

# CYK分析算法

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

詹卫东

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

# CYK分析算法

- 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:
    - $A \rightarrow BC$
    - $A \rightarrow x$       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

# 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
		张三	是	县长	派	来	的
							位置

# CYK算法—数据结构

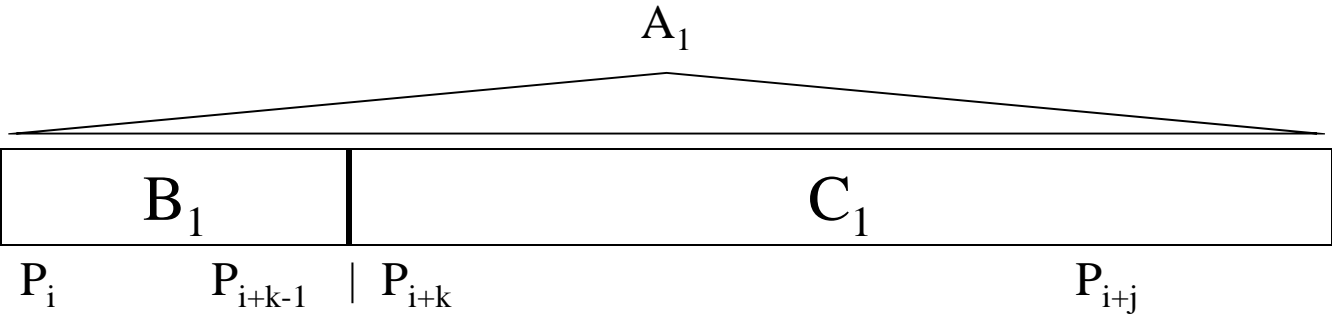
- 一个二维矩阵：{  $P(i, j)$  }
  - 每一个元素 $P(i, j)$ 对应于输入句子中某一个跨度（Span）上所有可能形成的短语的非终结符的集合
  - 横坐标  $i$ ：该跨度左侧第一个词的位置
  - 纵坐标  $j$ ：该跨度包含的词数
- $P(3,1)=\{NP,N\}$ 表示“县长”可以归约成N和NP  
3表示从第3个词位置开始，1表示NP,N覆盖1个词。
- $P(3,3)=\{CS\}$ 表示“县长+派+来”可以规约成CS  
3表示从第3个词位置开始，3表示CS覆盖3个词

# CYK算法：算法描述

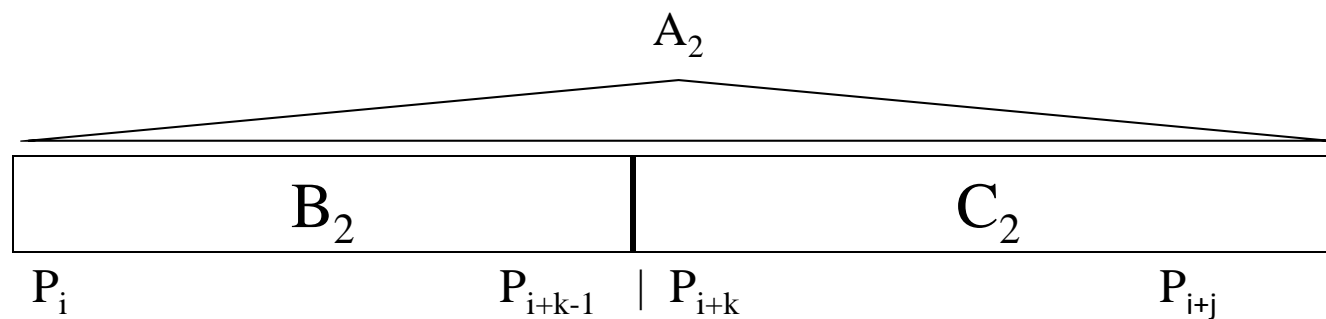
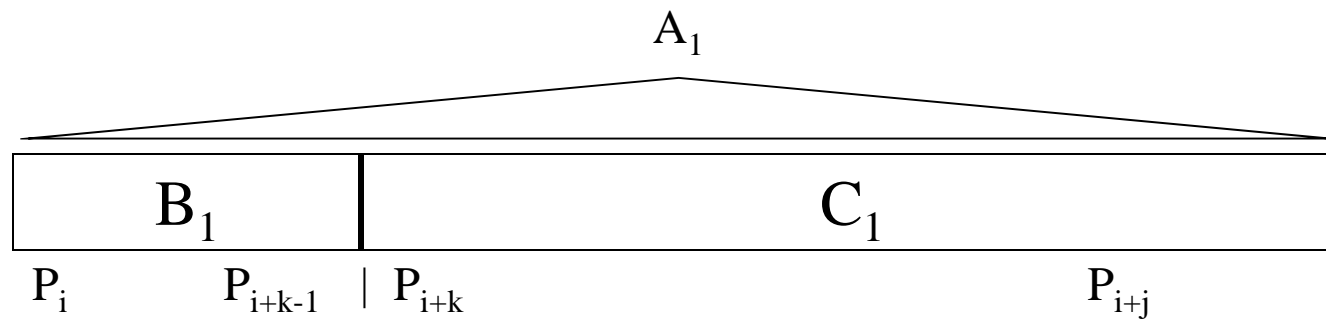
$i$  从1到 $n$ 的循环

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



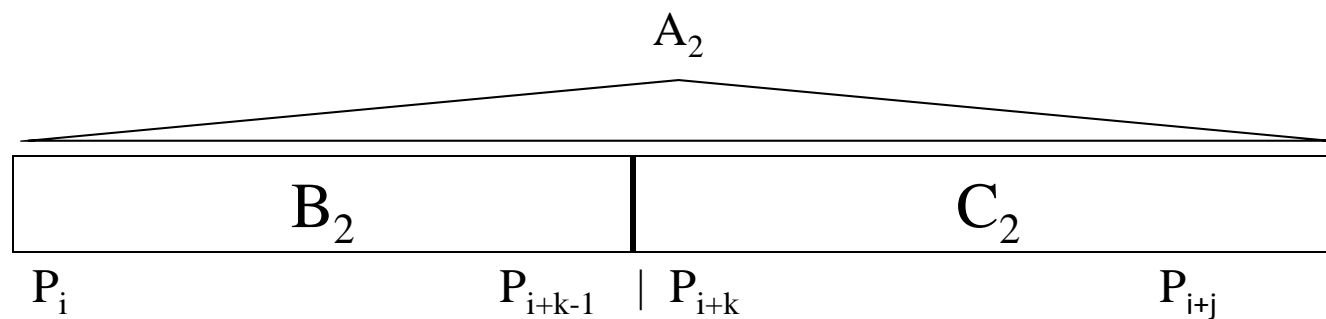
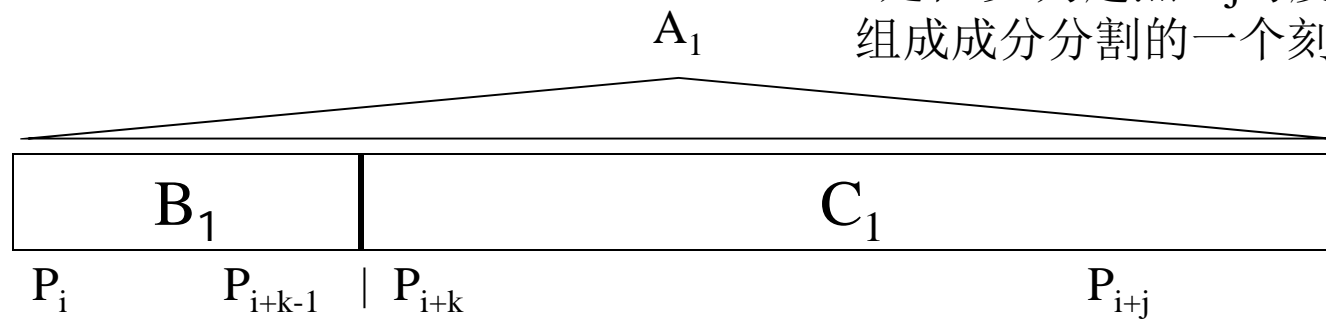
# CYK算法成分组合示意



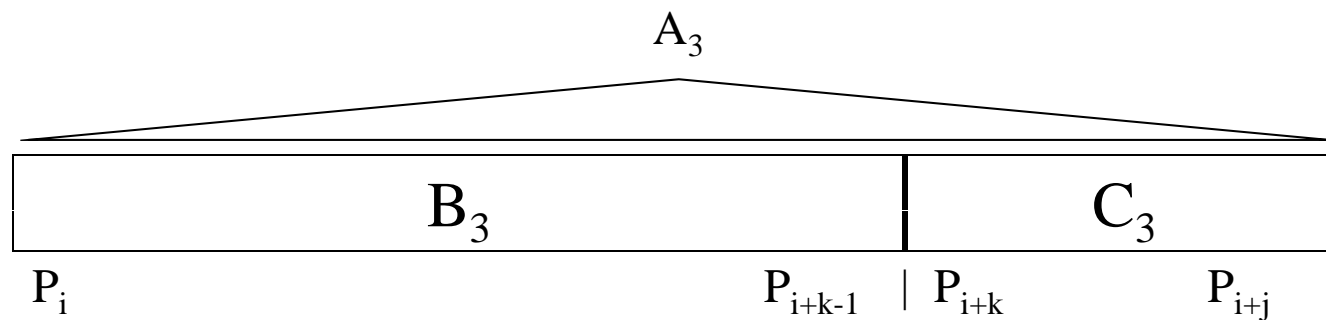
.....

# 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



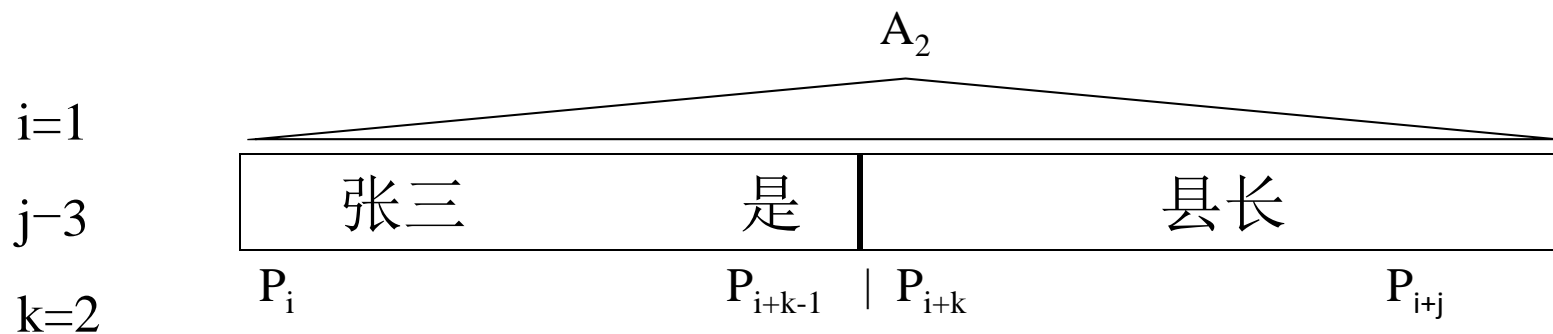
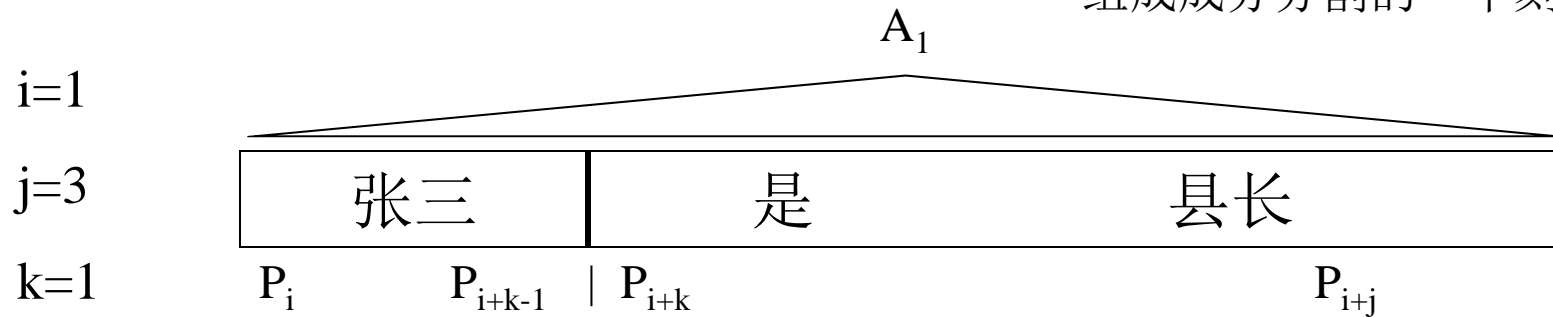
.....





# 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



# CYK算法示例-填写第1行

- (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	N	V	N	V	V	de
	1	2	3	4	5	6
	张三	是	县长	派	来	的

位置

# 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
	1	2	3	4	5	6
	张三	是	县长	派	来	的

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

# 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
	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)

# CYK算法示例-填写第2行

- (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
	张三	是	县长	派	来	的

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

位置

# 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
	1	2	3	4	5	6
	张三	是	县长	派	来	的

位置

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

# CYK算法示例-填写第2行（续）

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

跨度	6					
5						
4						
3						
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

位置

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

# CYK算法示例-填写第3行

- (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					
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

位置

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



# CYK算法示例-填写第3行（续）

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

跨度	6					
5						
4						
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

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

位置

# CYK算法示例-填写第4行

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

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

# CYK算法示例-填写第5行

- (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		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

位置

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

# CYK算法示例-填写第6行

- (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	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
		张三	是	县长	派	来	的

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

位置

# CYK算法：特点

- 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。