Computing Theory

COMP 147

Chapter 2: Context-Free Languages
CYK algorithm

Notes on CFL and CFG

- 2 definitions for Context Free Languages
    - Languages definable by CFG
        - nondeterministic PDA defines all the CFL's.
- Equivalence between PDA's and CFG's
    - Given a PDA can construct a CFG and vice versa
- Hierarchy of grammars (context-sensitive, unrestricted)

Non-Context Free Languages

- Example: $L = \{0^n 1^n 2^n | n \geq 0\}$
- How to show a language is not context-free?
  - There is a pumping lemma

Pumping Lemma for CFLs

If A is a CFL,
then $\exists p$, such that $\forall s \in A$,
if $|s| \geq p$, then $\exists u, v, x, y, z$, such that

- $s = uvxyz$
- for each $i \geq 0$, $uv^ixy^iz \in A$
- $|vy| > 0$
- $|vxy| \leq p$

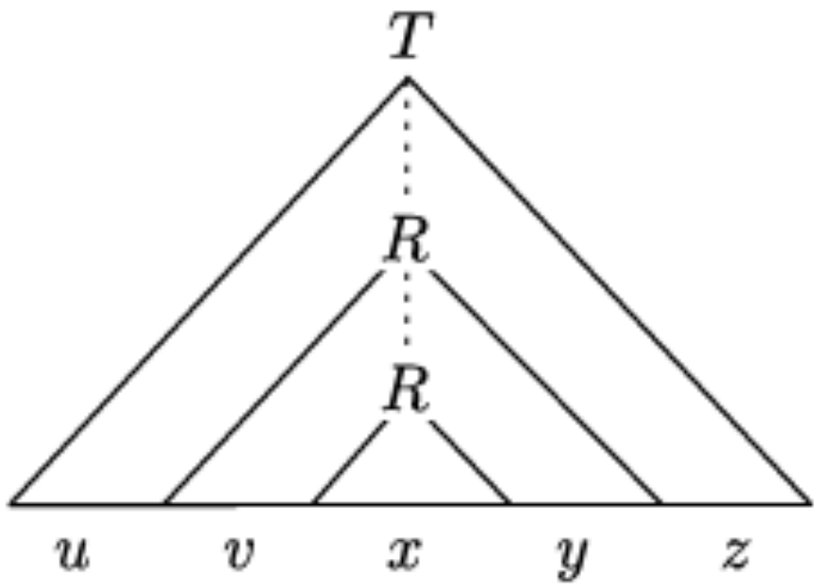
PL for CFL: Proof Idea

Consider a parse tree for $s \in A$.

- For a sufficiently deep tree,
  there must be some repeated variable.

$$R \overset{*}{\Rightarrow} vRy$$

$$R \overset{*}{\Rightarrow} x$$



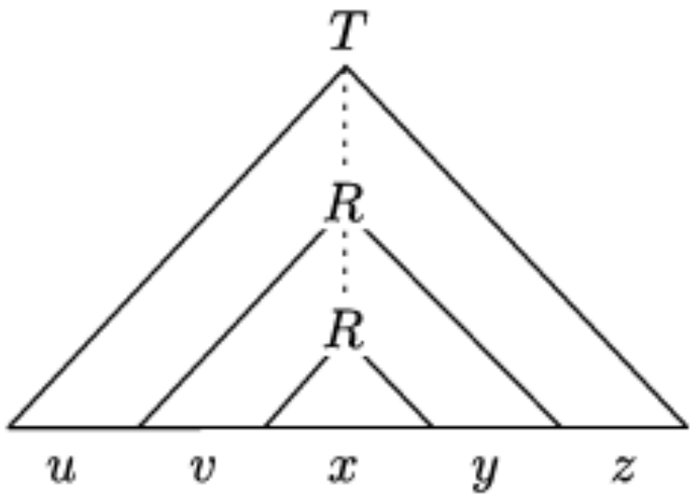
PL for CFL: Proof Idea

Consider a parse tree for $s \in A$.

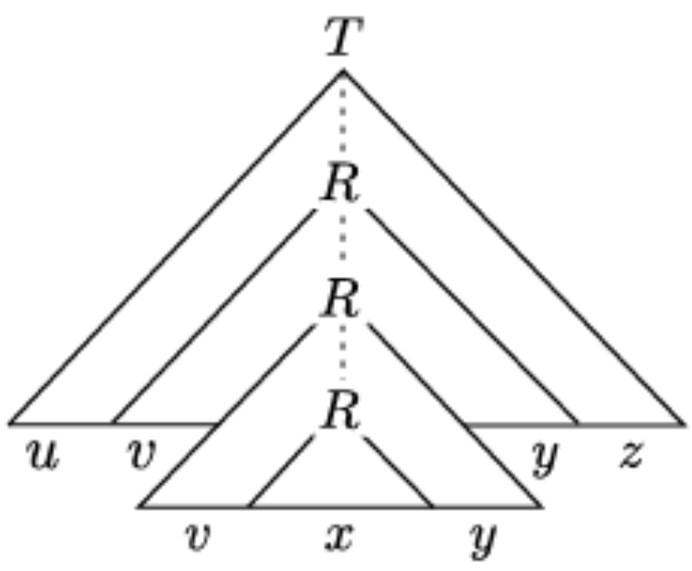
- For a sufficiently deep tree, there must be some repeated variable.
- Pumping adds or removes subtrees

$$R \overset{*}{\Longrightarrow} vRy$$

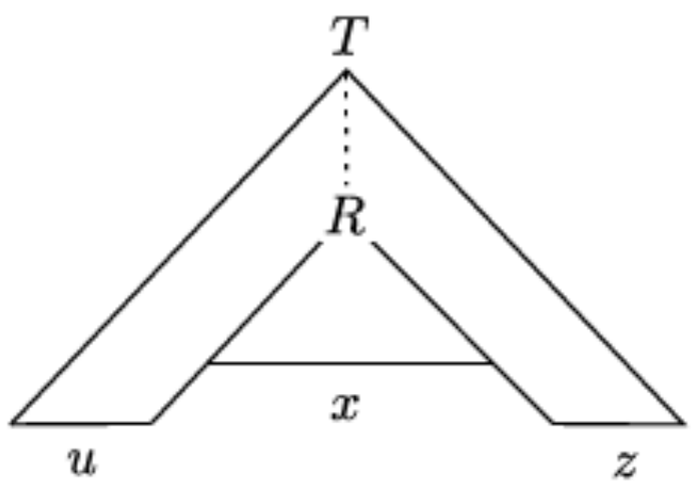
$$R \overset{*}{\Longrightarrow} x$$



uv^1xy^1z



Pump up: uv^2xy^2z



Pump down: uv^0xy^0z

CYK Parsing Algorithm

Cocke-**Y**oung-**K**asami Algorithm for parsing

The CYK Algorithm

- Membership problem:
    - Given grammar and string w, is w in L(G)?
- Bottom-up parsing: start with the string
    - Top-down parsing: Number of subtrees can be exponential
- Dynamic programming:
    - save the results in a table/chart
      re-use these results in finding larger constituents
- Complexity: $O(n^3|G|)$
  n: length of string, $|G|$: size of grammar)
- Presumes G is in CNF

Construct a triangular table

- Each row corresponds to one length of substrings
  - a Bottom Row – Strings of length 1
  - Second from Bottom Row – Strings of length 2
  .
  .
- Top Row – string ‘w’

The CYK Algorithm

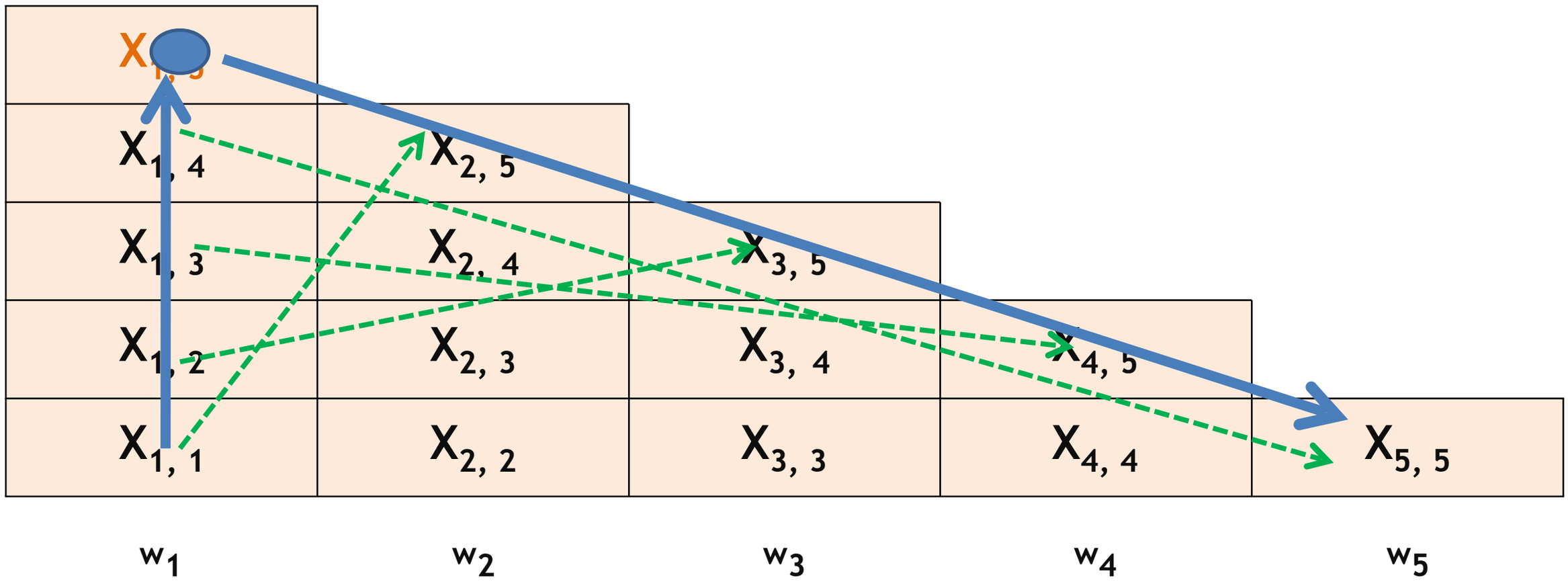
- Notation X_{ij} : set of variables that generate symbols starting from symbol i to symbol j
- A -> BC | AB | 1, B-> AA | 0, C -> CB| 1| 0
- String 110100
- $X_{4,4}$: {A,C}
- $X_{3,3}$:{B, C}
- find $X_{1,6}$ and see if it has A
- Compare at most n pairs of previously computed sets:
  $(X_{i,i}, X_{i+1,j}), (X_{i,i+1}, X_{i+2,j}) \dots (X_{i,j-1}, X_{j,j})$

Construct a Triangular Table

$X_{1,5}$				
$X_{1,4}$	$X_{2,5}$			
$X_{1,3}$	$X_{2,4}$	$X_{3,5}$		
$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,5}$	
$X_{1,1}$	$X_{2,2}$	$X_{3,3}$	$X_{4,4}$	$X_{5,5}$
w_1	w_2	w_3	w_4	w_5

Table for string **‘w’** that has length 5

Construct a Triangular Table



Looking for pairs to compare

Example CYK Algorithm

- Show the CYK Algorithm with the following example:
  - CNF grammar G
    - S ->AB | BC
    - A -> BA | a
    - B -> CC | b
    - C -> AB | a
  - w is baaba
  - Question Is baaba in L(G)?

Constructing The Triangular Table

S → AB | BC
A → BA | a
B → CC | b
C → AB | a

{B}	**{A, C}**	**{A, C}**	**{B}**	**{A, C}**
b	**a**	**a**	**b**	**a**

Calculating the Bottom ROW

Constructing The Triangular Table

S → AB | BC
A → BA | a
B → CC | b
C → AB | a

{S, A}				
{B}	{A, C}	{A, C}	{B}	{A, C}
b	a	a	b	a

Constructing The Triangular Table

S → AB | BC
A → BA | a
B → CC | b
C → AB | a

{S, A}	{B}			
{B}	{A, C}	{A, C}	{B}	{A, C}
b	a	a	b	a

Constructing The Triangular Table

S → AB | BC
A → BA | a
B → CC | b
C → AB | a

{S, A}	{B}	{S, C}		
{B}	{A, C}	{A, C}	{B}	{A, C}
b	a	a	b	a

Constructing The Triangular Table

S → AB | BC
A → BA | a
B → CC | b
C → AB | a

{S, A}	{B}	{S, C}	{S, A}	
{B}	{A, C}	{A, C}	{B}	{A, C}
b	a	a	b	a

Constructing The Triangular Table

S → AB | BC
A → BA | a
B → CC | b
C → AB | a

Ø				
{S, A}	{B}	{S, C}	{S, A}	
{B}	{A, C}	{A, C}	{B}	{A, C}
b	a	a	b	a

Constructing The Triangular Table

S → AB | BC
A → BA | a
B → CC | b
C → AB | a

Ø	{B}			
{S, A}	{B}	{S, C}	{S, A}	
{B}	{A, C}	{A, C}	{B}	{A, C}
b	a	a	b	a

Constructing The Triangular Table

S → AB | BC
A → BA | a
B → CC | b
C → AB | a

Ø	{B}	{B}		
{S, A}	{B}	{S, C}	{S, A}	
{B}	{A, C}	{A, C}	{B}	{A, C}
b	a	a	b	a

Final Triangular Table

S → AB | BC
A → BA | a
B → CC | b
C → AB | a

{S, A, C}	← $X_{1,5}$			
Ø	{S, A, C}			
Ø	{B}	{B}		
{S, A}	{B}	{S, C}	{S, A}	
{B}	{A, C}	{A, C}	{B}	{A, C}
b	a	a	b	a

- Table for string ‘w’ that has length 5
- The algorithm populates the triangular table