CSCI 361 Lecture 9: Pumping Lemma for CFGs

Shikha Singh

Announcements & Logistics

- No exercise for next Tuesday
- **HW 4** was due yesterday: solutions will be posted on GLOW today
- **Reminder:** Midterm 1 in-class on Oct 7
  - Closed book but can ask clarification on definitions
  - Several textbooks will be available for referencing
  - Everything up to HW 4 included
- **Plan for today:**
  - Use some lecture time to go over Pumping Lemma for CFGs
  - Do problems on the practice midterm

Last Time

- Practice with push-down automata
- Equivalence of CFGs and non-deterministic push-down automata

Non-Context-Free Languages

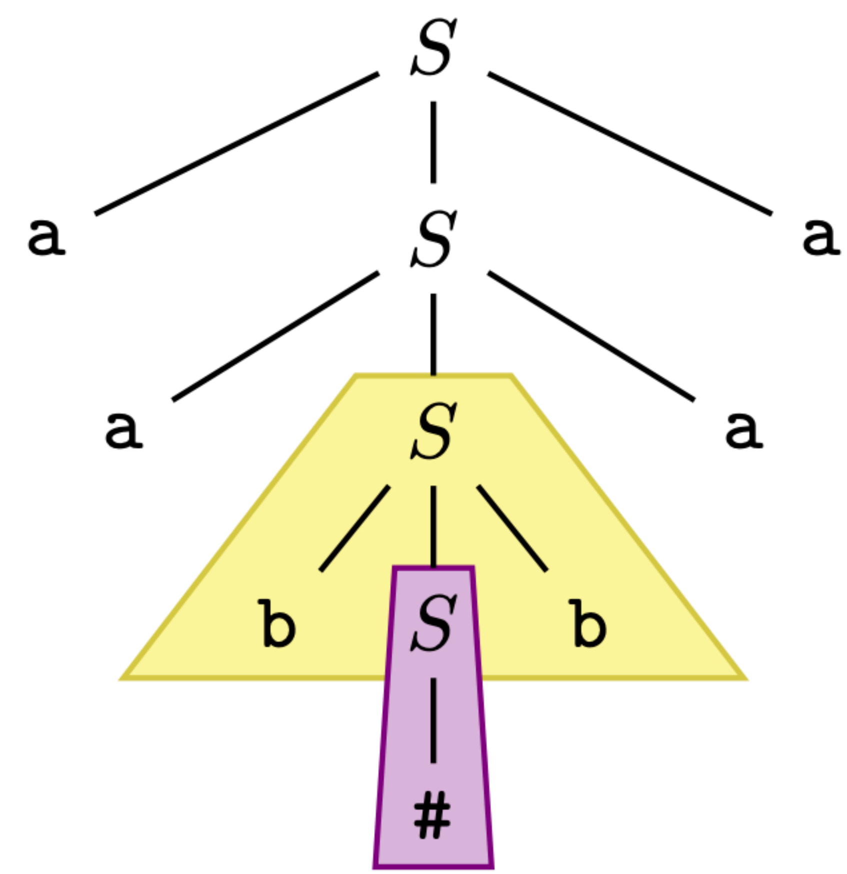
- Proved using a similar "pumping lemma" as regular languages
- With respect to regular languages:
    - pumping lemma exploits the fact that if a string is long enough, a state is repeated in the DFA for the language (loop)
- With respect to CFLs:
    - pumping lemma exploits the fact that if a string is long enough, deriving it requires recursion (repeated use of a variable)
- Lemma based length of parse trees for derivations

Parse Trees and CFGs

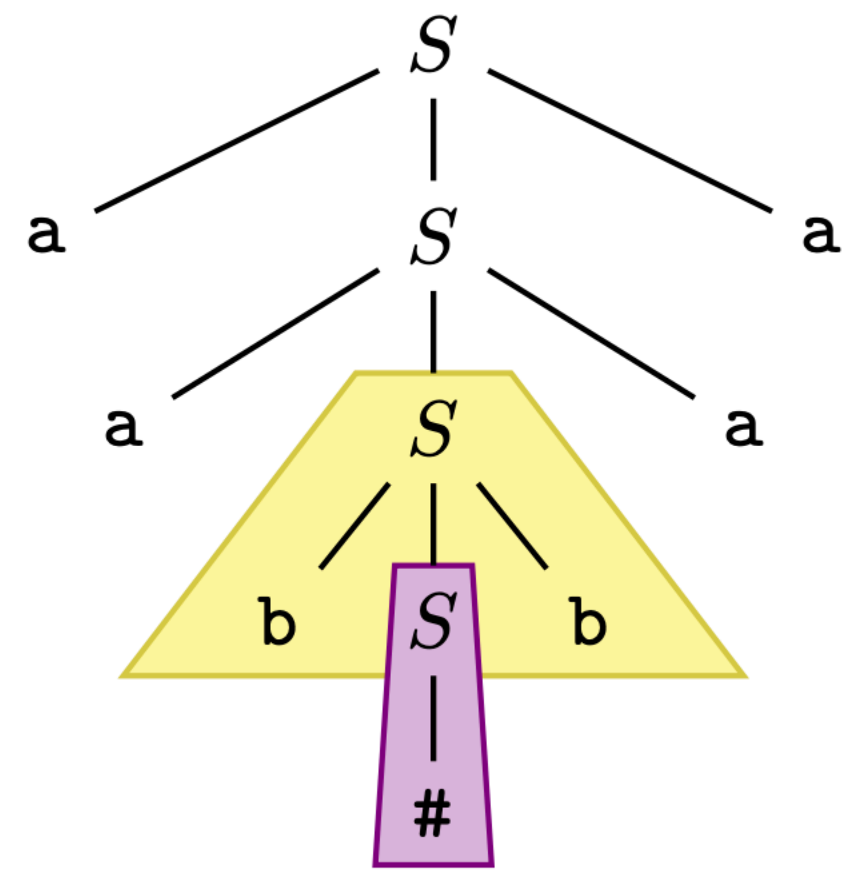
- Consider the CFG for $A = \{w\#w^R \mid w \in \{a,b\}^*\}$:

  $S \rightarrow aSa \mid bSb \mid \#$

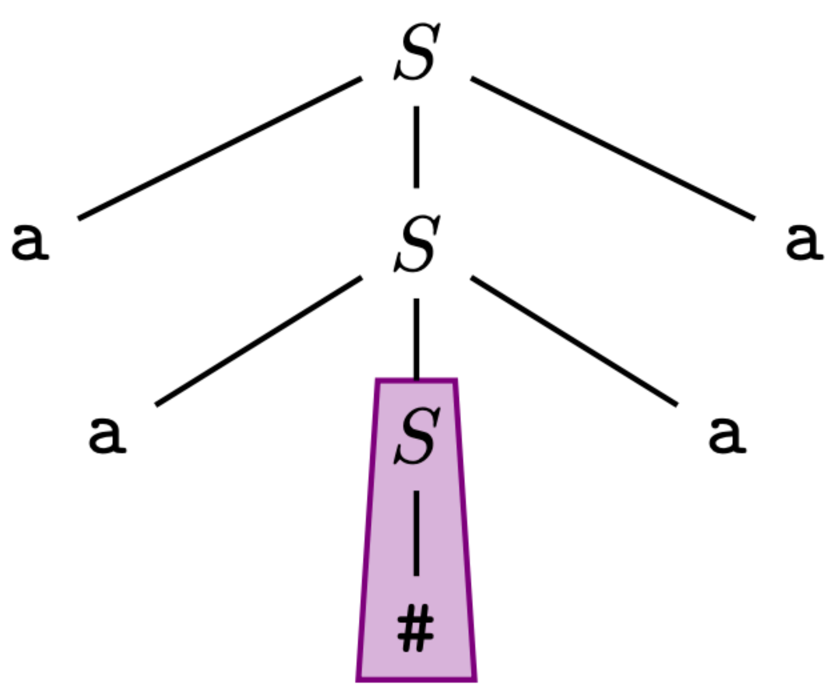
- Consider a parse tree for $w =$ aab#baa

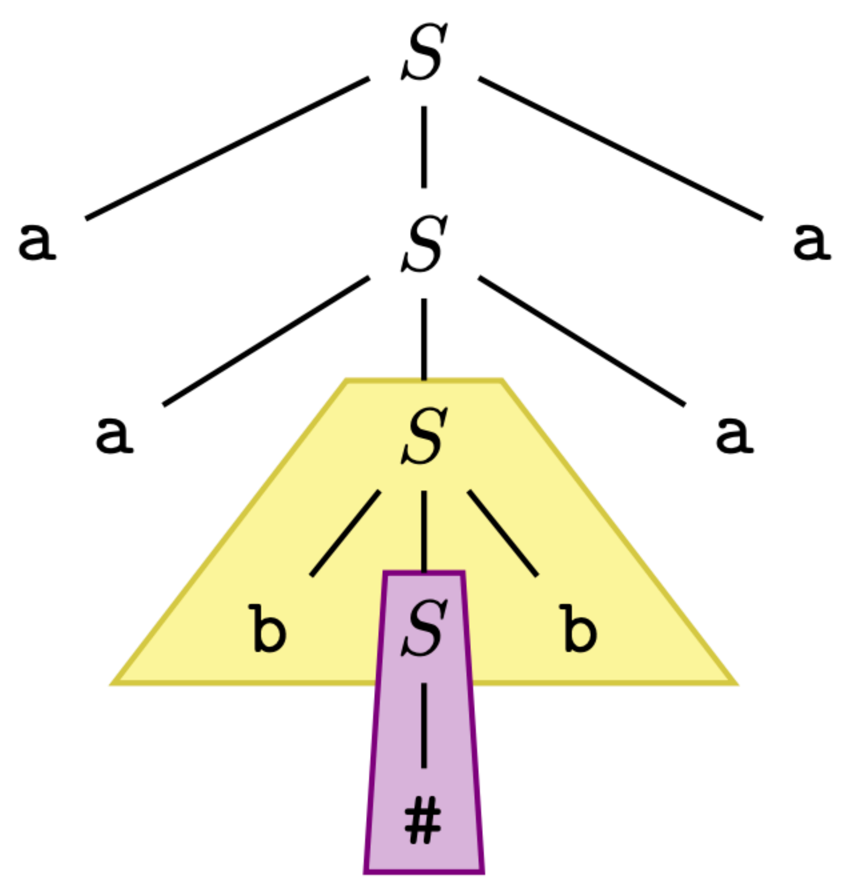


Parse Trees and CFGs



- Variable S is repeated
- Can "pump up" or "pump down" to create strings in the language
  - Replace yellow with violet: aa#aa
  - Replace violet with yellow: aabb#bba





Pumping Lemma: CFLs

- **Statement**: If L is a CFL, then there exists a number p (the pumping length) such that for all $s \in L$ of length at least p, it is possible to divide s into five pieces $s = uvxyz$ satisfying the conditions

    1. $|vy| > 0$
    2. $|vxy| \leq p$
    3. For each $i \geq 0$, $uv^i xy^i z \in L$

- Note that vxy can appear anywhere in the string as long as they are no longer than p symbols long

Proving L cannot be Context-Free

- **Statement**: Consider a language L. If for any number p, there exists an $s \in L$ of length at least p such that it is **impossible** to divide s into five pieces $s = uvxyz$ satisfying all three conditions below

    1. $|vy| > 0$
    2. $|vxy| \leq p$
    3. For each $i \geq 0$, $uv^ixy^iz \in L$

  then L cannot be regular.

Pumping Lemma: Game View

- Defender claims L satisfies pumping lemma
- Challenger claims L does not satisfy pumping lemma

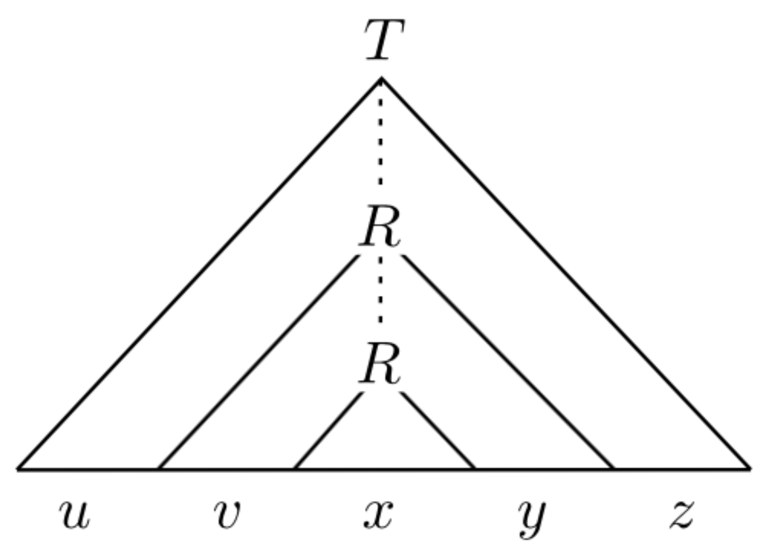
Defender		**Challenger**
Pick pumping length p	$\xrightarrow{p}$	
	$\xleftarrow{z}$	Pick $z \in L$ s.t. $\lvert z \rvert \geq p$
Divide z into u, v, w, x, y s.t. $\lvert vwx \rvert \leq p$, and $\lvert vx \rvert > 0$	$\xrightarrow{u,v,w,x,y}$	
	$\xleftarrow{i}$	Pick i, s.t. $uv^iwx^iy \notin L$

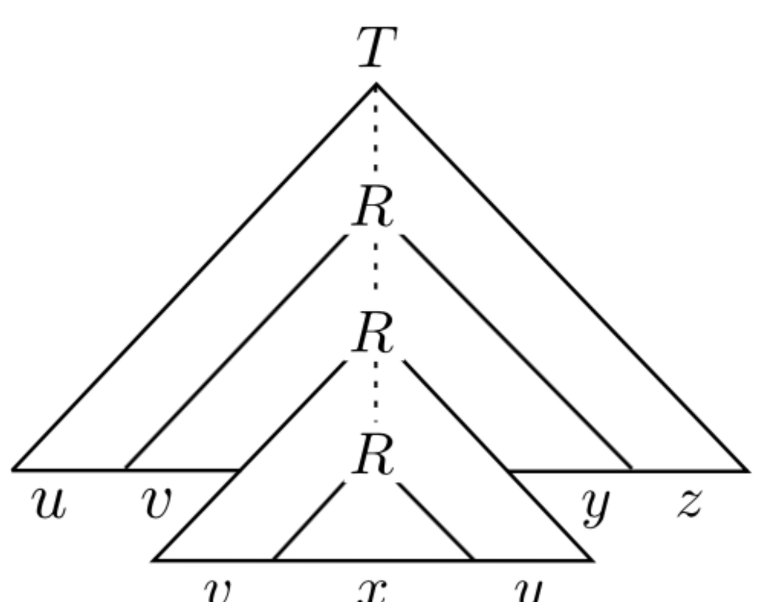
Pumping Lemma: Game View

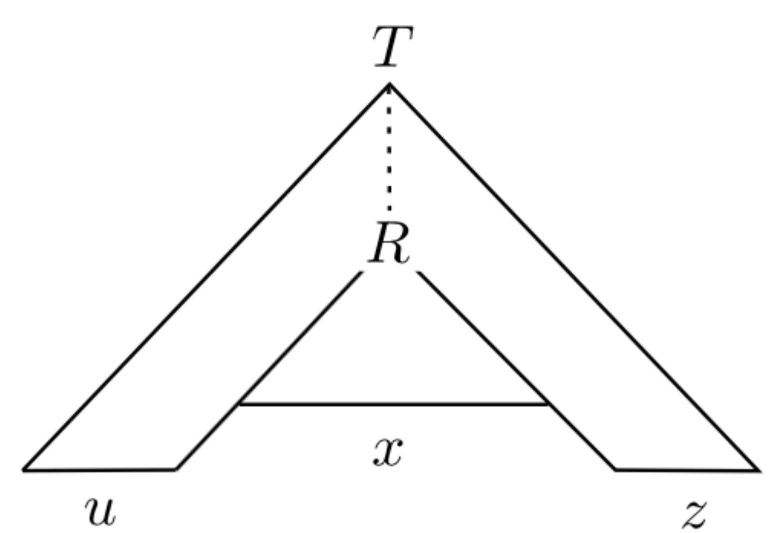
- If L is a CFL: defender has a winning strategy, challenger gets stuck
- If challenger has a winning strategy, L cannot be a CFL

Defender		**Challenger**
Pick pumping length p	$\xrightarrow{p}$	
	$\xleftarrow{z}$	Pick $z \in L$ s.t. $\lvert z \rvert \geq p$
Divide z into u, v, w, x, y s.t. $\lvert vwx \rvert \leq p$, and $\lvert vx \rvert > 0$	$\xrightarrow{u,v,w,x,y}$	
	$\xleftarrow{i}$	Pick i, s.t. $uv^iwx^iy \notin L$

Non-Context-Free Languages







Pumping Lemma (CFL): Intuition

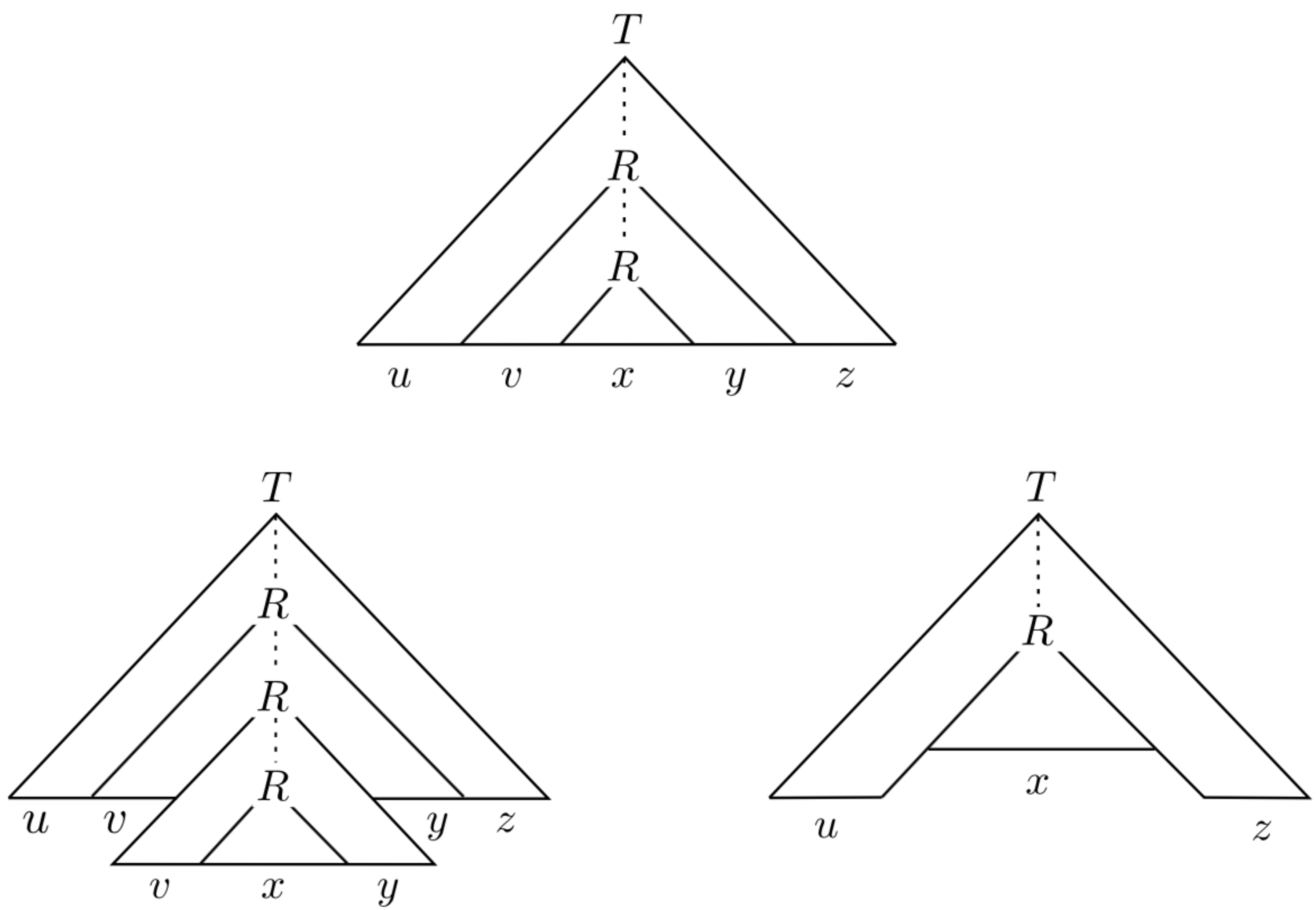
- If the grammar generates a long enough string then the parse tree for that derivation must be "tall enough"
- Let $|V|$ be the number of variables in the CFG and b be the max number of symbols in the RHS of any rule
  - Each node in a parse tree has at most b children
- If the parse tree has height h, what is the max num of leaves it can have?
  - b^h
- If a tree has at least $b^{|V|+1}$ leaves and each node has degree at most b, what can we say about the height?
  - At least $|V| + 1$

Pumping Lemma (CFL): Proof

- Consider a CFG G and let b be the maximum number of symbols on the RHS of G
- Let $|V|$ be the number of variables
- Consider a $w \in L(G)$ of length at least $b^{|V|+1}$
- Consider the derivation of w in the smallest parse tree
  - Each node has at most b children
  - Num of leaves = $|w| \geq b^{|V|+1}$
- What can we conclude about the height of the parse tree?
  - Longest path from root to leaf (height) is at least $|V| + 1$

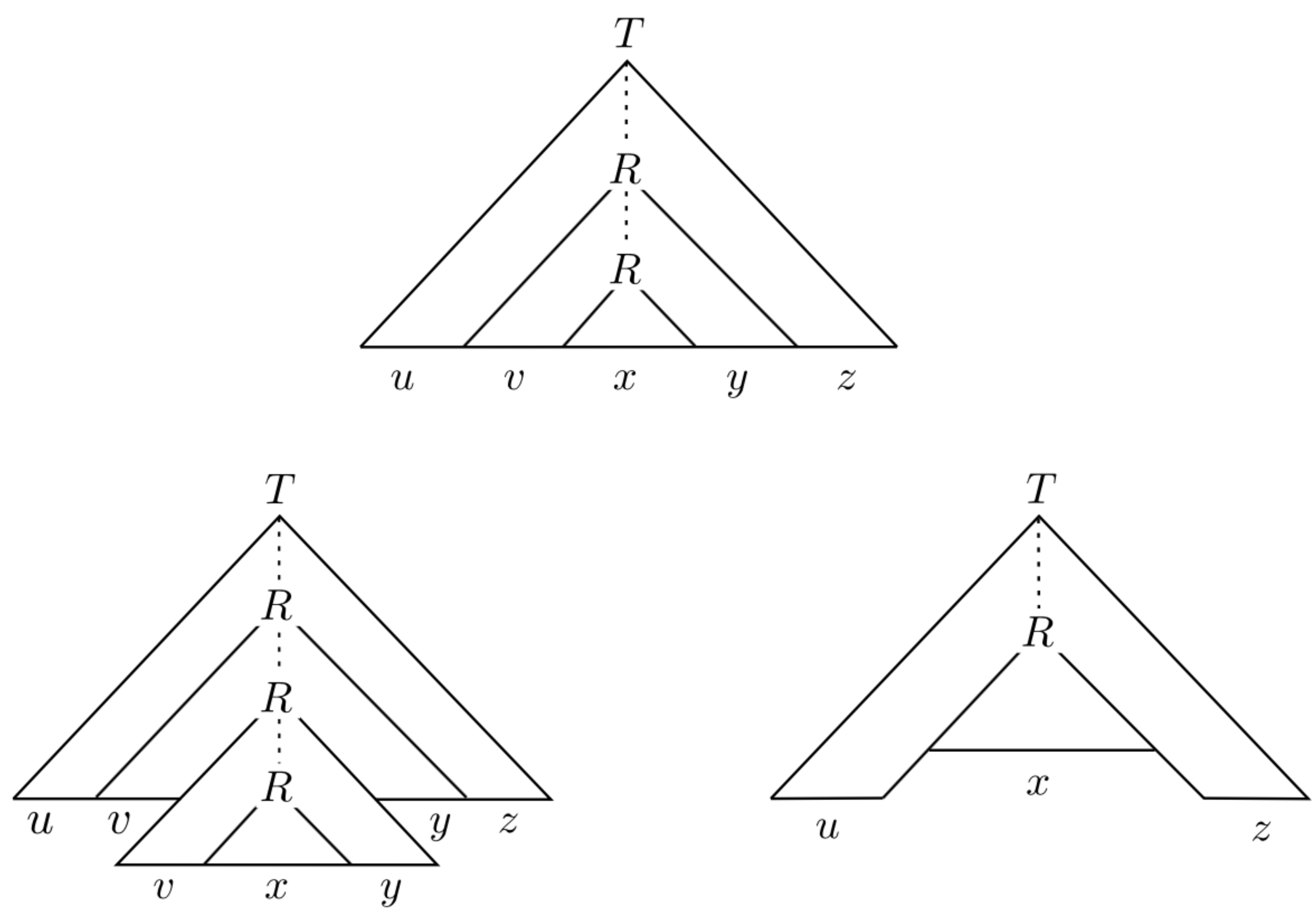
Non-Context-Free Languages

- Number of variables in a path with $|V| + 1$ edges is $|V| + 1$
- Some variable must be repeated in this derivation



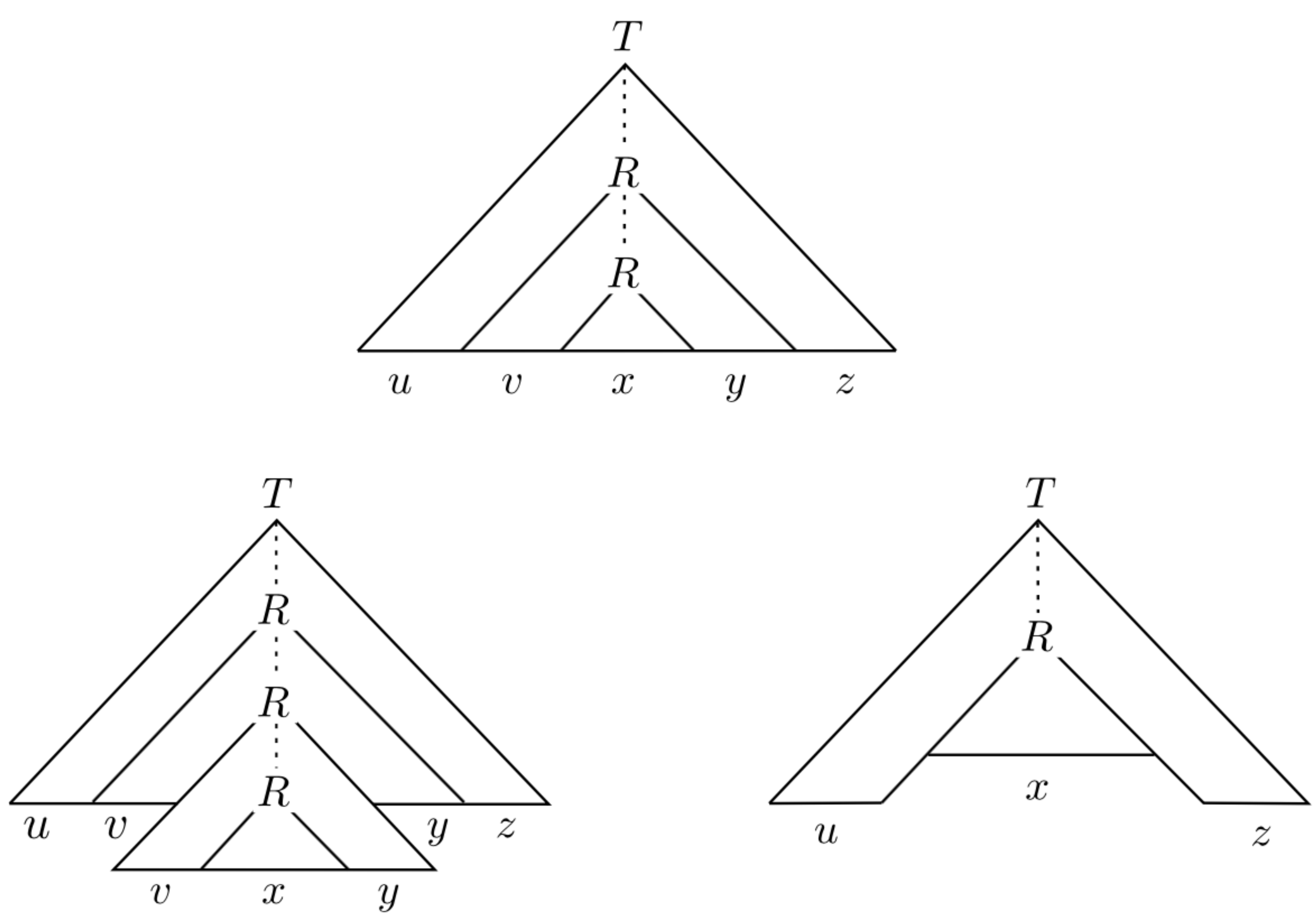
Non-Context-Free Languages

- Consider the smallest-parse tree generating s and let R be the a variable that repeats among the lowest $|V|+1$ variables onthe longest root to leaf path in the parse tree



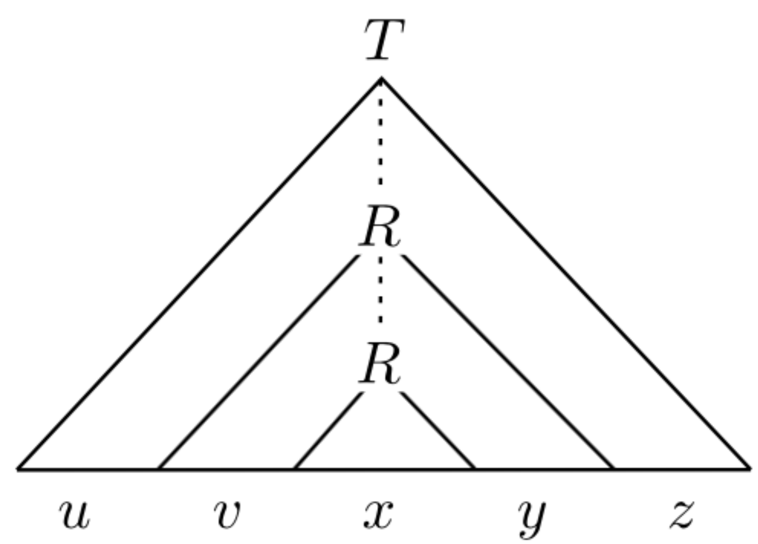
Non-Context-Free Languages

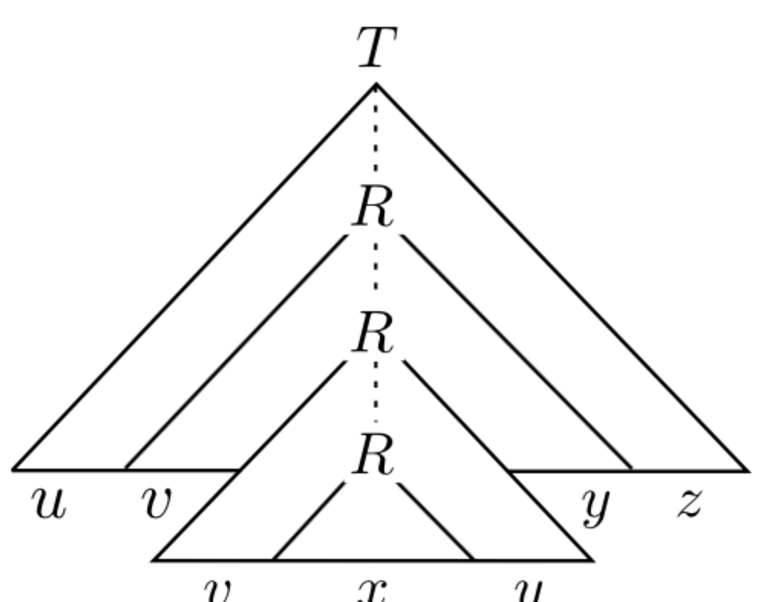
- Let the upper occurrence generate a substring of ***s*** of the form ***vxy***
- Overall the string ***s*** must contain ***vxy*** and is of the form ***uvxyv***

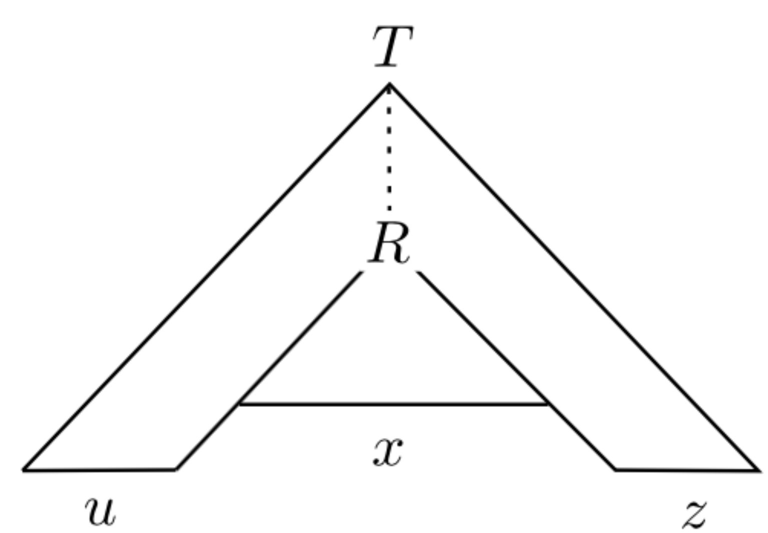


Non-Context-Free Languages

- **Takeaway:** Can replace the smaller subtree under the second occurrence of R with the larger one and still have a valid derivation

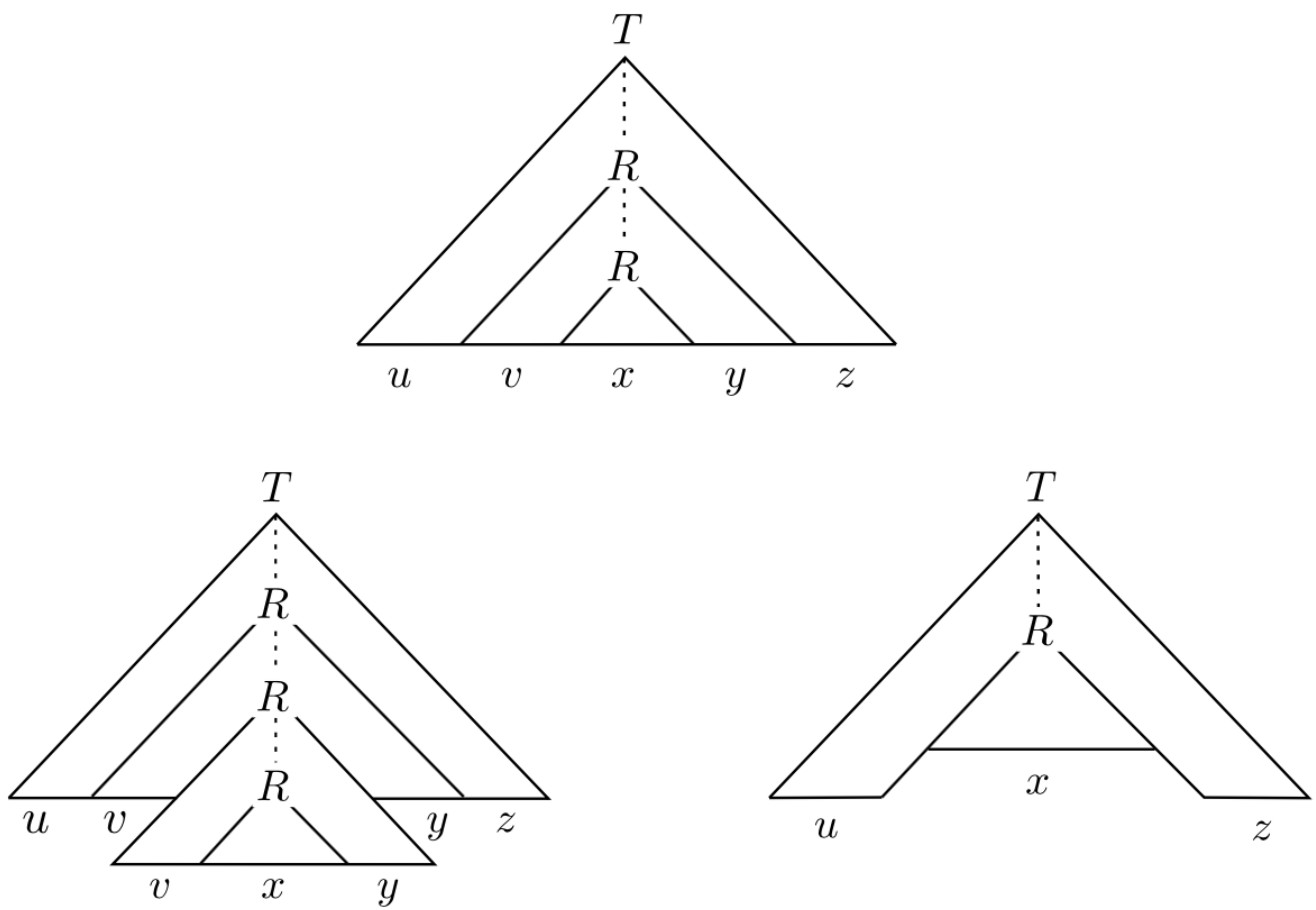






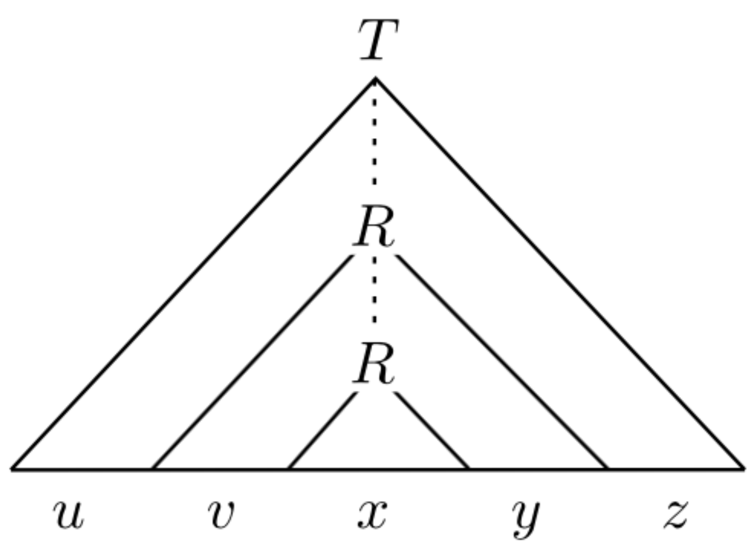
Non-Context-Free Languages

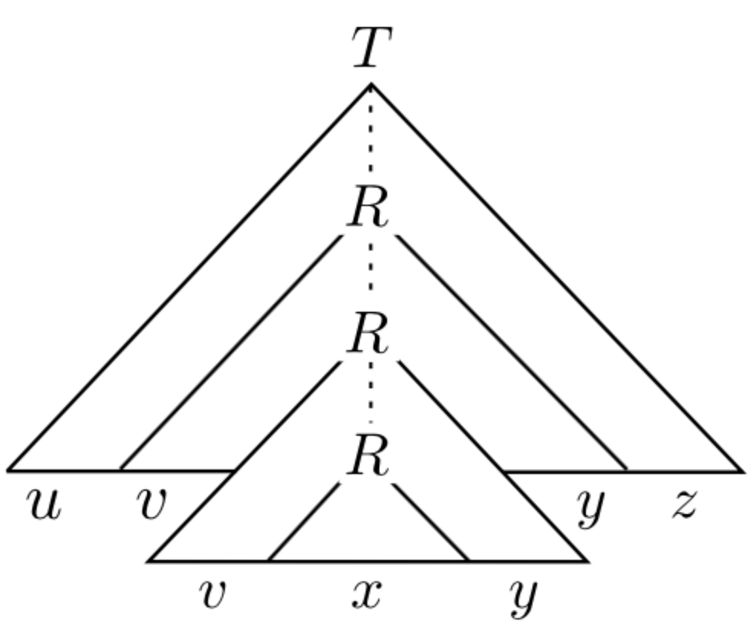
- **Condition 3:** Strings of the form uv^ixy^iz and uxy should all be valid strings in the language

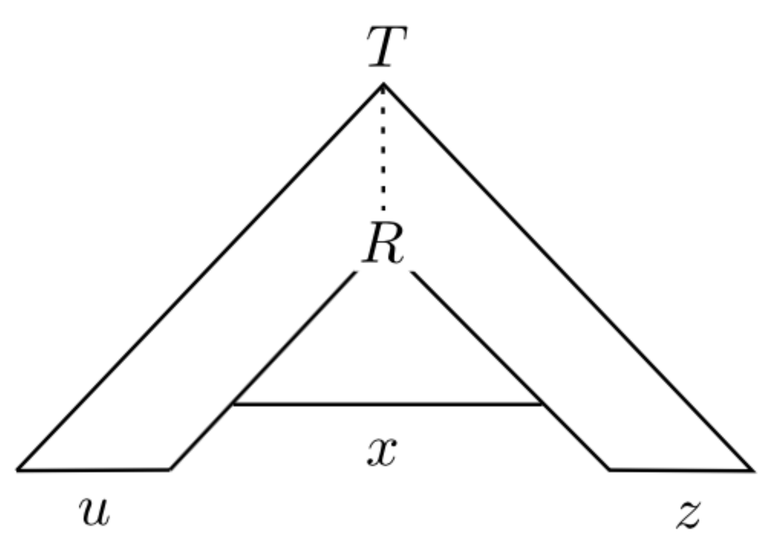


Non-Context-Free Languages

- **Condition 1:** Both v and y should not be ε. If they were both ε then then smaller parse tree generating uxz generates w but this violates our assumption that we started with the smallest parse tree.

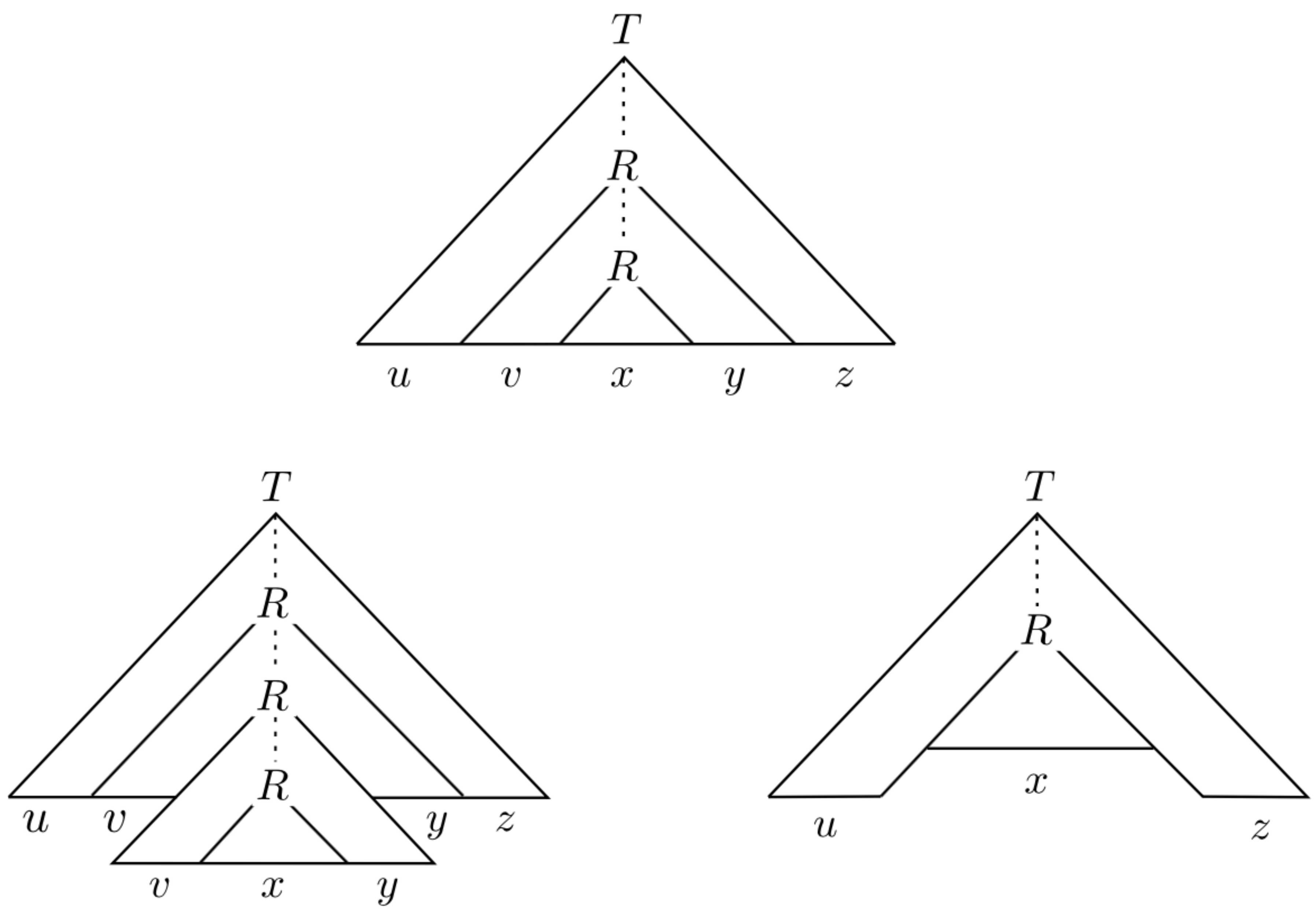






Non-Context-Free Languages

- **Condition 2:** $|vxy| \leq p$**:** R is chosen to be among the bottom $|V| + 1$ variables and is the longest path in the parse tree, then the subtree vxy is at most $|V| + 1$ high and thus $|vxy| \leq 2^{|V|+1} = p$



Using the Pumping Lemma

- **Problem.** Apply the pumping lemma to prove that the language $\{a^n b^n c^n \mid n \geq 0\}$ is not context-free.
- Proof. Assume L is context-free with pumping length p.
- Select $w = a^p b^p c^p \in L$ and has length $3p \geq p$
- Consider all possible ways to partition w into $uvxyz$ s.t. condition (2) and (3) hold: $|vy| > 0$ and $|vxy| \leq p$
    - Notice that vxy cannot be made up of all three letters (why?)

Using the Pumping Lemma

- **Problem.** Apply the pumping lemma to prove that the language $\{a^n b^n c^n \mid n \geq 0\}$ is not context-free.
- Proof. Assume L is context-free with pumping length p.
- Select $w = a^p b^p c^p \ \in L$ and has length $3p \geq p$
- Consider all possible ways to partition w into $uvxyz$ s.t. condition (2) and (3) hold: $|vy| > 0$ and $|vxy| \leq p$
  - **Case 1.** At least one of v or y contains two distinct symbols. Then xv^2xy^2z contains symbols out of order and $\notin L$
  - **Case 2.** Both v and y contain the same symbol (both are $a's$ or both $b's$ or both $c's$ then $uxz \notin B$

Pumping Lemma Questions

- **Question**. What does it mean for a L to satisfy the pumping lemma?
- **Question**. What does it mean to show that L does not satisfy PL?
- **Question**. If a language satisfies PL for CFLs, does it mean it is context-free?
- **Question**. If a language is context-free, does it have to satisfy PL?

Why context-free?

- **Question.** What is the meaning of being "context-free"?
- In CFGs, left-hand side of rules can only contain a single variable say T (no context around when to replace T in a derivation)
- "Context-sensitive" grammars are more general
- A context-sensitive grammar for $\{a^n b^n c^n \mid n > 0\}$:

$$S \to abc \mid aBSc$$

$$Ba \to aB$$

$$Bb \to bb$$

$$S \to aBSc \to aBaBScc \to aaBScc \to aabbcc$$

$$S \to aBSc \to aBaBScc \to aaBScc \to aaBabccc \to aaaBbccc \to aaabbbccc$$