CKY (11/10)

CS 585, Fall 2015

Introduction to Natural Language Processing

http://people.cs.umass.edu/~brenocon/inlp2015/

Brendan O'Connor

College of Information and Computer Sciences

University of Massachusetts Amherst

CFGs

- A context-free grammar has
  - a start symbol
  - production rules: A -> B C D ...
    - One symbol on left
    - One or more symbols on right
  - Non-terminals vs Terminal symbols
    - non-terminals: S, NP, VP
    - terminals: e.g. words (leaves of tree)
- CFG can "generate" a set of strings (often infinite) via a rewrite process.
  - Multiple rules starting with e.g. A:
    multiple possible rewrites of A
- More general goal: *generative model of language*
  - other examples?

Ambiguity

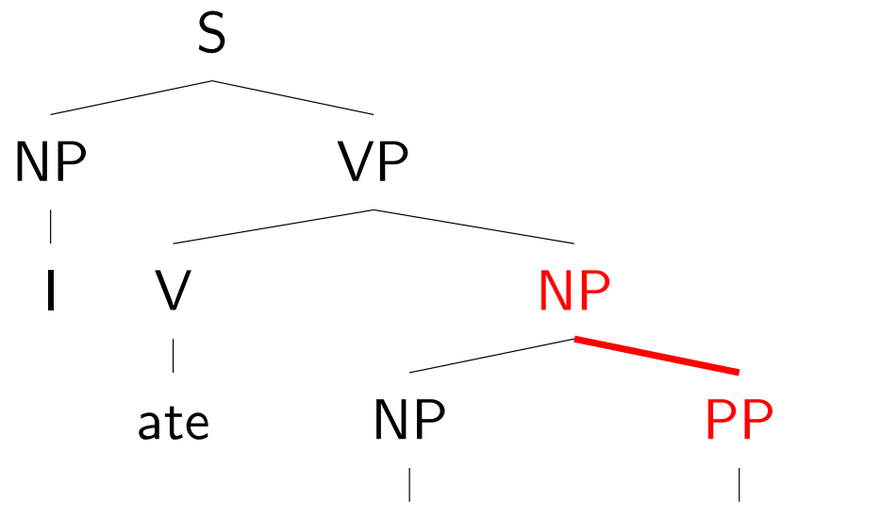
- There may be many parses for a single sentence

Prepositional attachment ambiguity

I ate some dessert with a fork.

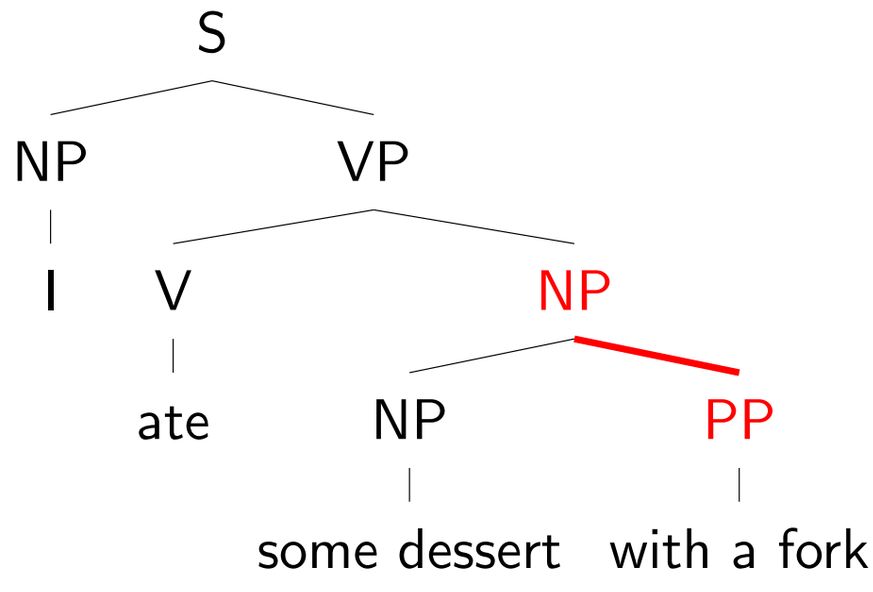
Prepositional attachment ambiguity

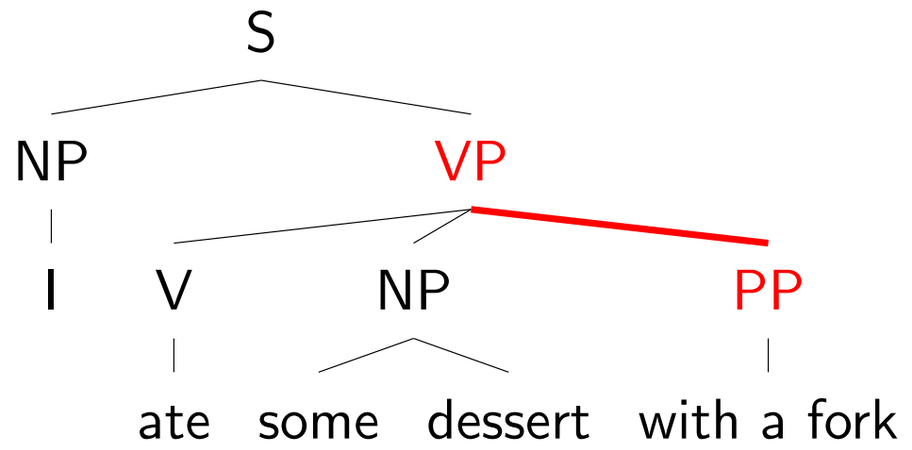
I ate some dessert with a fork.



Prepositional attachment ambiguity

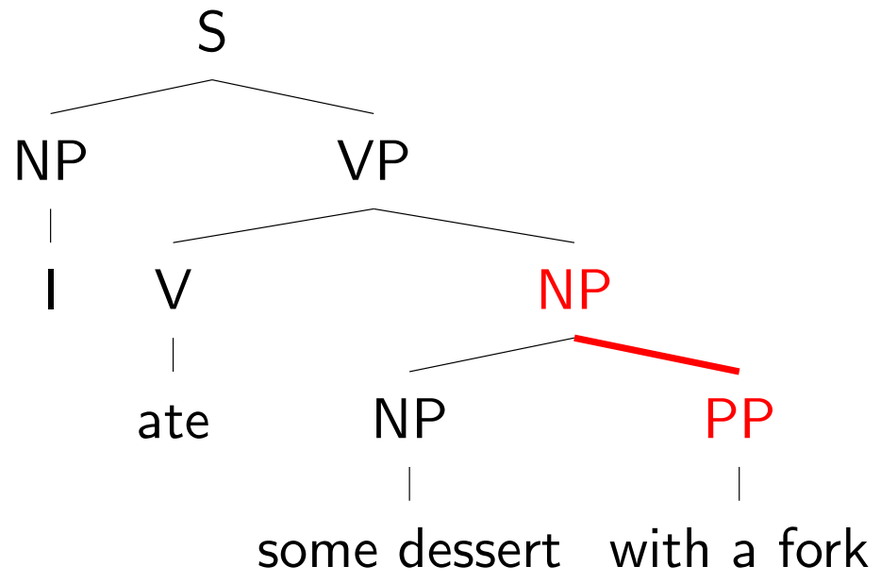
I ate some dessert with a fork.

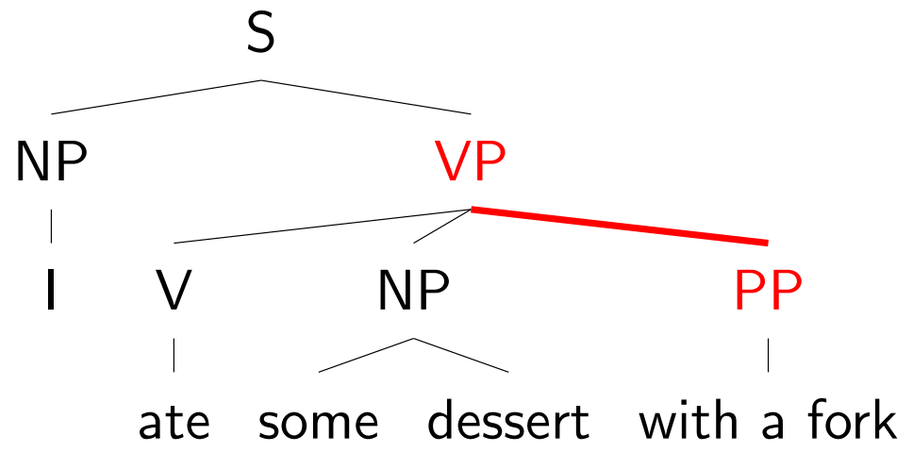




Prepositional attachment ambiguity

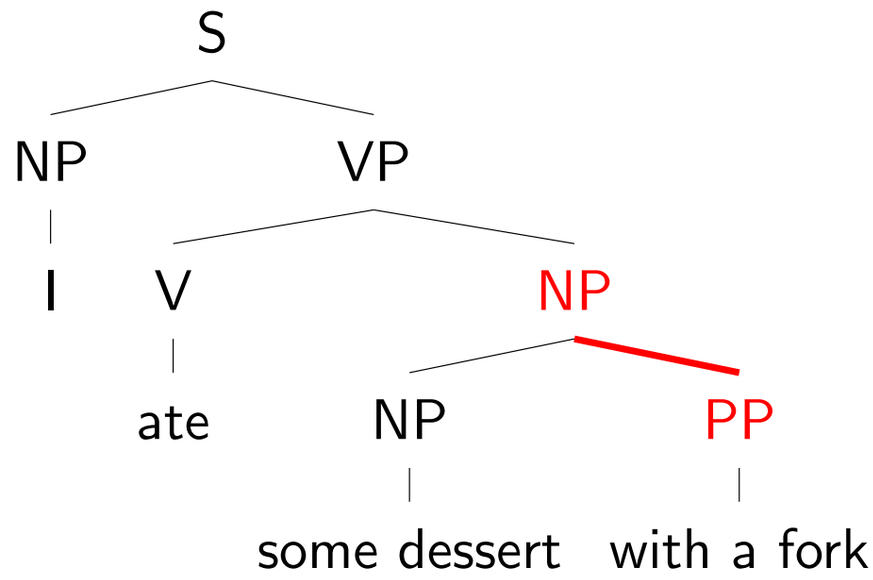
I ate some dessert with a fork.

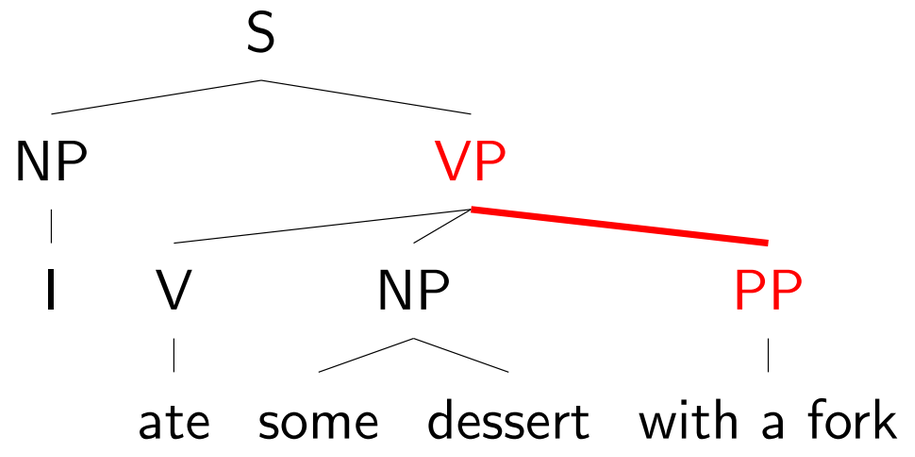




Prepositional attachment ambiguity

I ate some dessert with a fork.





Both are grammatical; is syntax enough to disambiguate?

Questions

- What is a good representation?
  - CFG
- What is an efficient algorithm to *parse*?
  - CKY

CKY Algorithm

Cocke–Younger–Kasami

- Given a CFG and a sentence, efficiently answer:
  - *recognizer:* Does a parse exist for it?
  - *parser:* Enumerate parses
- *Weighted CKY* with a weighted/prob CFG:
  - Find the most probable parse

- Dynamic programming!
  - We can construct possible *local subtrees*
  - Maintain these partial hypotheses, *bottom-up*
  - Infer the "parse forest" of *all* possible trees

Binarized rules

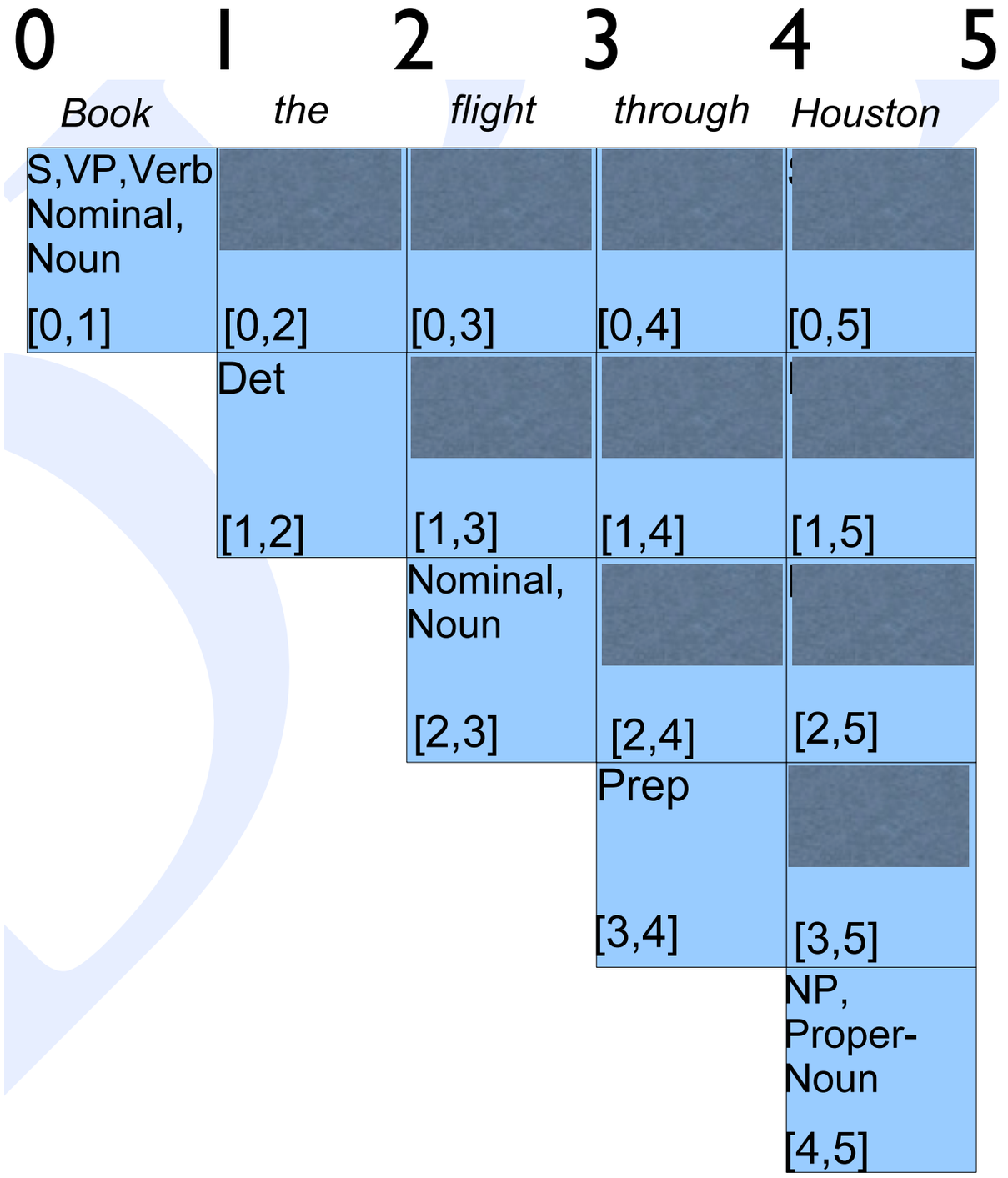
Necessary for CKY algorithm
Can convert to equivalent binarized grammar

$$S \rightarrow NP\ VP$$
$$S \rightarrow Aux\ NP\ VP$$

$$\left\| \begin{array}{l} S \rightarrow NP\ VP \\ S \rightarrow X1\ VP \\ X1 \rightarrow Aux\ NP \end{array} \right.$$

CKY recognizer

- Fill in all length-1 spans with possible nonterminals.
- Go bottom-up: progressively fill each cell with possible states, based on possible combinations below.
- If the top cell [0,5] can expand from ROOT, then accept!
- To get one of possible parses: trace backpointers
- Dynamic programming: what's below the cell does not matter



(J&M has a slightly different cell ordering. Both OK.)

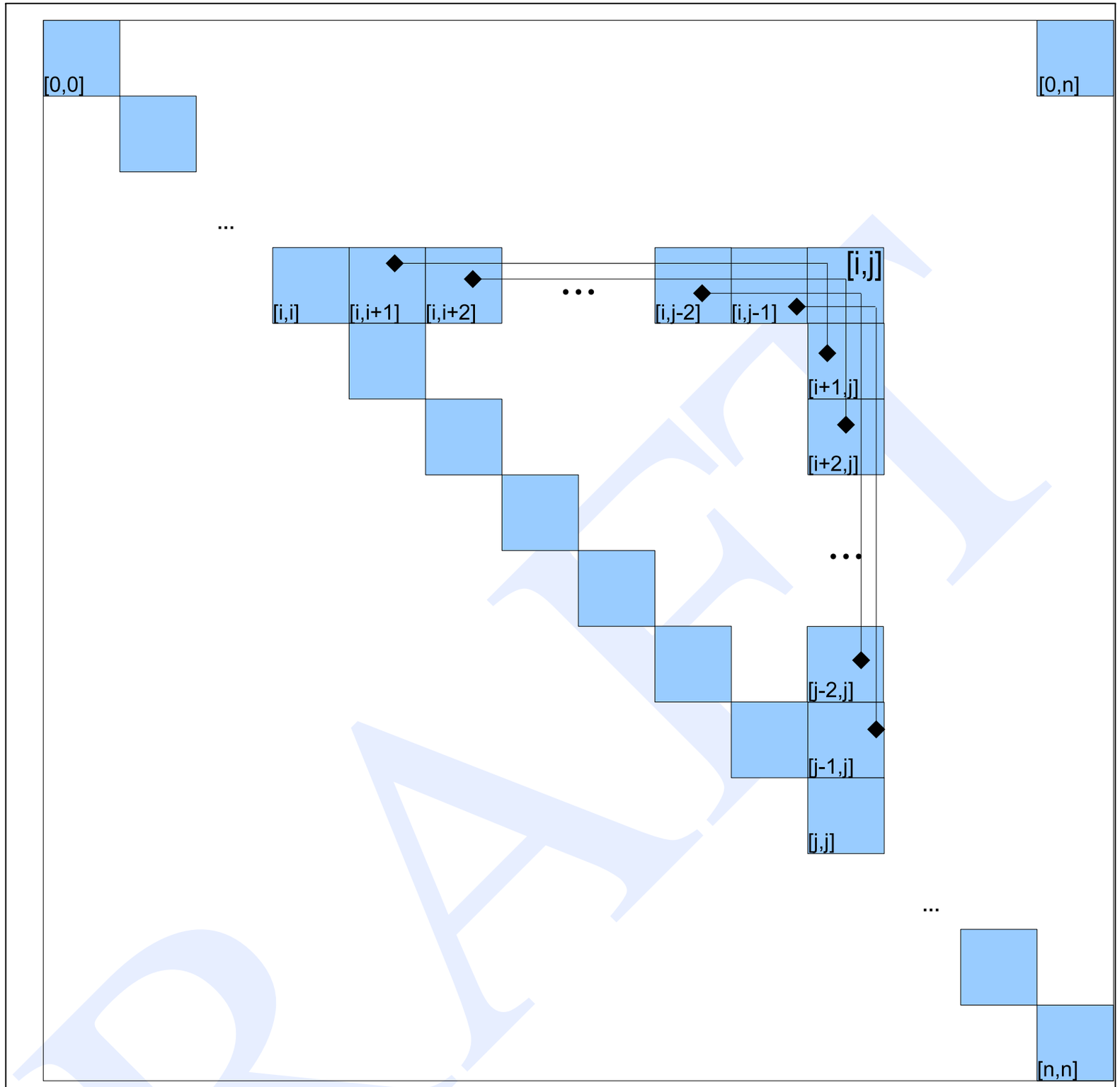
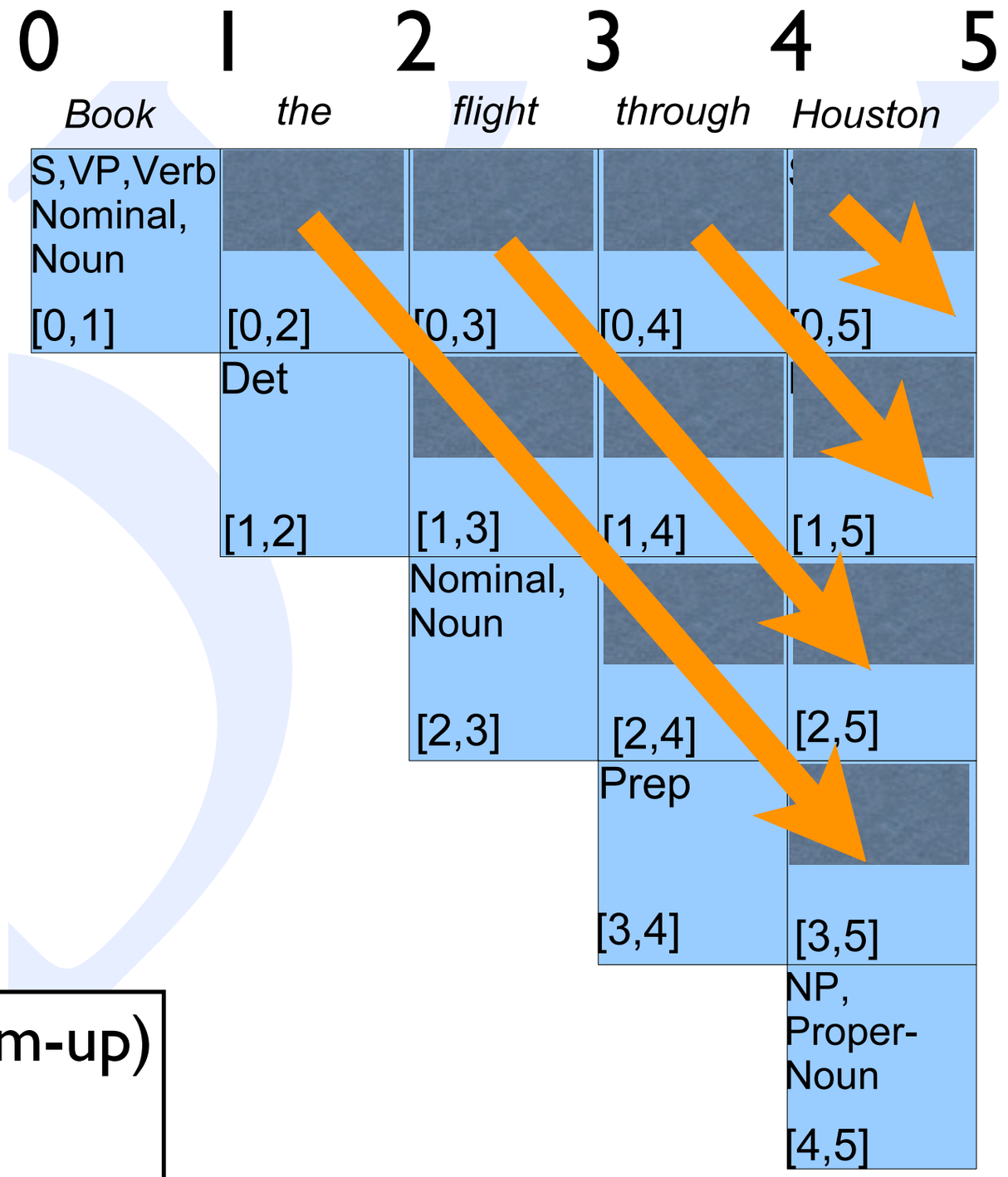


Figure 13.11 All the ways to fill the [i,j]th cell in the CKY table.

CKY recognizer

$S \rightarrow NP\ VP$
$S \rightarrow Aux\ NP\ VP$
$S \rightarrow VP$
$NP \rightarrow Pronoun$
$NP \rightarrow Proper\text{-}Noun$
$NP \rightarrow Det\ Nominal$
$Nominal \rightarrow Noun$
$Nominal \rightarrow Nominal\ Noun$
$Nominal \rightarrow Nominal\ PP$
$VP \rightarrow Verb$
$VP \rightarrow Verb\ NP$
$VP \rightarrow Verb\ NP\ PP$
$VP \rightarrow Verb\ PP$
$VP \rightarrow VP\ PP$
$PP \rightarrow Preposition\ NP$

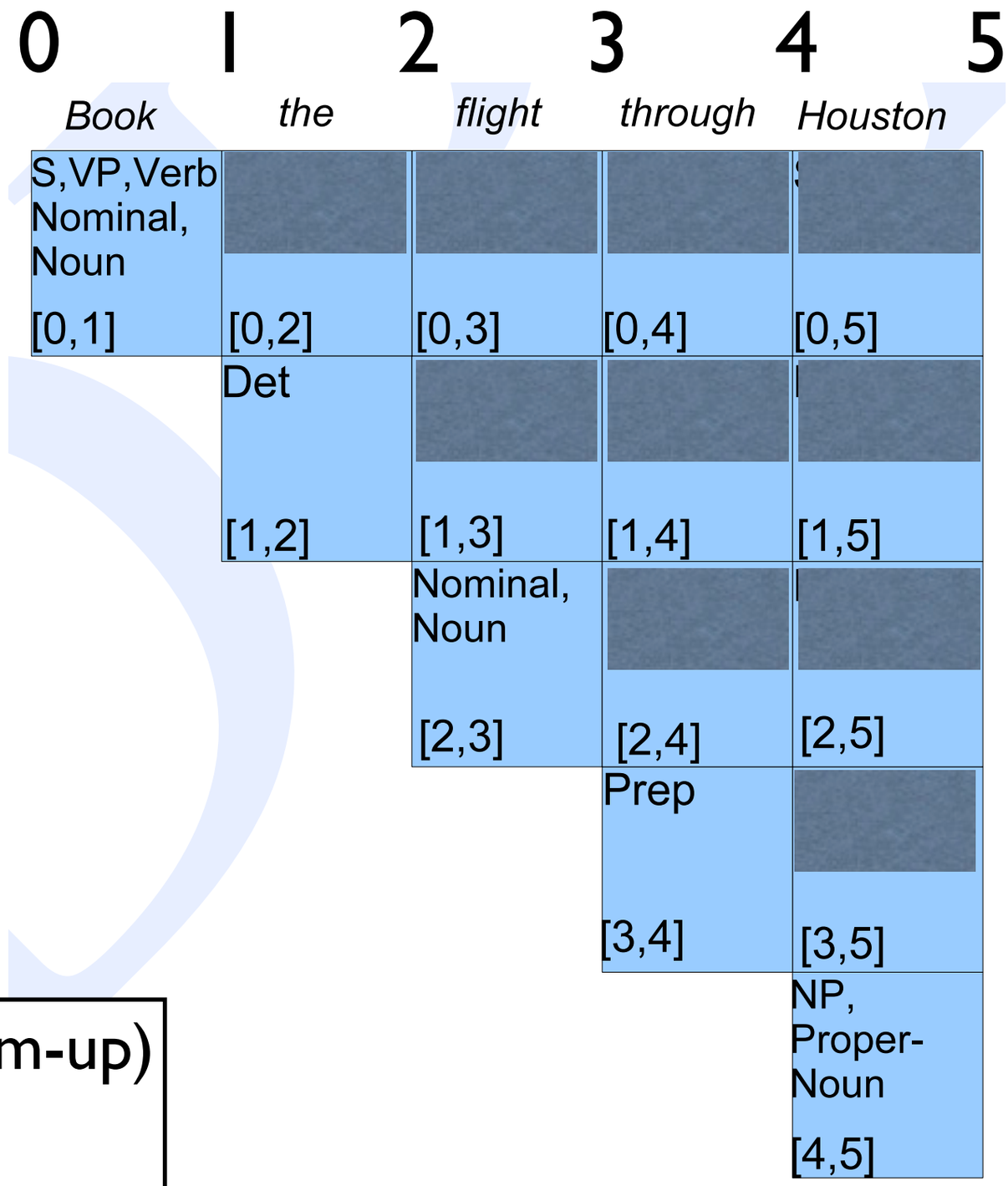


For cell [i,j] (loop through them bottom-up)
  For possible splitpoint k=(i+1)..(j-1):
    For every B in [i,k] and C in [k,j],
      If exists rule A -> B C,
        add A to cell [i,j]

(J&M has a slightly different cell ordering. Both OK.)

CKY recognizer

$S \rightarrow NP\ VP$
$S \rightarrow Aux\ NP\ VP$
$S \rightarrow VP$
$NP \rightarrow Pronoun$
$NP \rightarrow Proper\text{-}Noun$
$NP \rightarrow Det\ Nominal$
$Nominal \rightarrow Noun$
$Nominal \rightarrow Nominal\ Noun$
$Nominal \rightarrow Nominal\ PP$
$VP \rightarrow Verb$
$VP \rightarrow Verb\ NP$
$VP \rightarrow Verb\ NP\ PP$
$VP \rightarrow Verb\ PP$
$VP \rightarrow VP\ PP$
$PP \rightarrow Preposition\ NP$

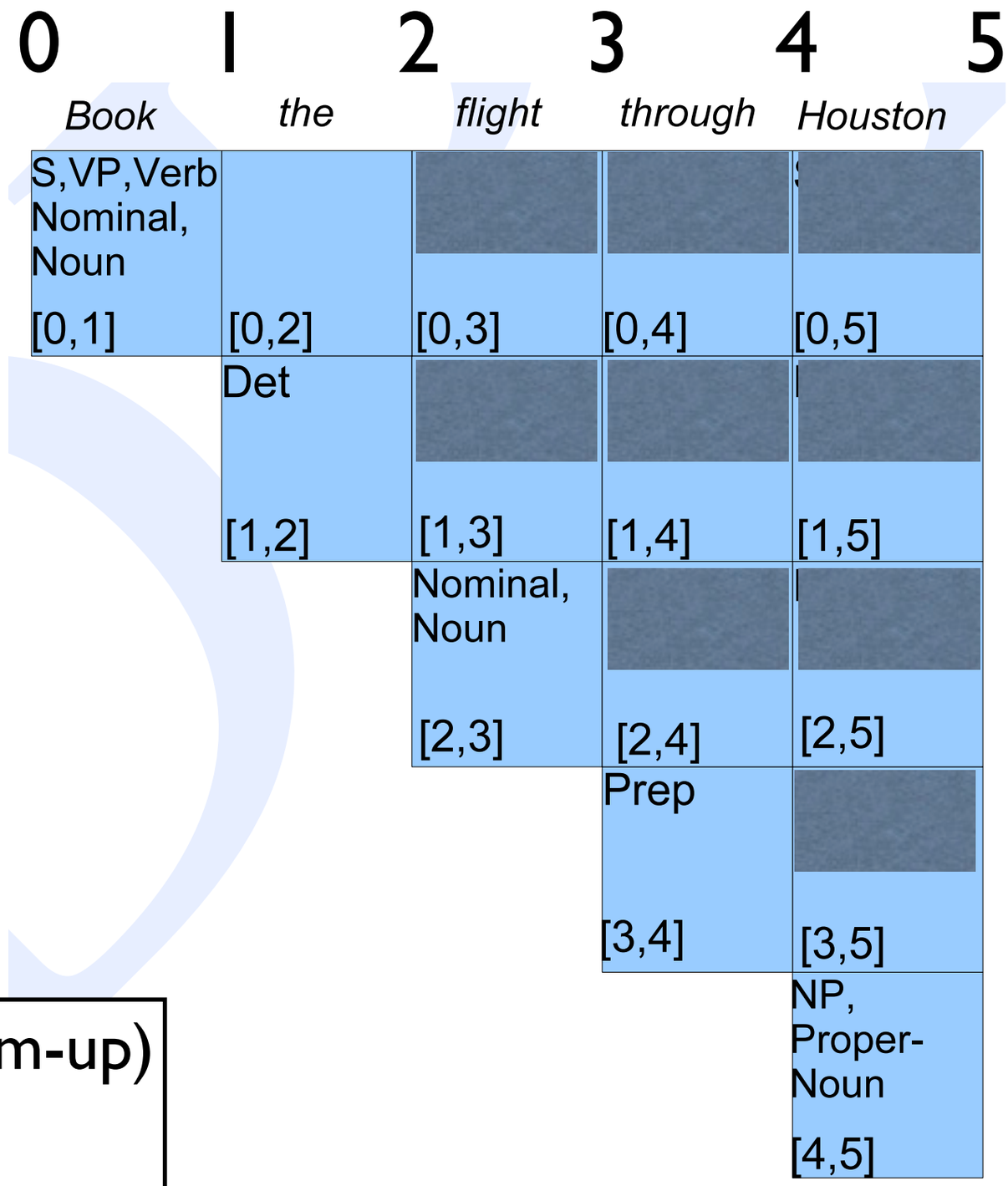


For cell [i,j] (loop through them bottom-up)
  For possible splitpoint k=(i+1)..(j-1):
    For every B in [i,k] and C in [k,j],
      If exists rule A -> B C,
        add A to cell [i,j]

(J&M has a slightly different cell ordering. Both OK.)

CKY recognizer

$S \rightarrow NP\ VP$
$S \rightarrow Aux\ NP\ VP$
$S \rightarrow VP$
$NP \rightarrow Pronoun$
$NP \rightarrow Proper\text{-}Noun$
$NP \rightarrow Det\ Nominal$
$Nominal \rightarrow Noun$
$Nominal \rightarrow Nominal\ Noun$
$Nominal \rightarrow Nominal\ PP$
$VP \rightarrow Verb$
$VP \rightarrow Verb\ NP$
$VP \rightarrow Verb\ NP\ PP$
$VP \rightarrow Verb\ PP$
$VP \rightarrow VP\ PP$
$PP \rightarrow Preposition\ NP$

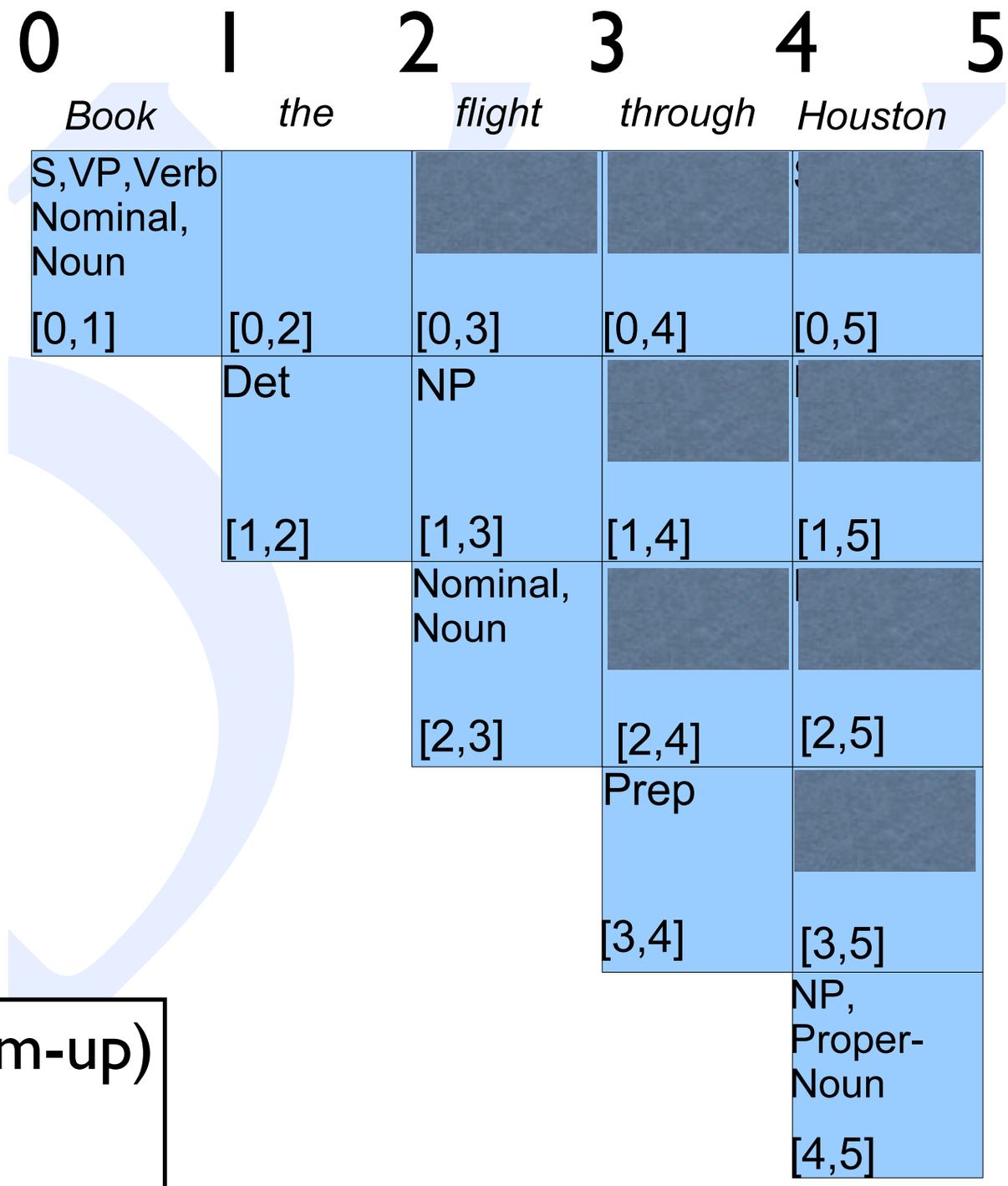


For cell [i,j] (loop through them bottom-up)
  For possible splitpoint k=(i+1)..(j-1):
    For every B in [i,k] and C in [k,j],
      If exists rule A -> B C,
        add A to cell [i,j]

(J&M has a slightly different cell ordering. Both OK.)

CKY recognizer

$S \rightarrow NP\ VP$
$S \rightarrow Aux\ NP\ VP$
$S \rightarrow VP$
$NP \rightarrow Pronoun$
$NP \rightarrow Proper\text{-}Noun$
$NP \rightarrow Det\ Nominal$
$Nominal \rightarrow Noun$
$Nominal \rightarrow Nominal\ Noun$
$Nominal \rightarrow Nominal\ PP$
$VP \rightarrow Verb$
$VP \rightarrow Verb\ NP$
$VP \rightarrow Verb\ NP\ PP$
$VP \rightarrow Verb\ PP$
$VP \rightarrow VP\ PP$
$PP \rightarrow Preposition\ NP$

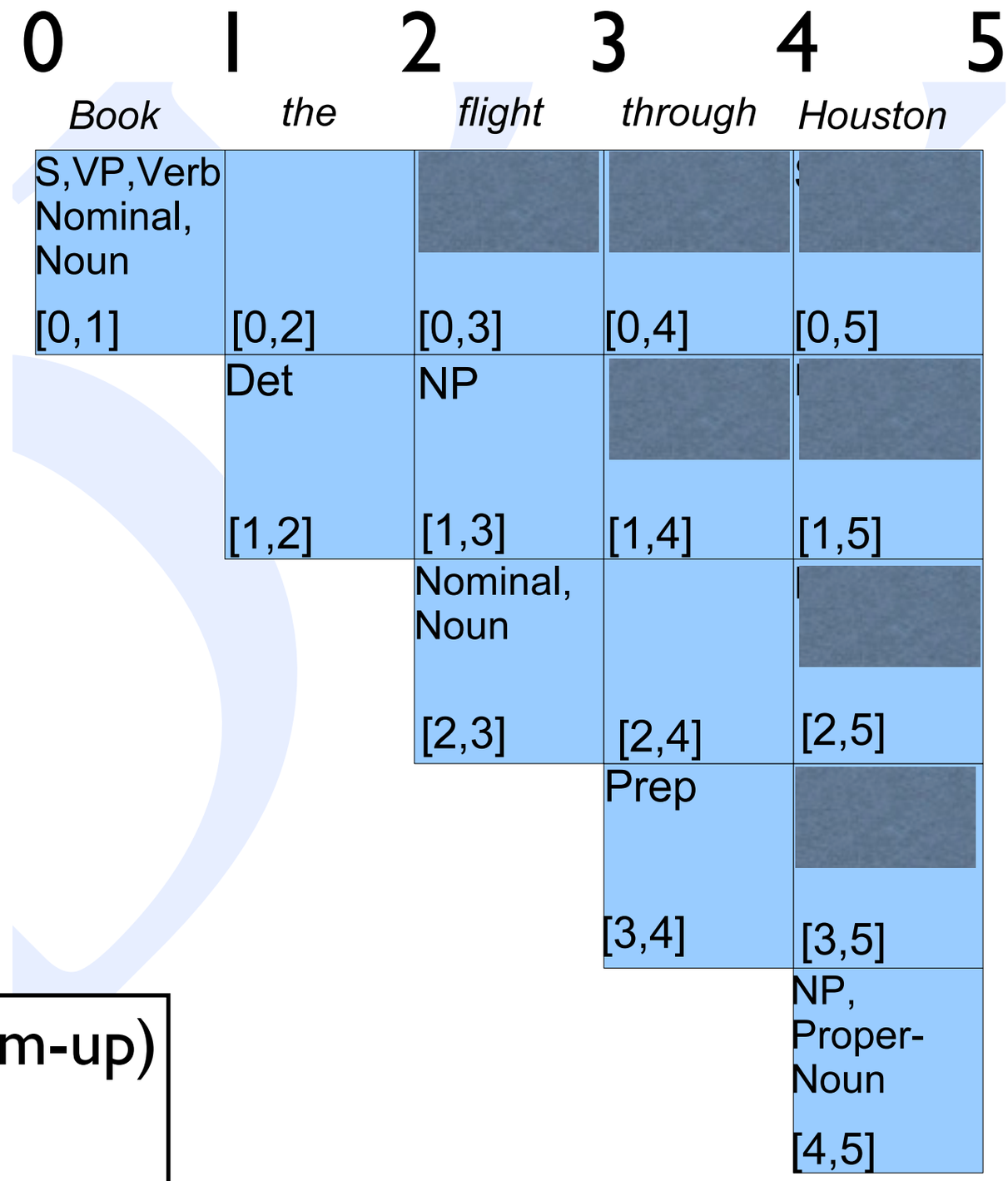


For cell [i,j] (loop through them bottom-up)
  For possible splitpoint k=(i+1)..(j-1):
    For every B in [i,k] and C in [k,j],
      If exists rule A -> B C,
        add A to cell [i,j]

(J&M has a slightly different cell ordering. Both OK.)

CKY recognizer

$S \rightarrow NP\ VP$
$S \rightarrow Aux\ NP\ VP$
$S \rightarrow VP$
$NP \rightarrow Pronoun$
$NP \rightarrow Proper\text{-}Noun$
$NP \rightarrow Det\ Nominal$
$Nominal \rightarrow Noun$
$Nominal \rightarrow Nominal\ Noun$
$Nominal \rightarrow Nominal\ PP$
$VP \rightarrow Verb$
$VP \rightarrow Verb\ NP$
$VP \rightarrow Verb\ NP\ PP$
$VP \rightarrow Verb\ PP$
$VP \rightarrow VP\ PP$
$PP \rightarrow Preposition\ NP$

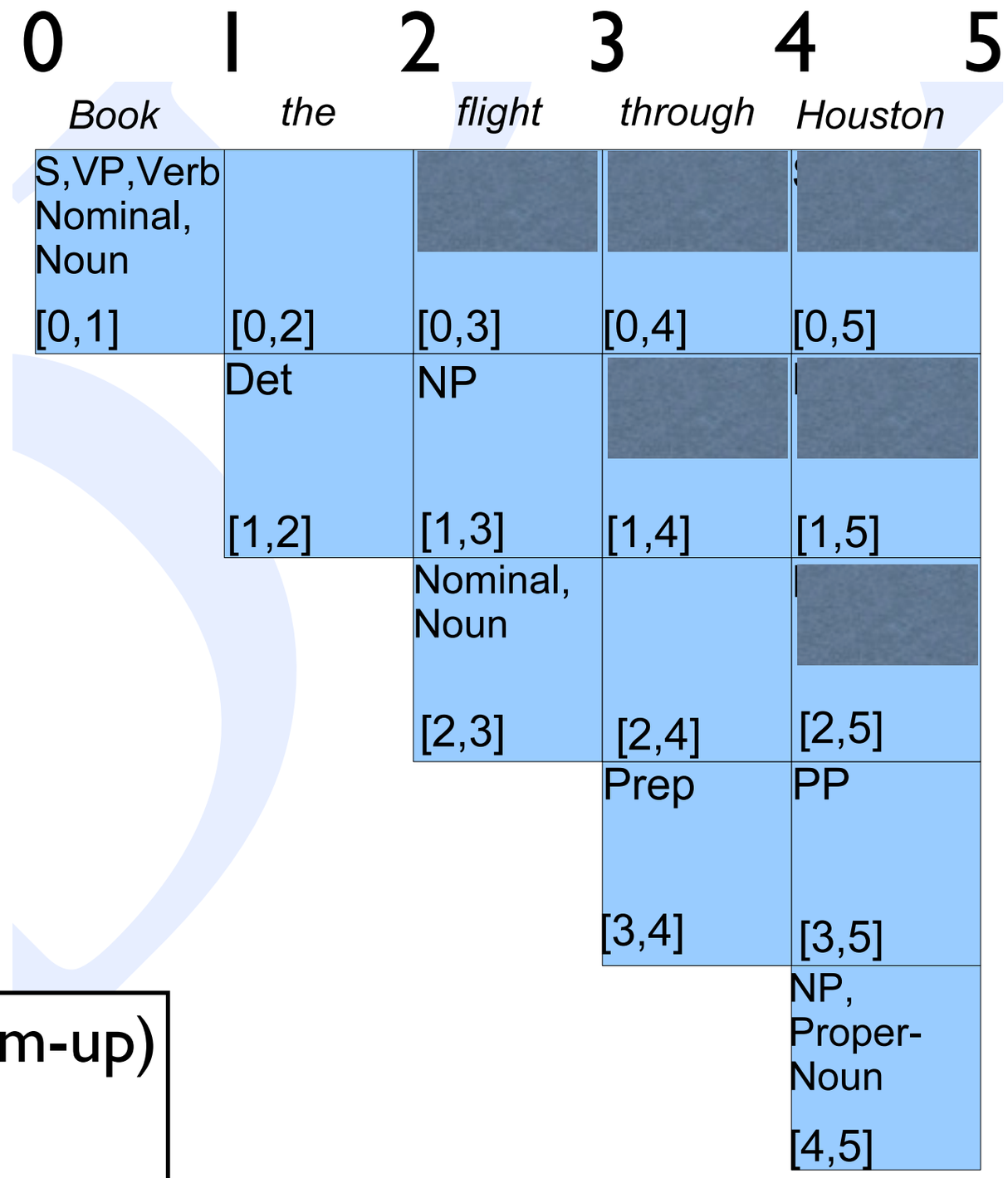


For cell [i,j] (loop through them bottom-up)
  For possible splitpoint k=(i+1)..(j-1):
    For every B in [i,k] and C in [k,j],
      If exists rule A -> B C,
        add A to cell [i,j]

(J&M has a slightly different cell ordering. Both OK.)

CKY recognizer

$S \rightarrow NP\ VP$
$S \rightarrow Aux\ NP\ VP$
$S \rightarrow VP$
$NP \rightarrow Pronoun$
$NP \rightarrow Proper\text{-}Noun$
$NP \rightarrow Det\ Nominal$
$Nominal \rightarrow Noun$
$Nominal \rightarrow Nominal\ Noun$
$Nominal \rightarrow Nominal\ PP$
$VP \rightarrow Verb$
$VP \rightarrow Verb\ NP$
$VP \rightarrow Verb\ NP\ PP$
$VP \rightarrow Verb\ PP$
$VP \rightarrow VP\ PP$
$PP \rightarrow Preposition\ NP$

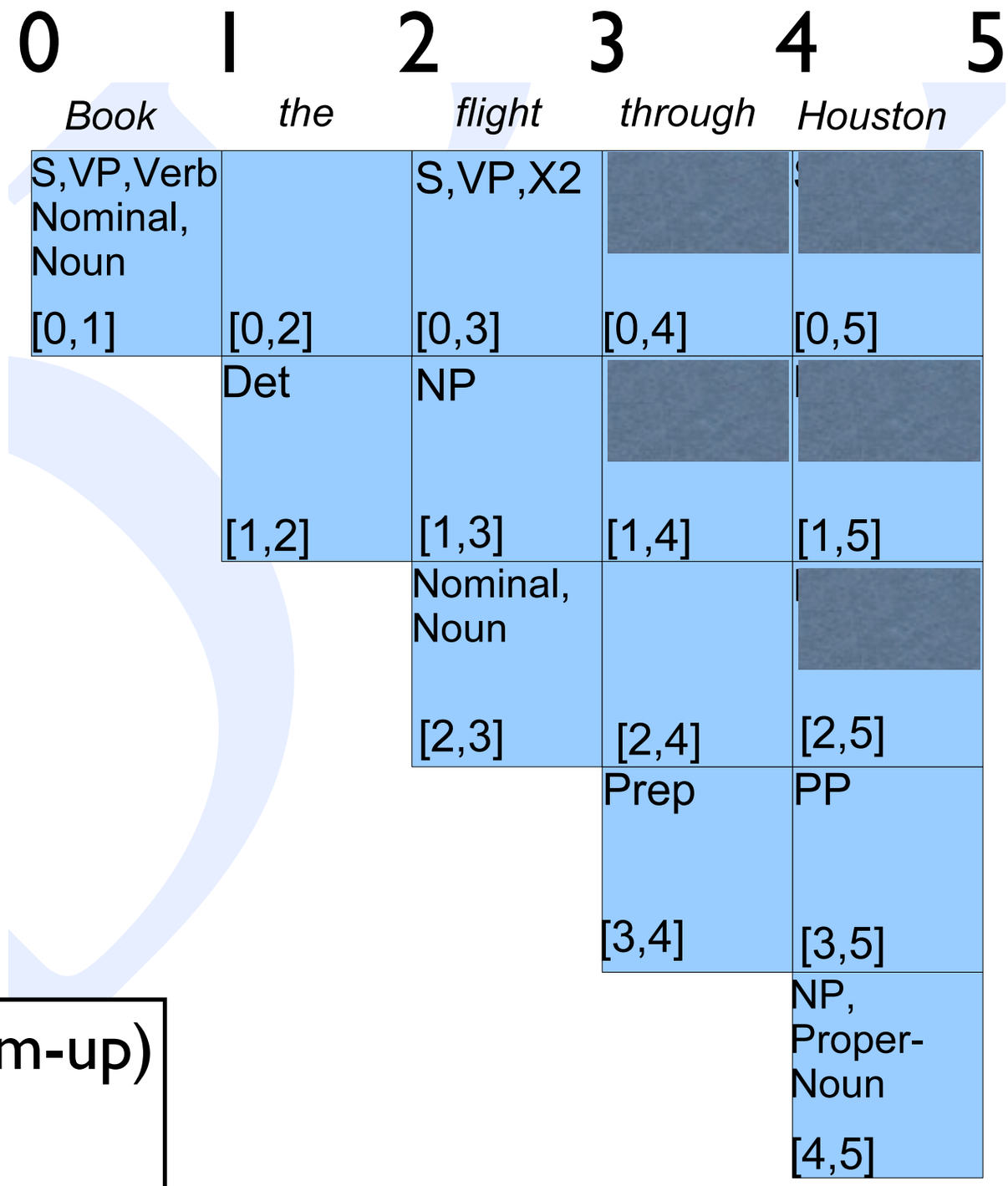


For cell [i,j] (loop through them bottom-up)
  For possible splitpoint k=(i+1)..(j-1):
    For every B in [i,k] and C in [k,j],
      If exists rule A -> B C,
        add A to cell [i,j]

(J&M has a slightly different cell ordering. Both OK.)

CKY recognizer

$S \rightarrow NP\ VP$
$S \rightarrow Aux\ NP\ VP$
$S \rightarrow VP$
$NP \rightarrow Pronoun$
$NP \rightarrow Proper\text{-}Noun$
$NP \rightarrow Det\ Nominal$
$Nominal \rightarrow Noun$
$Nominal \rightarrow Nominal\ Noun$
$Nominal \rightarrow Nominal\ PP$
$VP \rightarrow Verb$
$VP \rightarrow Verb\ NP$
$VP \rightarrow Verb\ NP\ PP$
$VP \rightarrow Verb\ PP$
$VP \rightarrow VP\ PP$
$PP \rightarrow Preposition\ NP$



For cell [i,j] (loop through them bottom-up)
  For possible splitpoint k=(i+1)..(j-1):
    For every B in [i,k] and C in [k,j],
      If exists rule A -> B C,
        add A to cell [i,j]

(J&M has a slightly different cell ordering. Both OK.)

CKY recognizer

$S \rightarrow NP\ VP$
$S \rightarrow Aux\ NP\ VP$
$S \rightarrow VP$
$NP \rightarrow Pronoun$
$NP \rightarrow Proper\text{-}Noun$
$NP \rightarrow Det\ Nominal$
$Nominal \rightarrow Noun$
$Nominal \rightarrow Nominal\ Noun$
$Nominal \rightarrow Nominal\ PP$
$VP \rightarrow Verb$
$VP \rightarrow Verb\ NP$
$VP \rightarrow Verb\ NP\ PP$
$VP \rightarrow Verb\ PP$
$VP \rightarrow VP\ PP$
$PP \rightarrow Preposition\ NP$

0	1	2	3	4	5
Book	*the*	*flight*	*through*	*Houston*	
S,VP,Verb Nominal, Noun [0,1]	[0,2]	S,VP,X2 [0,3]	[0,4]	[0,5]	
	Det [1,2]	NP [1,3]	[1,4]	[1,5]	
		Nominal, Noun [2,3]	[2,4]	[2,5]	
			Prep [3,4]	PP [3,5]	
				NP, Proper-Noun [4,5]	

For cell [i,j] (loop through them bottom-up)
  For possible splitpoint k=(i+1)..(j-1):
    For every B in [i,k] and C in [k,j],
      If exists rule A -> B C,
        add A to cell [i,j]

(J&M has a slightly different cell ordering. Both OK.)

CKY recognizer

$S \rightarrow NP\ VP$
$S \rightarrow Aux\ NP\ VP$
$S \rightarrow VP$
$NP \rightarrow Pronoun$
$NP \rightarrow Proper\text{-}Noun$
$NP \rightarrow Det\ Nominal$
$Nominal \rightarrow Noun$
$Nominal \rightarrow Nominal\ Noun$
$Nominal \rightarrow Nominal\ PP$
$VP \rightarrow Verb$
$VP \rightarrow Verb\ NP$
$VP \rightarrow Verb\ NP\ PP$
$VP \rightarrow Verb\ PP$
$VP \rightarrow VP\ PP$
$PP \rightarrow Preposition\ NP$

0	1	2	3	4	5
Book	*the*	*flight*	*through*	*Houston*	
S,VP,Verb Nominal, Noun [0,1]	[0,2]	S,VP,X2 [0,3]	[0,4]	[0,5]	
	Det [1,2]	NP [1,3]	[1,4]	[1,5]	
		Nominal, Noun [2,3]	[2,4]	Nominal [2,5]	
			Prep [3,4]	PP [3,5]	
				NP, Proper-Noun [4,5]	

For cell [i,j] (loop through them bottom-up)
  For possible splitpoint k=(i+1)..(j-1):
    For every B in [i,k] and C in [k,j],
      If exists rule A -> B C,
        add A to cell [i,j]

(J&M has a slightly different cell ordering. Both OK.)

CKY recognizer

$S \rightarrow NP\ VP$
$S \rightarrow Aux\ NP\ VP$
$S \rightarrow VP$
$NP \rightarrow Pronoun$
$NP \rightarrow Proper\text{-}Noun$
$NP \rightarrow Det\ Nominal$
$Nominal \rightarrow Noun$
$Nominal \rightarrow Nominal\ Noun$
$Nominal \rightarrow Nominal\ PP$
$VP \rightarrow Verb$
$VP \rightarrow Verb\ NP$
$VP \rightarrow Verb\ NP\ PP$
$VP \rightarrow Verb\ PP$
$VP \rightarrow VP\ PP$
$PP \rightarrow Preposition\ NP$

0	1	2	3	4	5
Book	*the*	*flight*	*through*	*Houston*	
S,VP,Verb Nominal, Noun [0,1]	[0,2]	S,VP,X2 [0,3]	[0,4]	[0,5]	
	Det [1,2]	NP [1,3]	[1,4]	[1,5]	
		Nominal, Noun [2,3]	[2,4]	Nominal [2,5]	
			Prep [3,4]	PP [3,5]	
				NP, Proper-Noun [4,5]	

For cell [i,j] (loop through them bottom-up)
  For possible splitpoint k=(i+1)..(j-1):
    For every B in [i,k] and C in [k,j],
      If exists rule A -> B C,
        add A to cell [i,j]

(J&M has a slightly different cell ordering. Both OK.)

CKY recognizer

$S \rightarrow NP\ VP$
$S \rightarrow Aux\ NP\ VP$
$S \rightarrow VP$
$NP \rightarrow Pronoun$
$NP \rightarrow Proper\text{-}Noun$
$NP \rightarrow Det\ Nominal$
$Nominal \rightarrow Noun$
$Nominal \rightarrow Nominal\ Noun$
$Nominal \rightarrow Nominal\ PP$
$VP \rightarrow Verb$
$VP \rightarrow Verb\ NP$
$VP \rightarrow Verb\ NP\ PP$
$VP \rightarrow Verb\ PP$
$VP \rightarrow VP\ PP$
$PP \rightarrow Preposition\ NP$

0	1	2	3	4	5
Book	*the*	*flight*	*through*	*Houston*	
S,VP,Verb Nominal, Noun [0,1]	[0,2]	S,VP,X2 [0,3]	[0,4]	[0,5]	
	Det [1,2]	NP [1,3]	[1,4]	NP [1,5]	
		Nominal, Noun [2,3]	[2,4]	Nominal [2,5]	
			Prep [3,4]	PP [3,5]	
				NP, Proper-Noun [4,5]	

For cell [i,j] (loop through them bottom-up)
  For possible splitpoint k=(i+1)..(j-1):
    For every B in [i,k] and C in [k,j],
      If exists rule A -> B C,
        add A to cell [i,j]

(J&M has a slightly different cell ordering. Both OK.)

CKY recognizer

$S \rightarrow NP\ VP$
$S \rightarrow Aux\ NP\ VP$
$S \rightarrow VP$
$NP \rightarrow Pronoun$
$NP \rightarrow Proper\text{-}Noun$
$NP \rightarrow Det\ Nominal$
$Nominal \rightarrow Noun$
$Nominal \rightarrow Nominal\ Noun$
$Nominal \rightarrow Nominal\ PP$
$VP \rightarrow Verb$
$VP \rightarrow Verb\ NP$
$VP \rightarrow Verb\ NP\ PP$
$VP \rightarrow Verb\ PP$
$VP \rightarrow VP\ PP$
$PP \rightarrow Preposition\ NP$

0	1	2	3	4	5
Book	*the*	*flight*	*through*	*Houston*	
S,VP,Verb Nominal, Noun [0,1]	[0,2]	S,VP,X2 [0,3]	[0,4]	S, VP [0,5]	
	Det [1,2]	NP [1,3]	[1,4]	NP [1,5]	
		Nominal, Noun [2,3]	[2,4]	Nominal [2,5]	
			Prep [3,4]	PP [3,5]	
				NP, Proper-Noun [4,5]	

For cell [i,j] (loop through them bottom-up)
  For possible splitpoint k=(i+1)..(j-1):
    For every B in [i,k] and C in [k,j],
      If exists rule A -> B C,
        add A to cell [i,j]

(J&M has a slightly different cell ordering. Both OK.)

CKY recognizer

$S \rightarrow NP\ VP$
$S \rightarrow Aux\ NP\ VP$
$S \rightarrow VP$
$NP \rightarrow Pronoun$
$NP \rightarrow Proper\text{-}Noun$
$NP \rightarrow Det\ Nominal$
$Nominal \rightarrow Noun$
$Nominal \rightarrow Nominal\ Noun$
$Nominal \rightarrow Nominal\ PP$
$VP \rightarrow Verb$
$VP \rightarrow Verb\ NP$
$VP \rightarrow Verb\ NP\ PP$
$VP \rightarrow Verb\ PP$
$VP \rightarrow VP\ PP$
$PP \rightarrow Preposition\ NP$

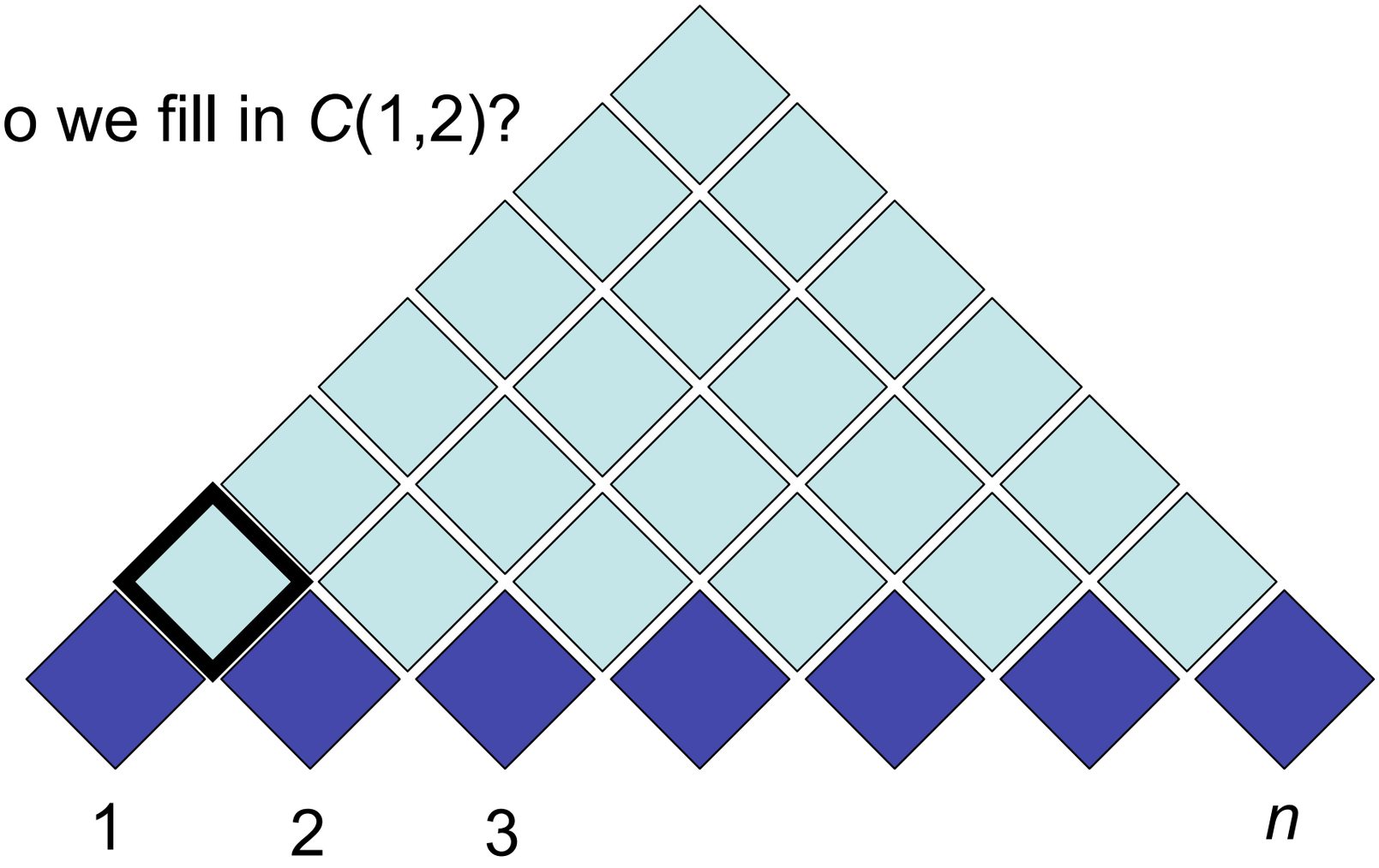
0	1	2	3	4	5
Book	*the*	*flight*	*through*	*Houston*	
S,VP,Verb Nominal, Noun [0,1]	[0,2]	S,VP,X2 [0,3]	[0,4]	S, VP [0,5]	
	Det [1,2]	NP [1,3]	[1,4]	NP [1,5]	
		Nominal, Noun [2,3]	[2,4]	Nominal [2,5]	
			Prep [3,4]	PP [3,5]	
				NP, Proper-Noun [4,5]	

For cell [i,j] (loop through them bottom-up)
  For possible splitpoint k=(i+1)..(j-1):
    For every B in [i,k] and C in [k,j],
      If exists rule A -> B C,
        add A to cell [i,j]

(J&M has a slightly different cell ordering. Both OK.)

For cell [i,j] (loop through them bottom-up)
  For possible splitpoint k=(i+1)..(j-1):
    For every B in [i,k] and C in [k,j],
      If exists rule A -> B C,
        add A to cell [i,j]

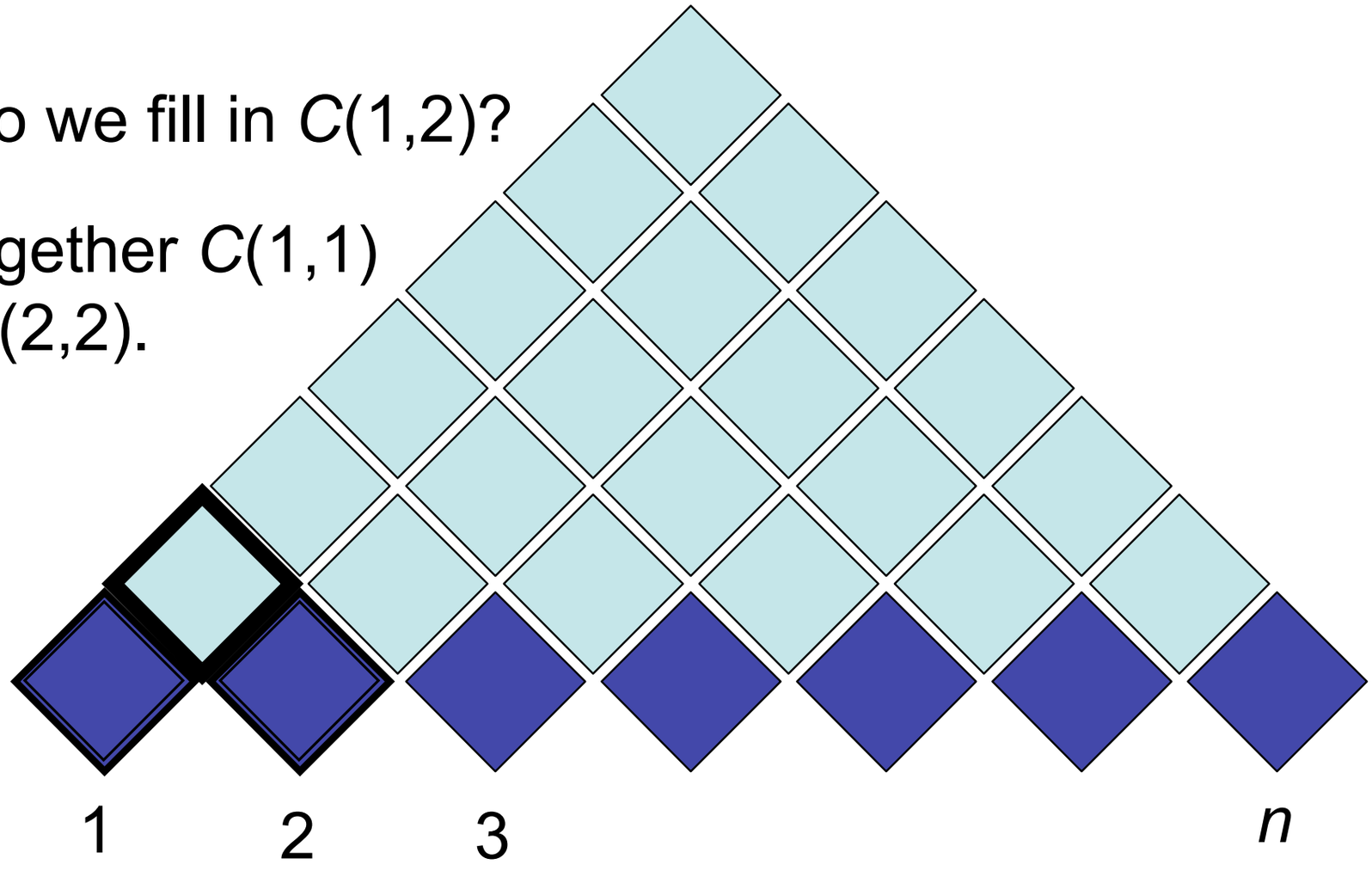
How do we fill in *C*(1,2)?



```
For cell [i,j]  (loop through them bottom-up)
    For possible splitpoint k=(i+1)..(j-1):
        For every B in [i,k] and C in [k,j],
            If exists rule A -> B C,
                add  A to cell [i,j]
```

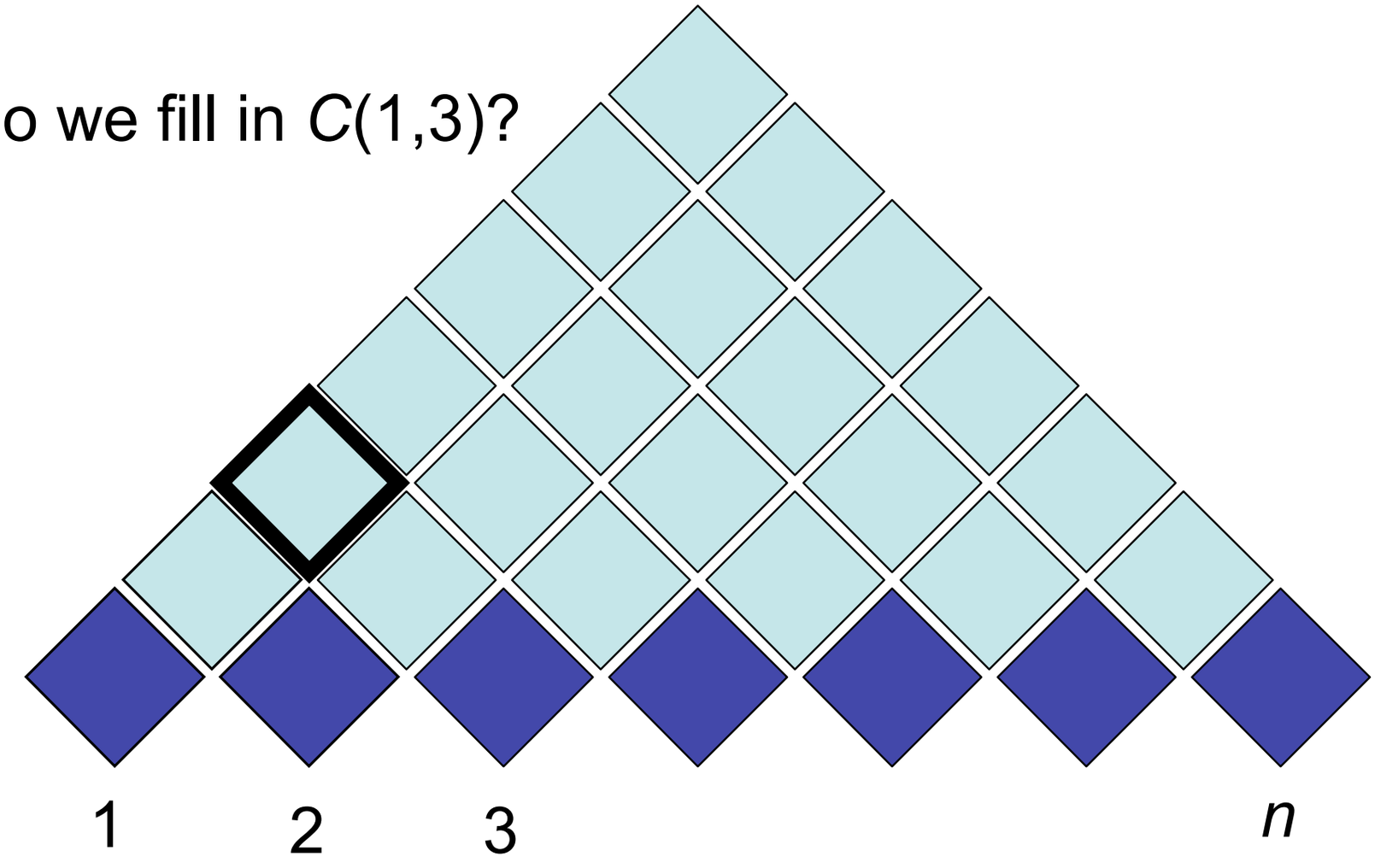
How do we fill in *C*(1,2)?

Put together *C*(1,1) and *C*(2,2).



For cell [i,j] (loop through them bottom-up)
  For possible splitpoint k=(i+1)..(j-1):
    For every B in [i,k] and C in [k,j],
      If exists rule A -> B C,
        add A to cell [i,j]

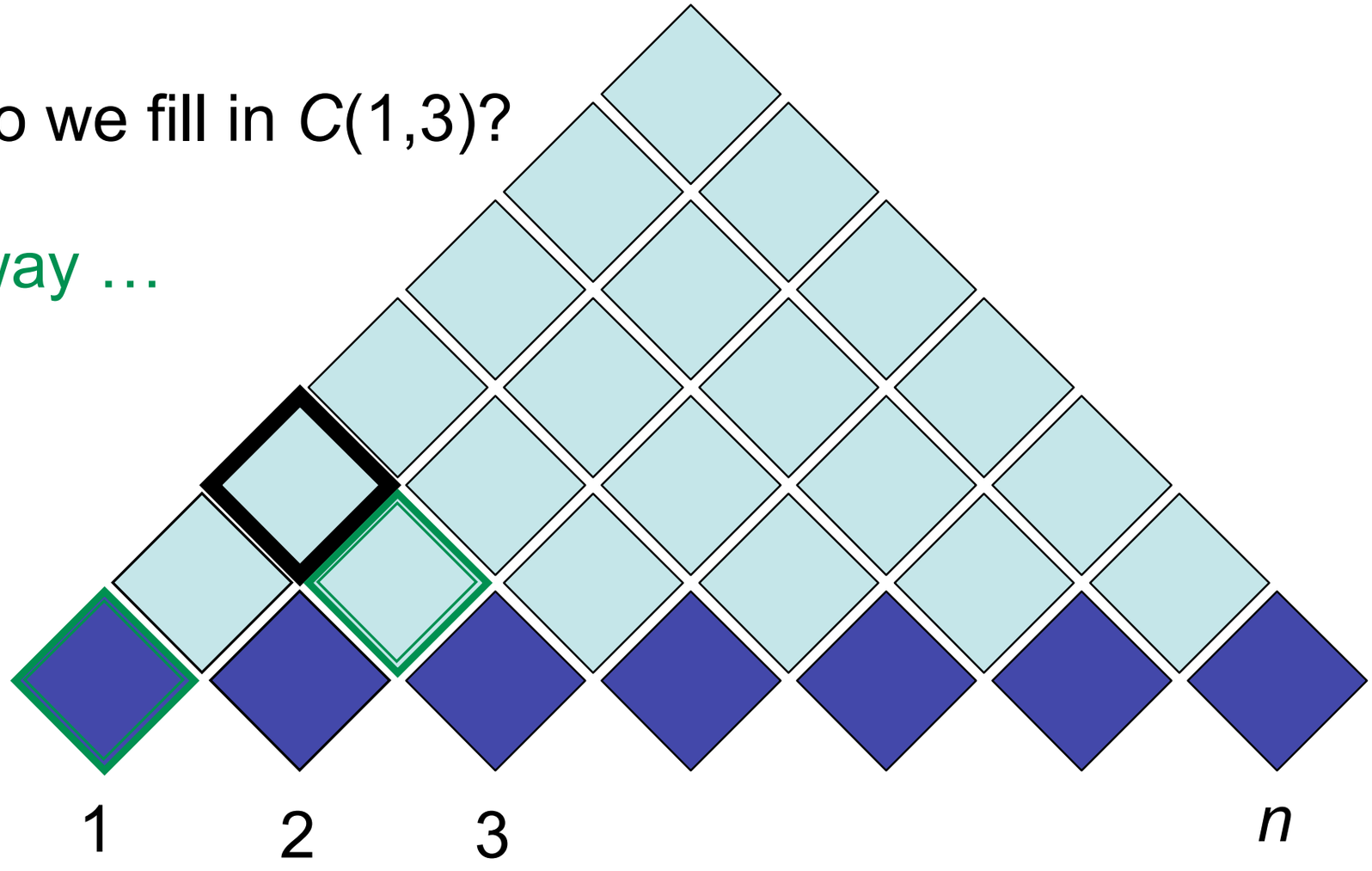
How do we fill in *C*(1,3)?



```
For cell [i,j]  (loop through them bottom-up)
    For possible splitpoint k=(i+1)..(j-1):
        For every B in [i,k] and C in [k,j],
            If exists rule A -> B C,
                add  A to cell [i,j]
```

How do we fill in *C*(1,3)?

One way …

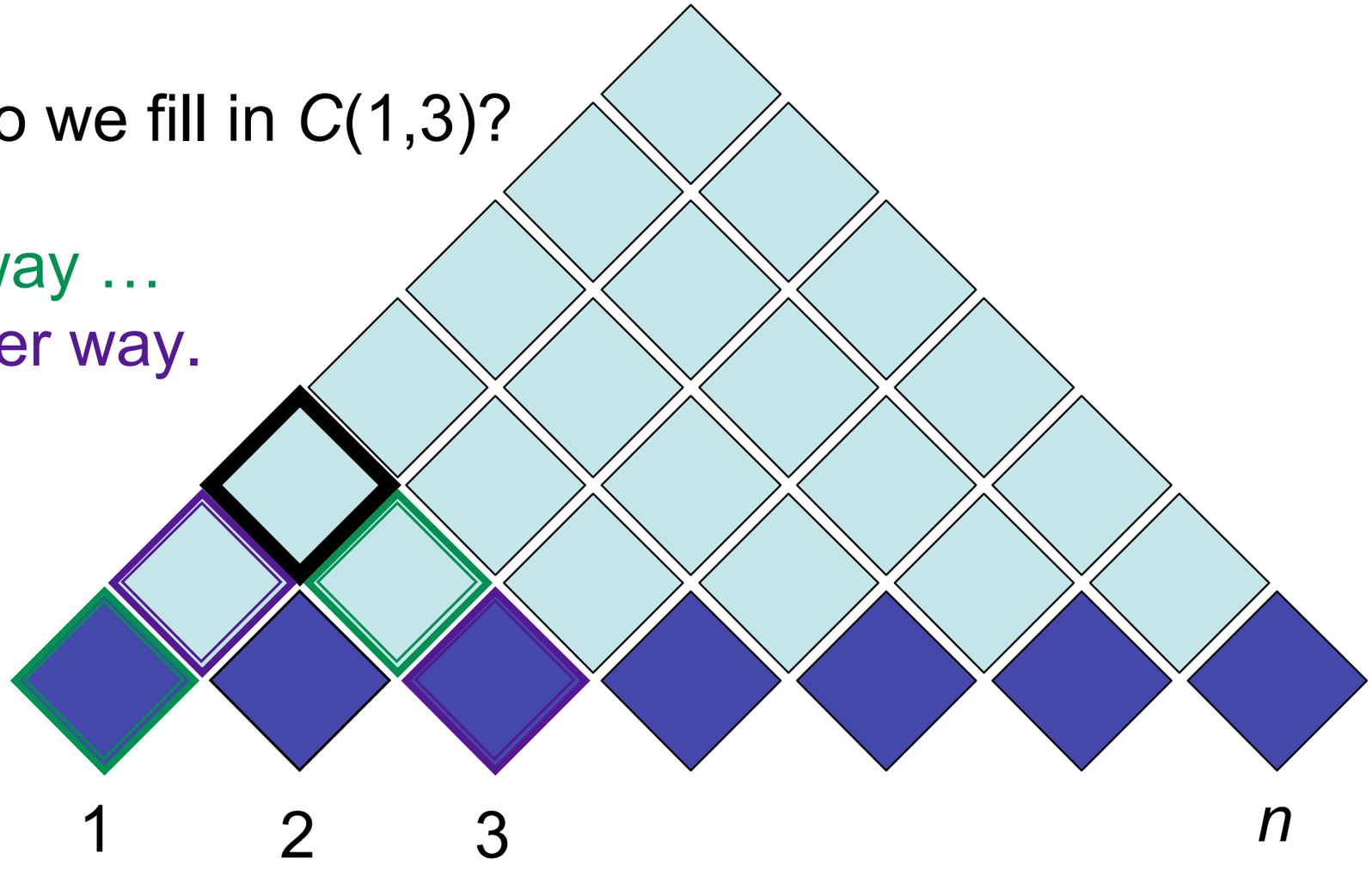


```
For cell [i,j]  (loop through them bottom-up)
    For possible splitpoint k=(i+1)..(j-1):
        For every B in [i,k] and C in [k,j],
            If exists rule A -> B C,
                add  A to cell [i,j]
```

How do we fill in *C*(1,3)?

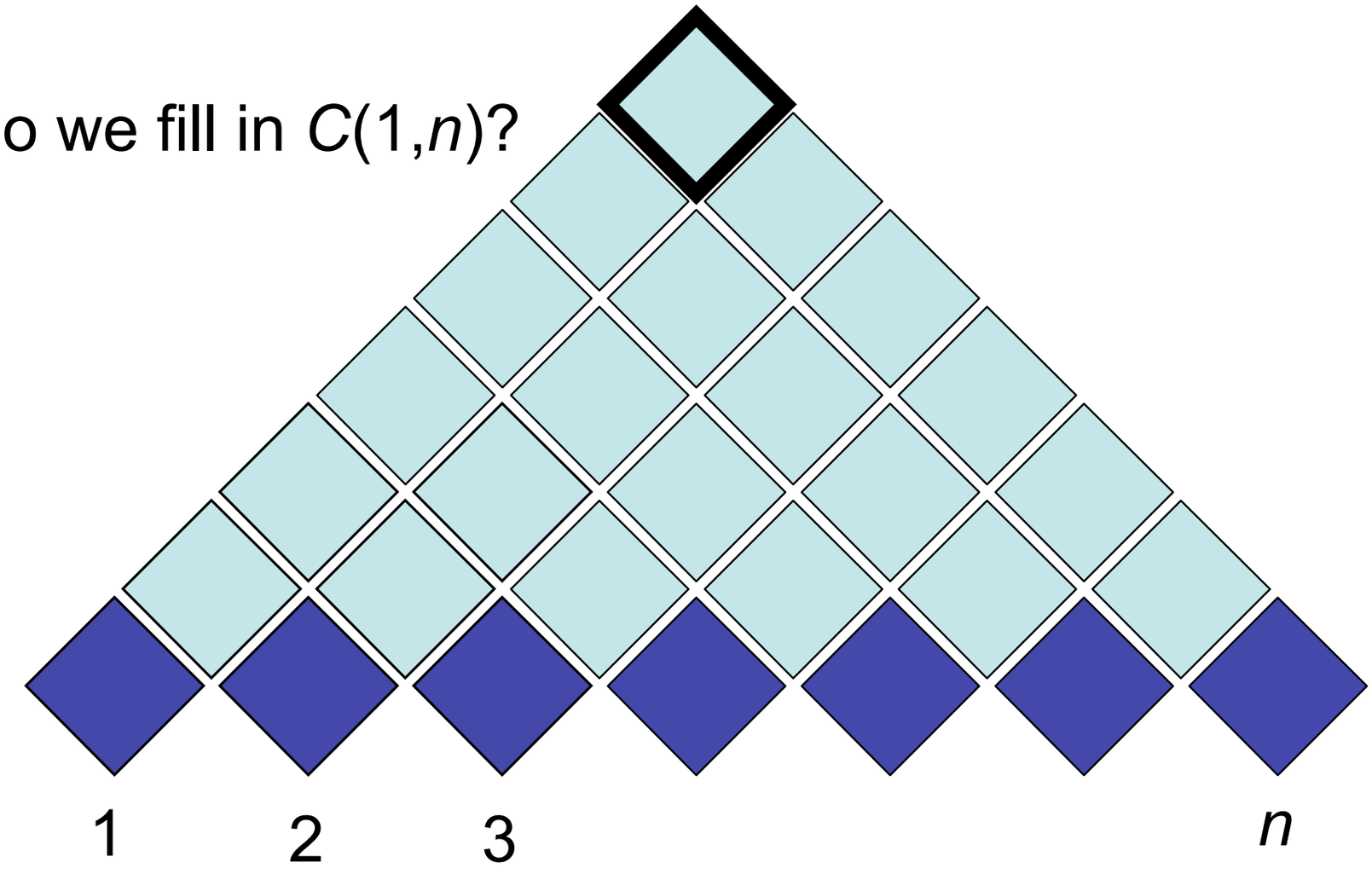
One way …

Another way.



```
For cell [i,j]  (loop through them bottom-up)
    For possible splitpoint k=(i+1)..(j-1):
        For every B in [i,k] and C in [k,j],
            If exists rule A -> B C,
                add  A to cell [i,j]
```

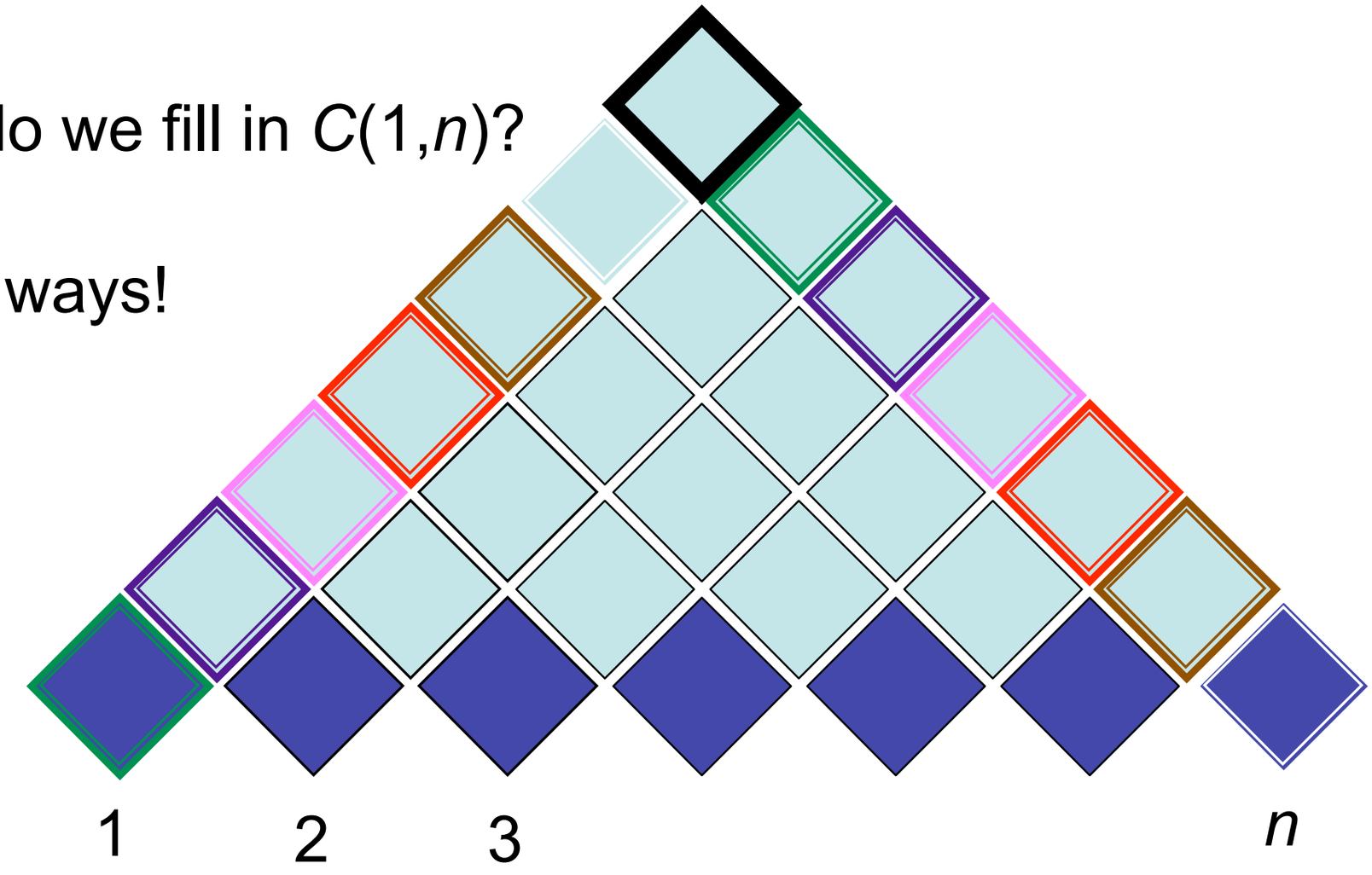
How do we fill in *C*(1,*n*)?



For cell [i,j] (loop through them bottom-up)
  For possible splitpoint k=(i+1)..(j-1):
    For every B in [i,k] and C in [k,j],
      If exists rule A -> B C,
        add A to cell [i,j]

How do we fill in *C*(1,*n*)?

n - 1 ways!



For cell [i,j] (loop through them bottom-up)
  For possible splitpoint k=(i+1)..(j-1):
    For every B in [i,k] and C in [k,j],
      If exists rule A -> B C,
        add A to cell [i,j]

Where N is the number of nonterminals in the grammar, and n is the length of the sentence

$O(n^2)$ cells to fill
up to N items per cell, thus
$O(N\ n^2)$ items to create

and
$O(N^2\ n)$ ways to fill a cell