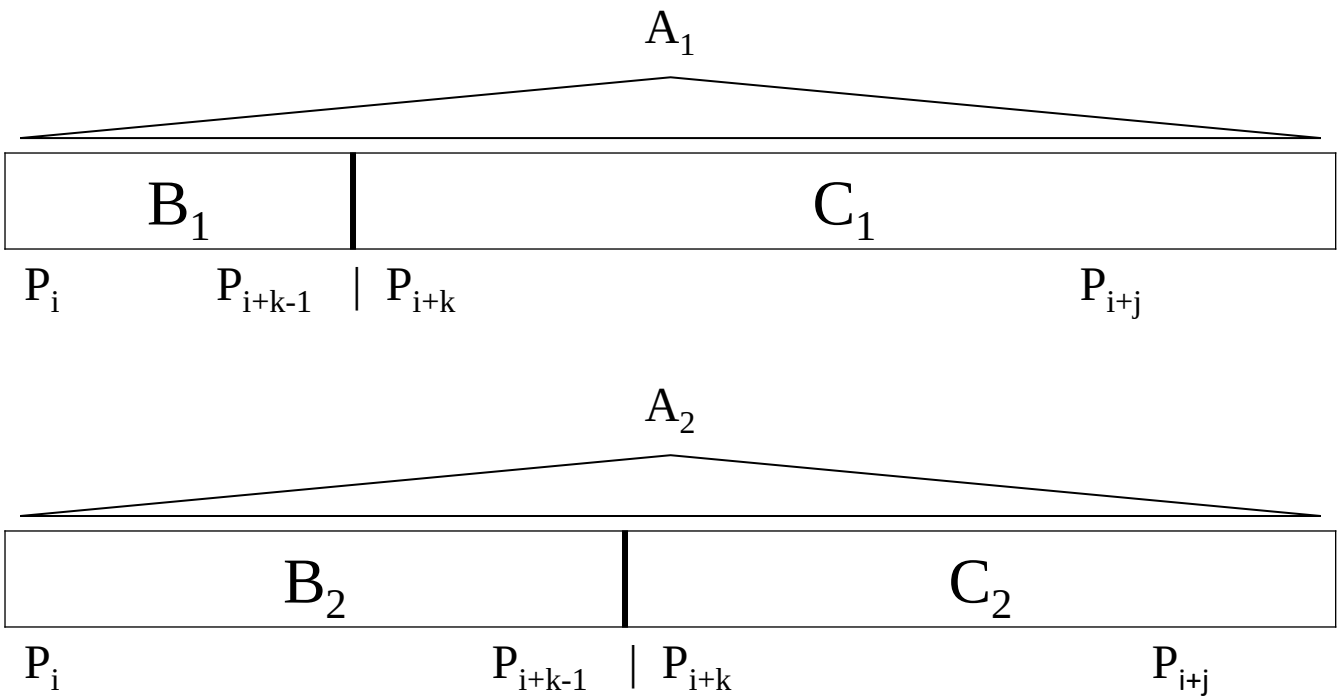
   那么将非终结符A加入集合P(i,j)

3. 如果S∈P(1,n)，那么分析成功，否则分析失败

CYK算法成分组合示意



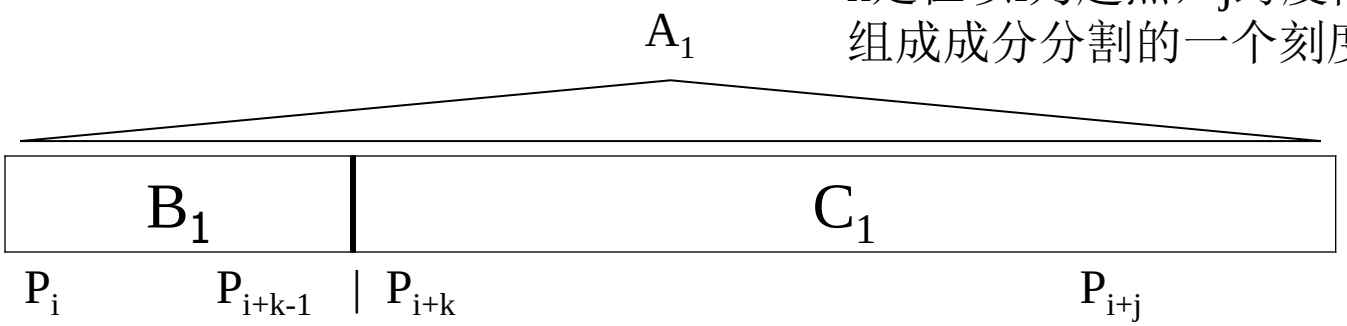
CYK算法成分组合示意

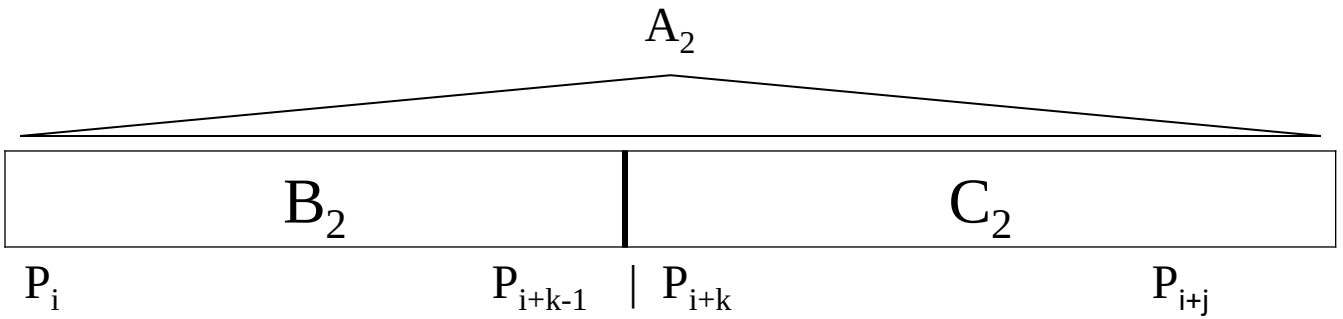


……

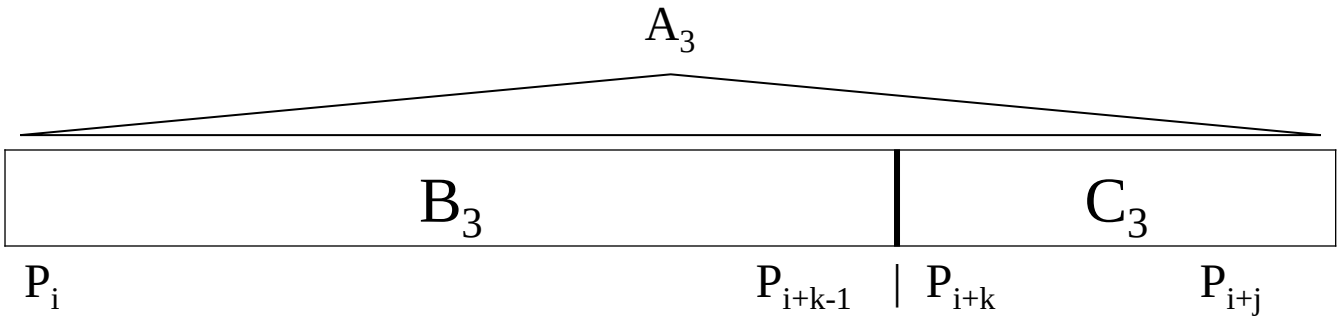
CYK算法成分组合示意

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



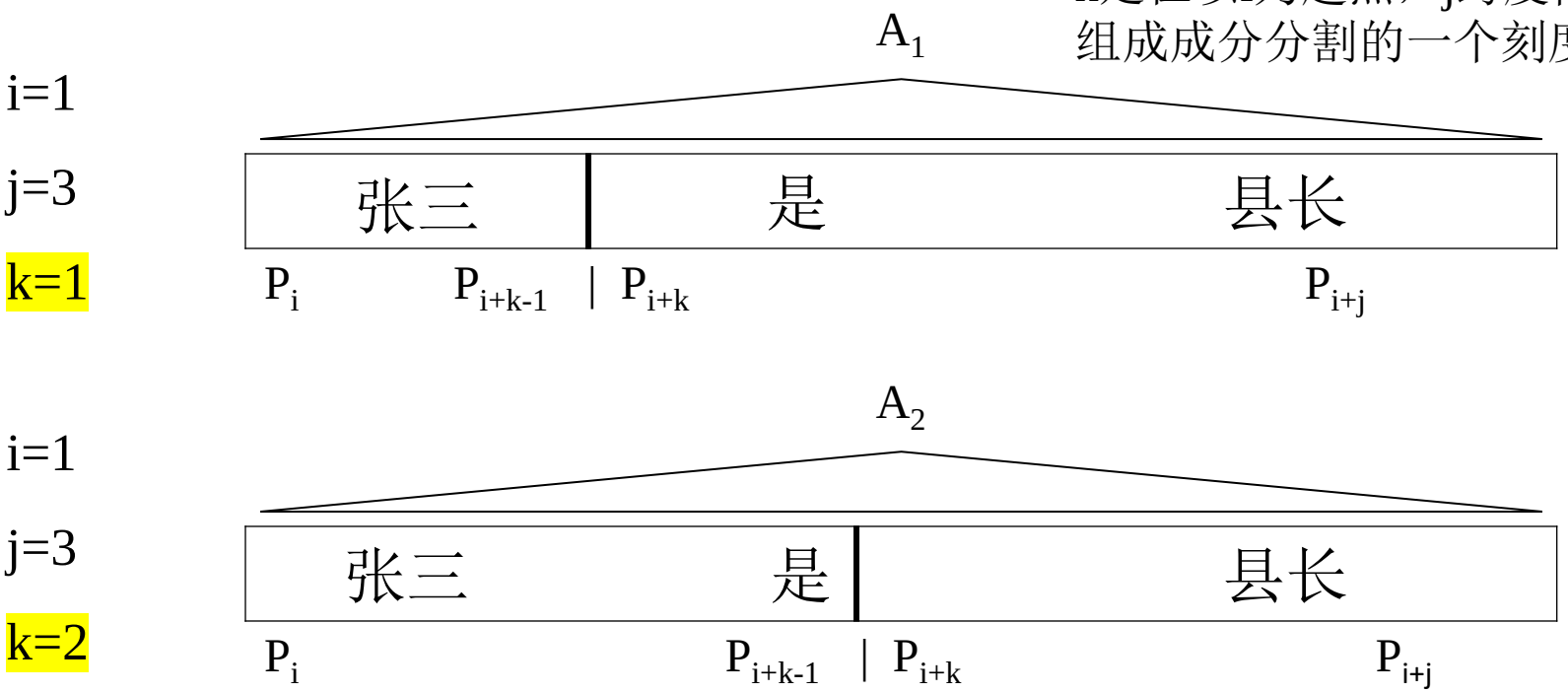


……



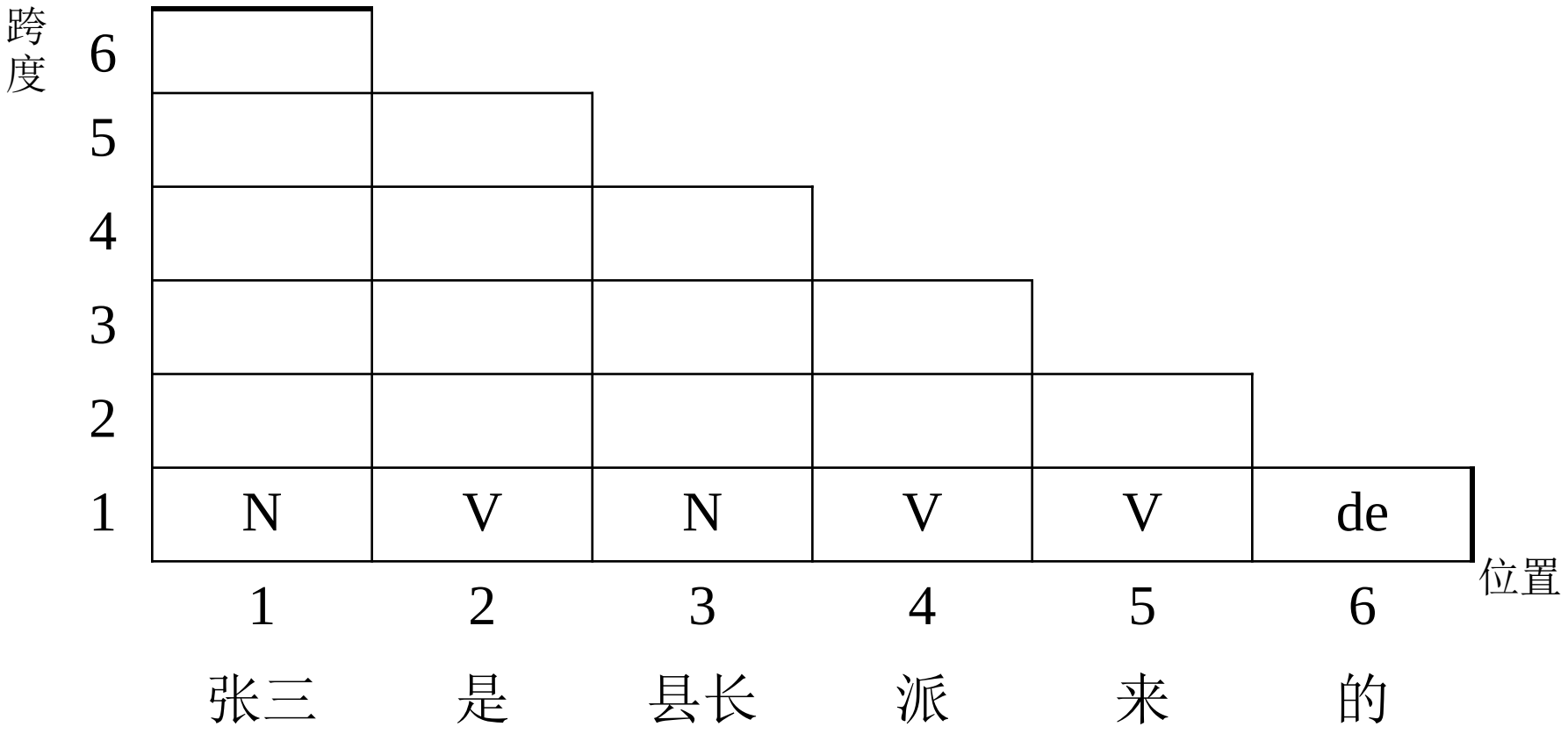
CYK算法成分组合示意

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



四 CYK算法示例-填写第1行

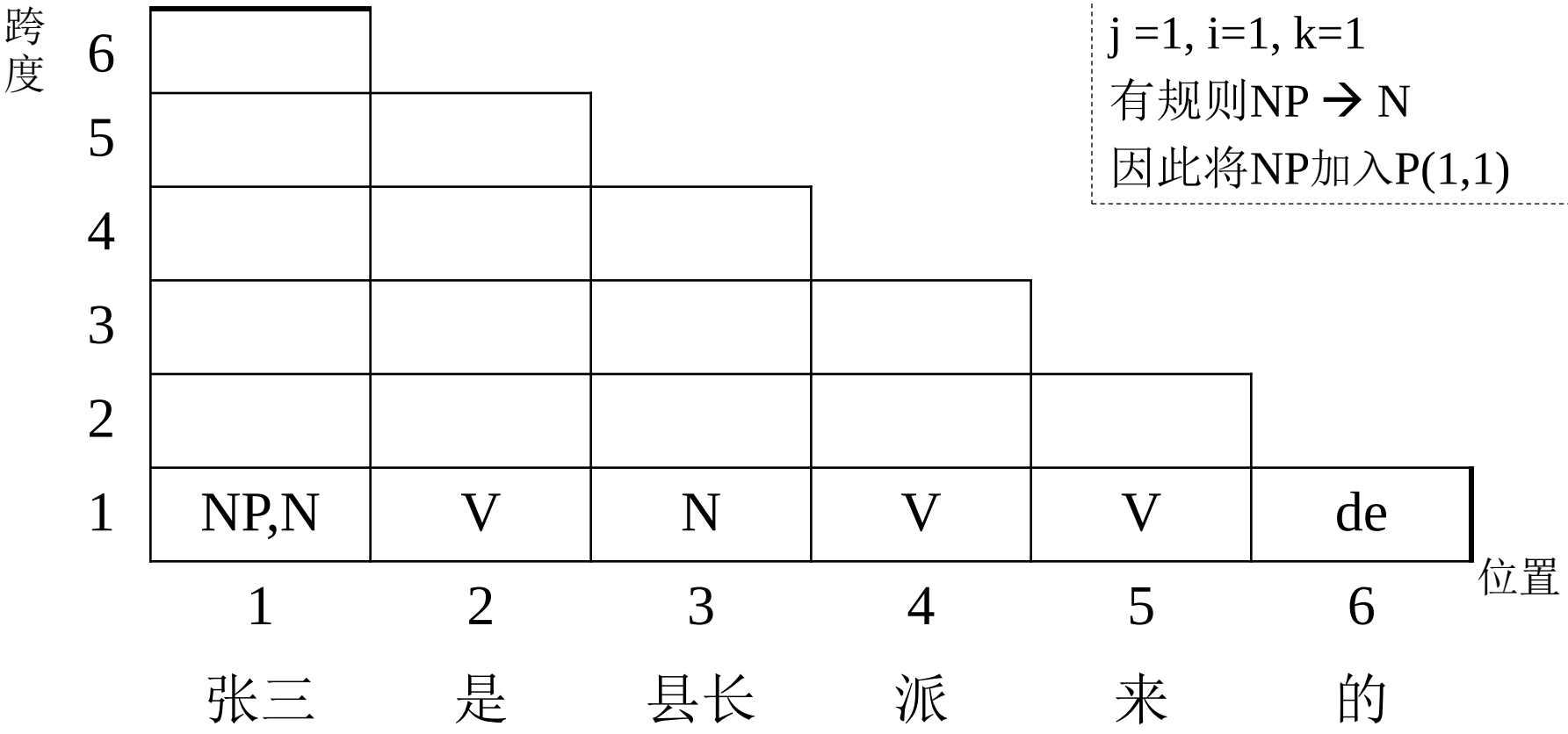
(1) S → NP VP
(2) NP → N
(3) NP → CS 的
(4) CS → NP V '
(5) VP → V NP
(6) V ' → V V



四 CYK算法示例-填写第1行

(1) S → NP VP
(2) NP → N
(3) NP → CS 的
(4) CS → NP V'
(5) VP → V NP
(6) V' → V V

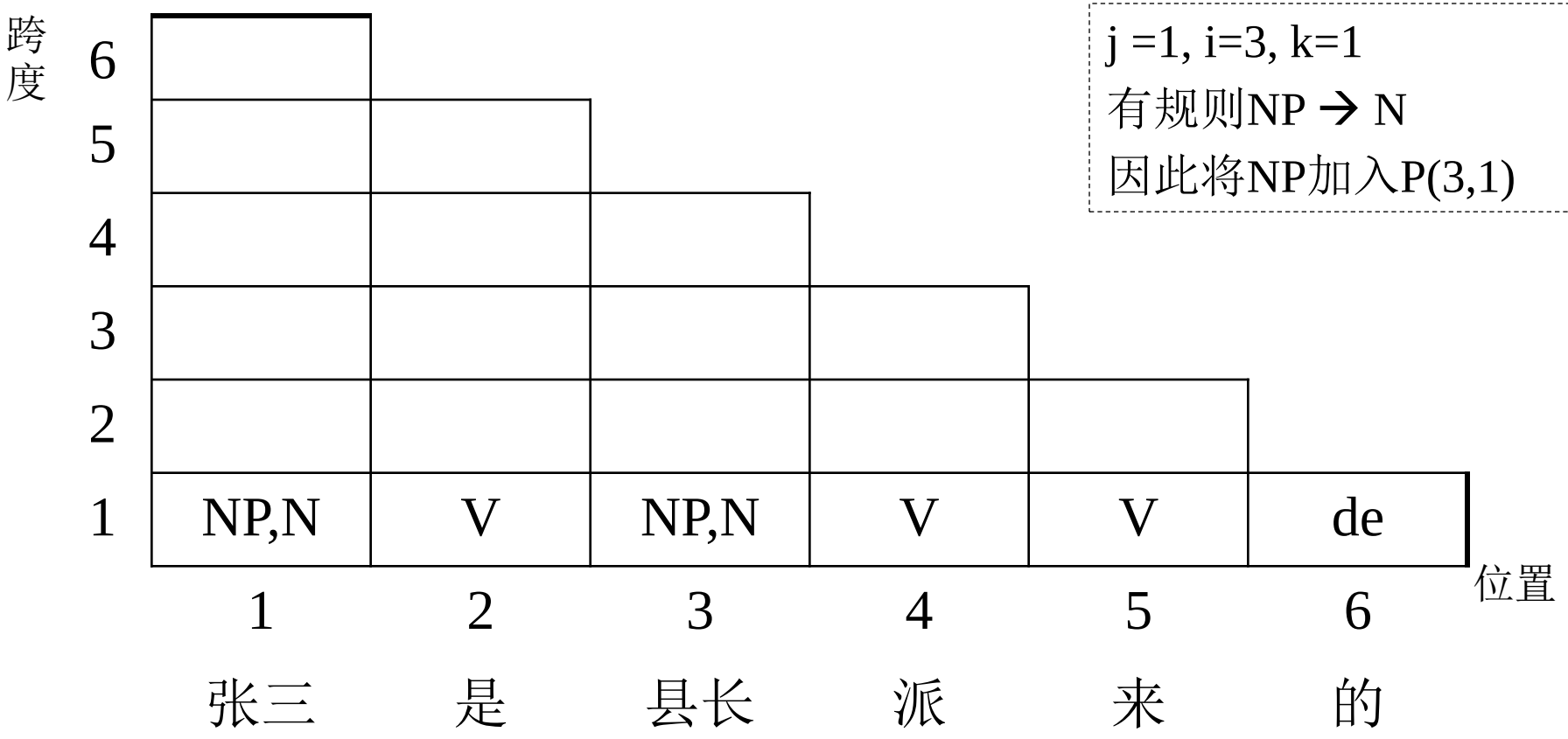
j =1, i=1, k=1
有规则NP → N
因此将NP加入P(1,1)



四 CYK算法示例-填写第1行

(1) S → NP VP
(2) NP → N
(3) NP → CS 的
(4) CS → NP V '
(5) VP → V NP
(6) V ' → V V

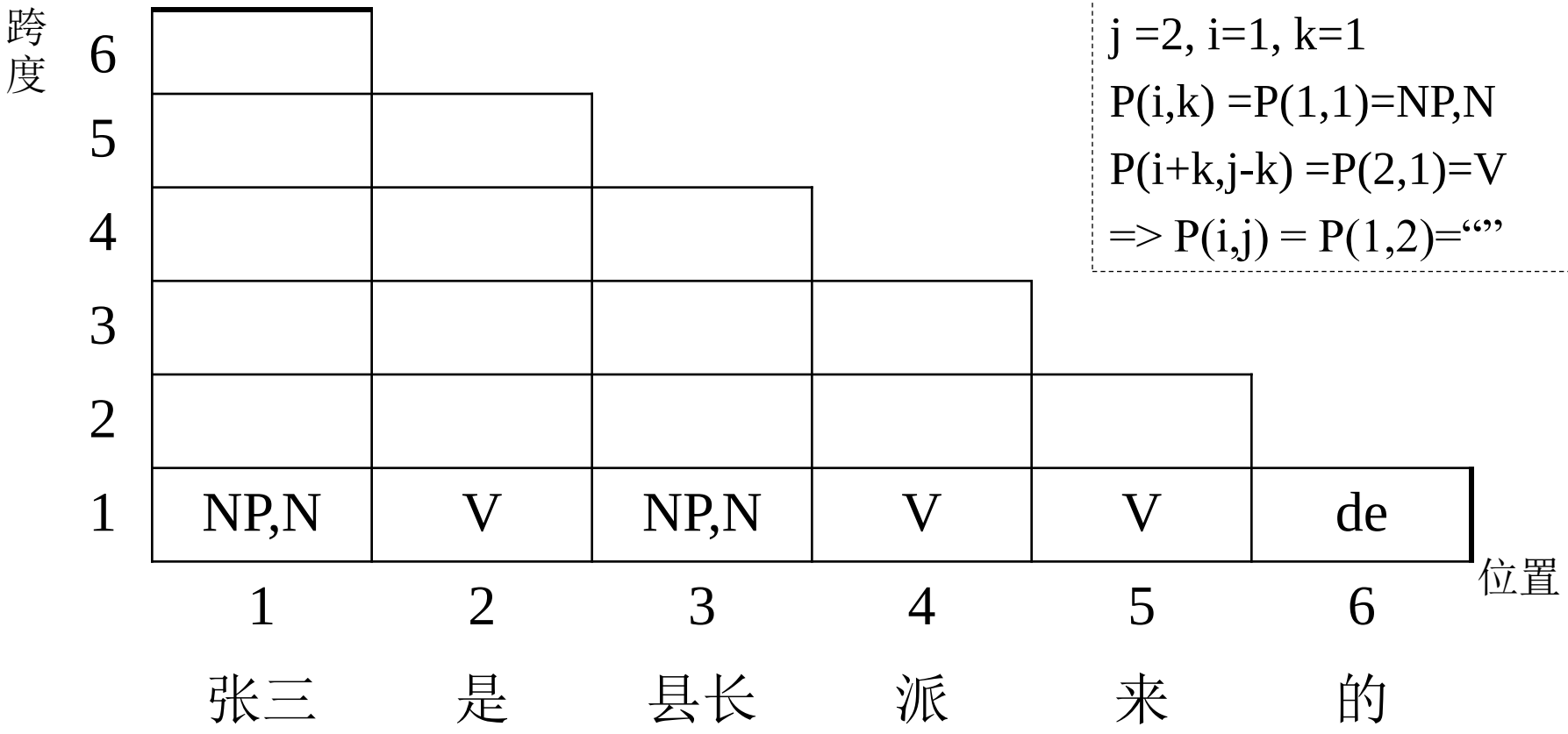
j =1, i=3, k=1
有规则NP → N
因此将NP加入P(3,1)



四 CYK算法示例-填写第2行

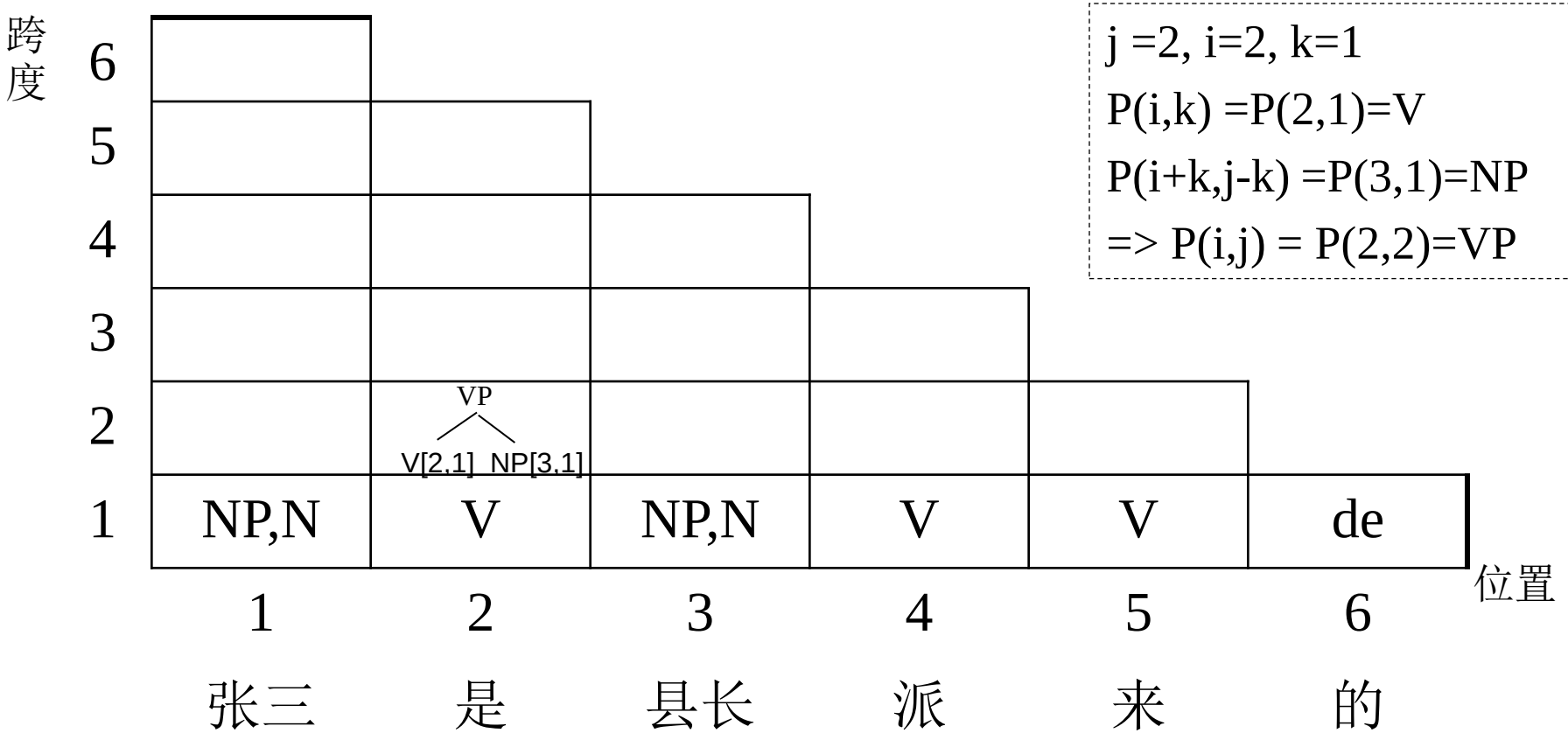
(1) S → NP VP
(2) NP → N
(3) NP → CS 的
(4) CS → NP V '
(5) VP → V NP
(6) V ' → V V

j =2, i=1, k=1
P(i,k) =P(1,1)=NP,N
P(i+k,j-k) =P(2,1)=V
=> P(i,j) = P(1,2)=“”



四 CYK算法示例-填写第2行

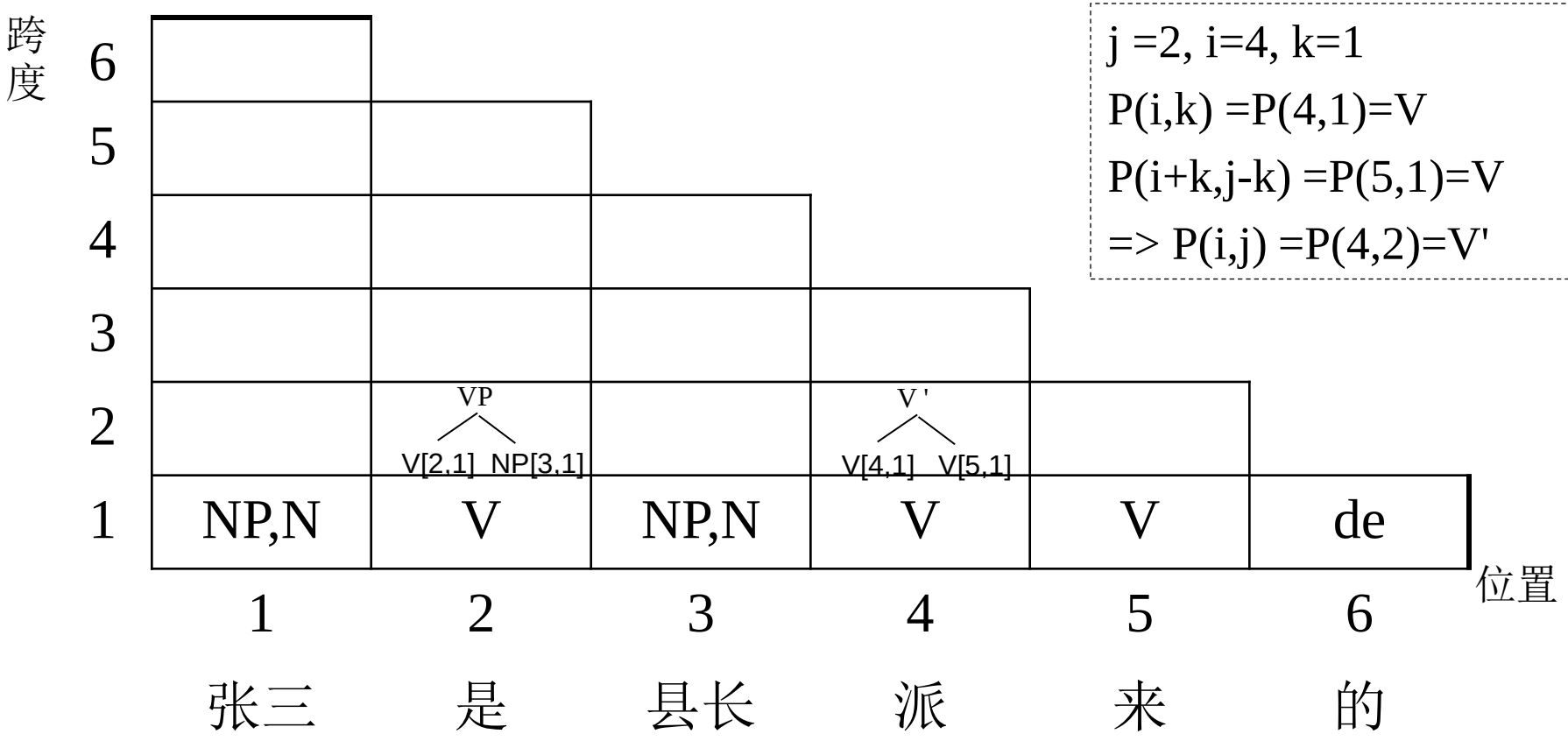
(1) S → NP VP
(2) NP → N
(3) NP → CS 的
(4) CS → NP V '
(5) VP → V NP
(6) V ' → V V



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

四 CYK算法示例-填写第2行

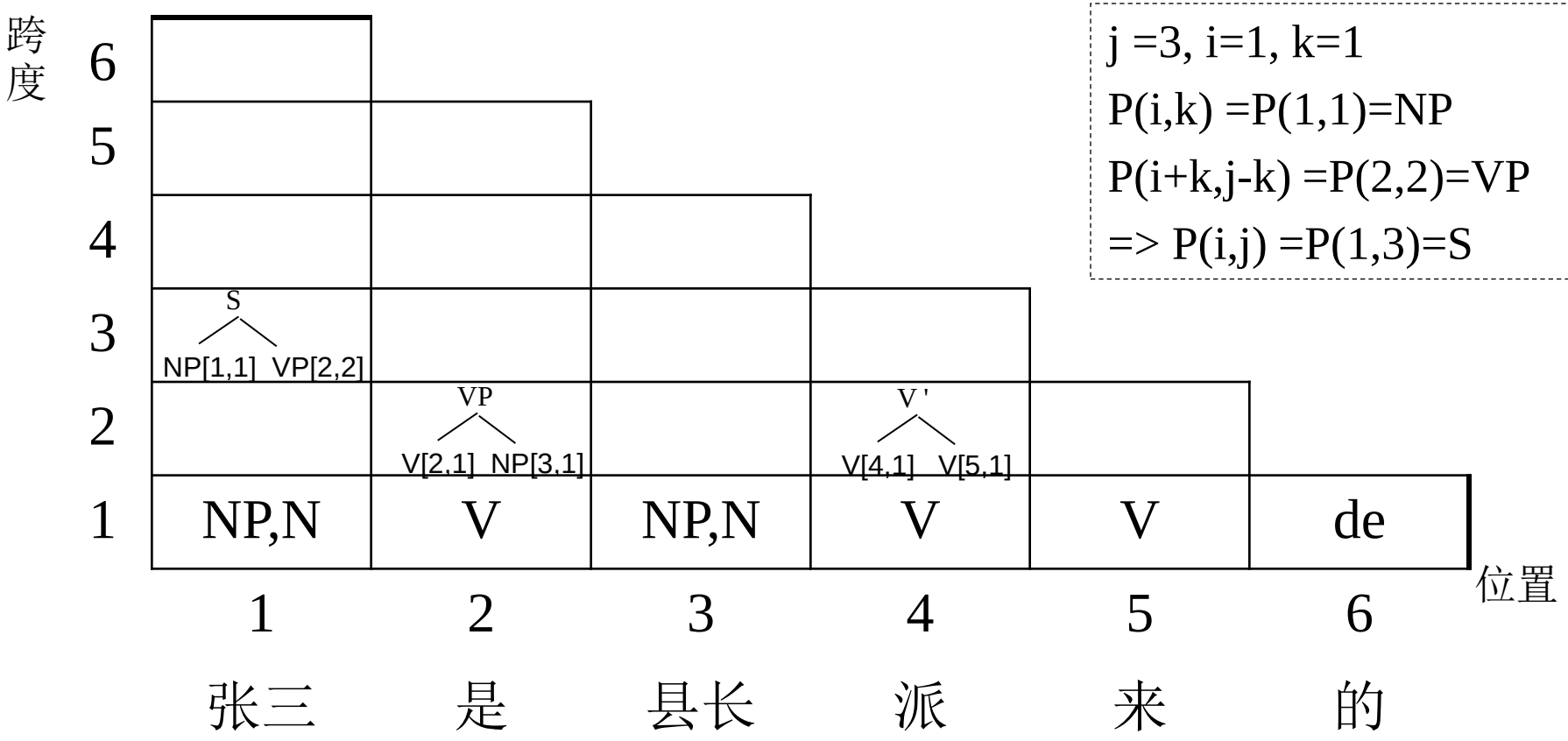
(1) S → NP VP
(2) NP → N
(3) NP → CS 的
(4) CS → NP V '
(5) VP → V NP
(6) V ' → V V



j =2, i=4, k=1
P(i,k) =P(4,1)=V
P(i+k,j-k) =P(5,1)=V
=> P(i,j) =P(4,2)=V'

四 CYK算法示例-填写第3行

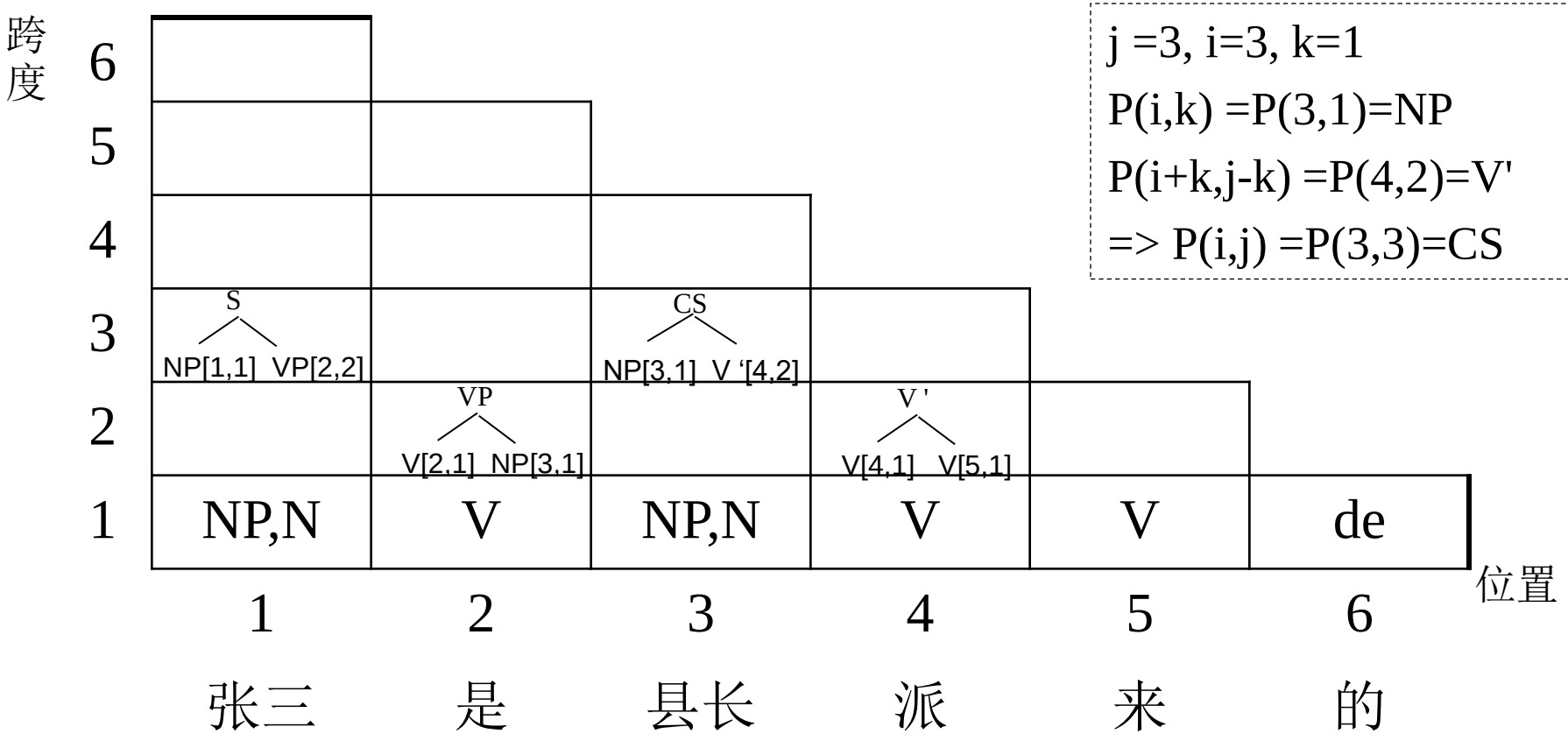
(1) S → NP VP
(2) NP → N
(3) NP → CS 的
(4) CS → NP V '
(5) VP → V NP
(6) V ' → V V



j =3, i=1, k=1
P(i,k) =P(1,1)=NP
P(i+k,j-k) =P(2,2)=VP
=> P(i,j) =P(1,3)=S

四 CYK算法示例-填写第3行

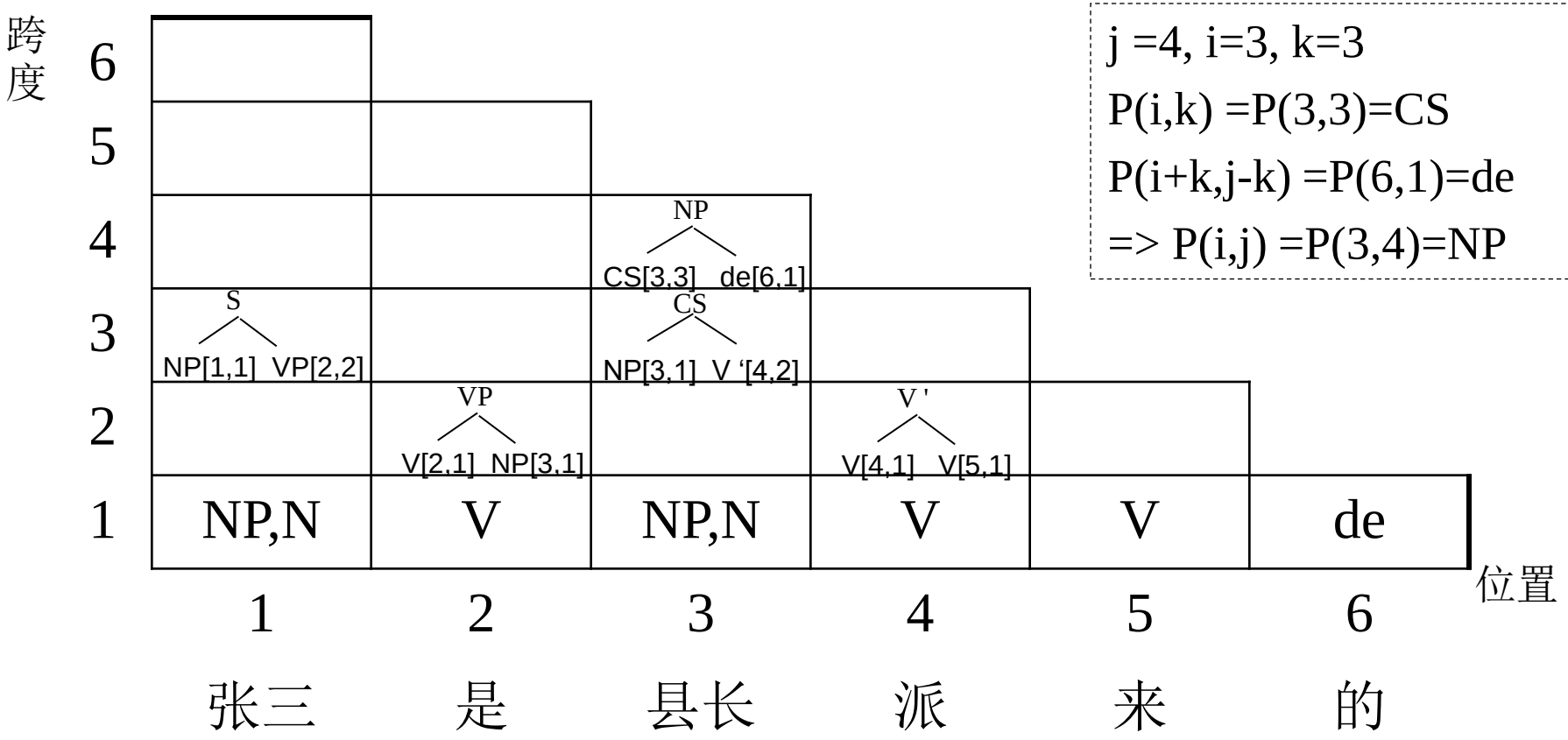
(1) S → NP VP
(2) NP → N
(3) NP → CS 的
(4) CS → NP V '
(5) VP → V NP
(6) V ' → V V



j =3, i=3, k=1
P(i,k) =P(3,1)=NP
P(i+k,j-k) =P(4,2)=V'
=> P(i,j) =P(3,3)=CS

四 CYK算法示例-填写第4行

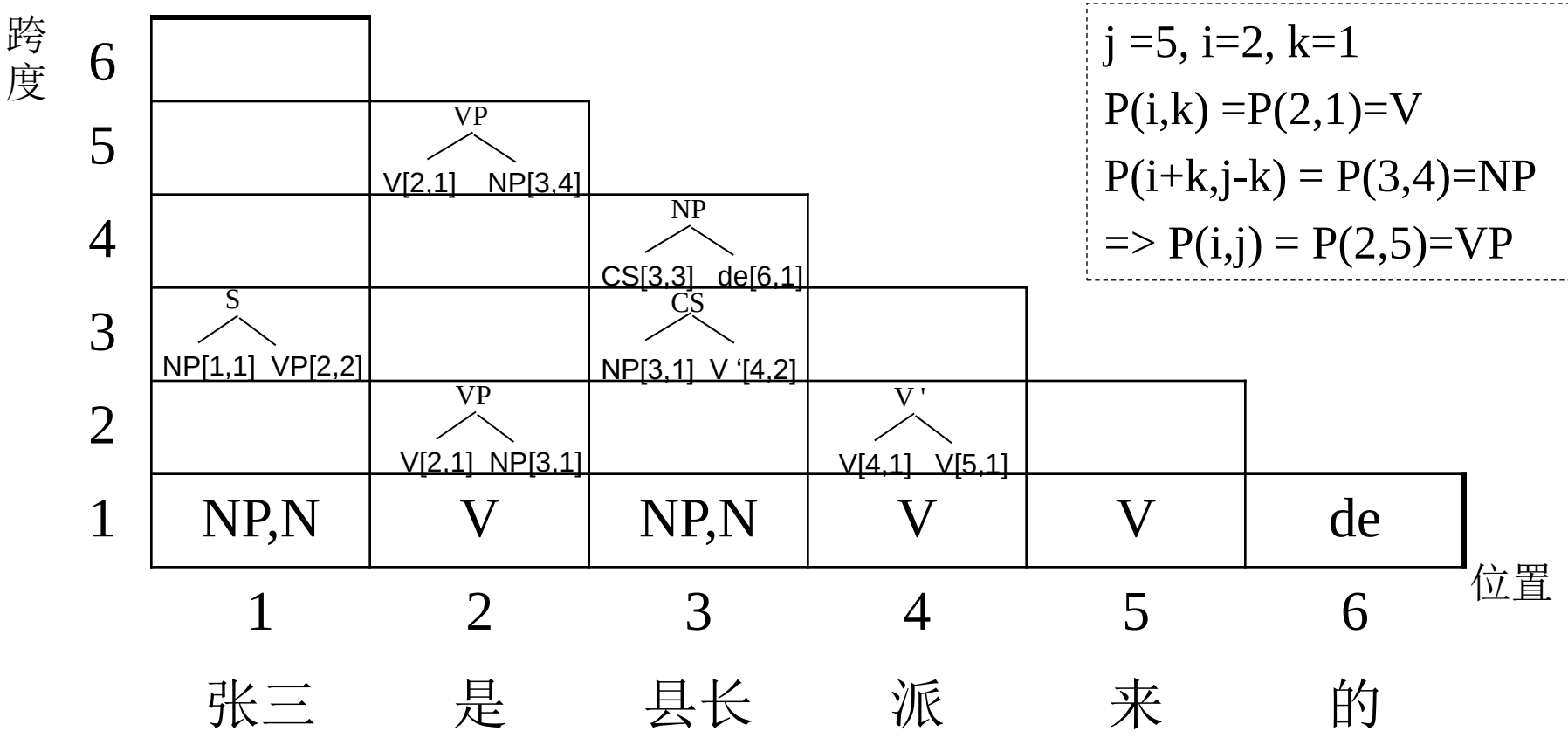
(1) S → NP VP
(2) NP → N
(3) NP → CS 的
(4) CS → NP V '
(5) VP → V NP
(6) V ' → V V



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

四 CYK算法示例-填写第5行

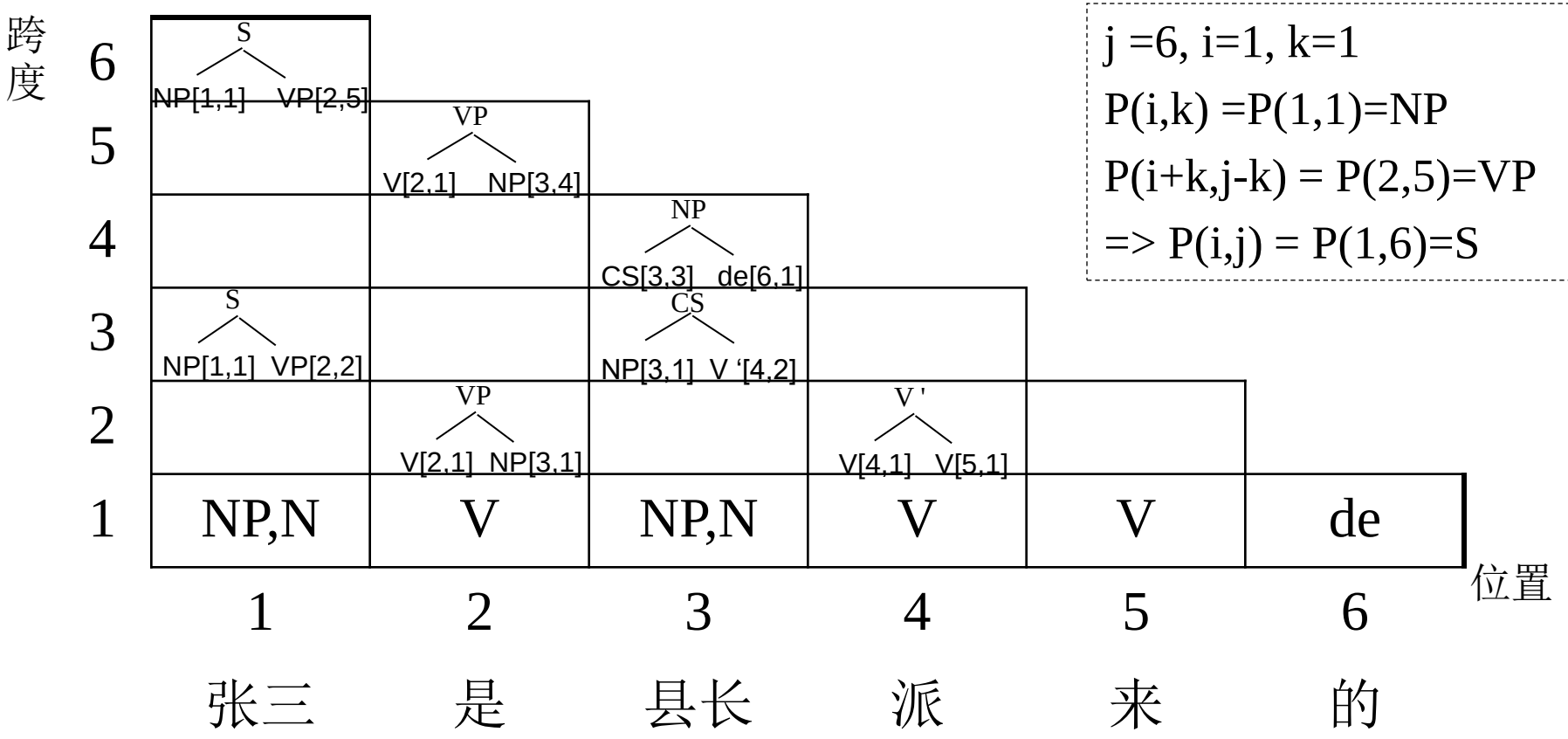
(1) S → NP VP
(2) NP → N
(3) NP → CS 的
(4) CS → NP V '
(5) VP → V NP
(6) V ' → V V



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

四 CYK算法示例-填写第6行

(1) S → NP VP
(2) NP → N
(3) NP → CS 的
(4) CS → NP V '
(5) VP → V NP
(6) V ' → V V



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

四 CYK算法示例-填写第6行

(1) S → NP VP
(2) NP → N
(3) NP → CS 的
(4) CS → NP V '
(5) VP → V NP
(6) V ' → V V

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

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

小结：CYK算法的特点

- 自底向上；
- 按跨度由短到长填表，处理顺序固定；
- 每个成分只算一次，没有重复计算；
- 文法需遵循乔姆斯基范式（CFG → CNF）；
- 时间复杂度$O(n^3)$；