Intro to Theory of Computation

CS 332

LECTURE 7

Last time:

- Pushdown automata (PDAs)
- Context-free grammars (CFG)

Today:

- Pumping lemma for CFLs
- Proving that a language is not CF

Sofya Raskhodnikova

GIVE A CFG FOR…

L_3 = { strings of balanced parens }

$L_4 = \{ a^i b^j c^k \mid i, j, k \geq 0 \text{ and } (i = j \text{ or } j = k) \}$

CFGs in the real world

The syntactic grammar for the Java programming language

BasicForStatement:
for *(; ;) Statement*
for *(; ; ForUpdate) Statement*
for *(; Expression ;) Statement*
for *(; Expression ; ForUpdate) Statement*
for *(ForInit ; ;) Statement*
for *(ForInit ; ; ForUpdate) Statement*
for *(ForInit ; Expression ;) Statement*
for *(ForInit ; Expression ; ForUpdate) Statement*

COMPILER MODULES

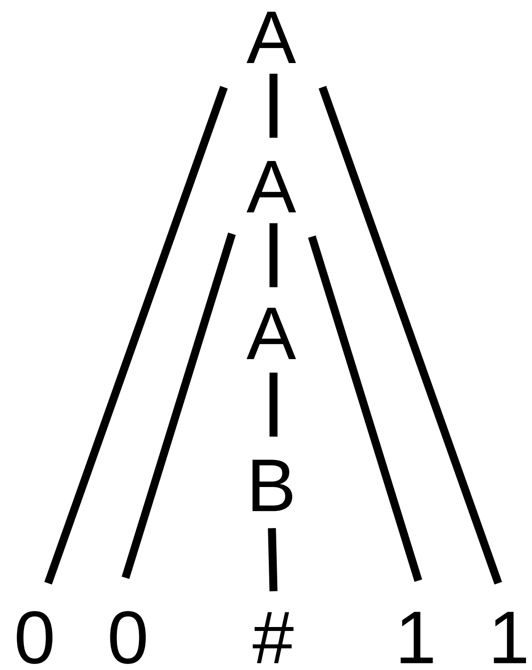
LEXER

PARSER

SEMANTIC ANALYZER

TRANSLATOR/INTERPRETER

Parse trees



$$A \Rightarrow 0A1 \Rightarrow 00A11 \Rightarrow 00B11 \Rightarrow 00\#11$$

Equivalence of CFGs & PDAs

A language is generated by a CFG

$\Leftrightarrow$

It is recognized by a PDA

Context-free or not?

NOT $L_1 = \{ xy \mid x,y \in \{0,1\}^* \text{ and } x=y\}$

YES $L_2 = \{xy \mid x,y \in \{0,1\}^*, |x|=|y| \text{ and } x \neq y\}$

CS 332

Pumping lemma for CFLs

Let L be a context-free language

Then there exists P such that
for every $w \in L$ with $|w| \geq P$

there exist uvxyz=w, where:

1. $|vy| > 0$
2. $|vxy| \leq P$
3. $uv^ixy^iz \in L$ for all $i \geq 0$

CS 332

Examples

there exist uvxyz=w, where:

1. $|vy| \geq 1$

2. $|vxy| \leq P$

3. $uv^ixy^iz \in L$ for all $i \geq 0$

Example: L = { w ∈ {0,1}* | w = w^R}.

w = 0; u,v,x,y,z = ?

w = 010; u,v,x,y,z=?

Example: L = { w ∈ {a,b}* | #a > #b in w}.

w = a; u,v,x,y,z = ?

w = aab; u,v,x,y,z=?

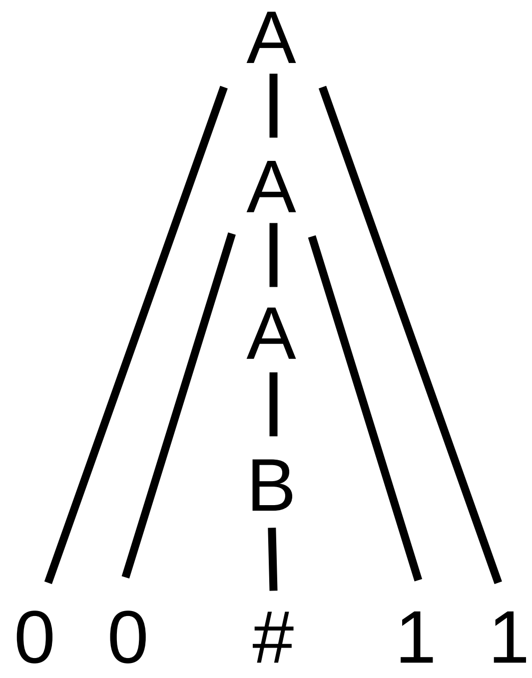
CS 332

Exercise

Example: L = { w ∈ {a,b}* | #a > #b in w}.
w = aab; u,v,x,y,z=?

A. $v = a, x = a, y = b$
B. $v = aa, x = \varepsilon, y = b$
C. $v = aa, x = \varepsilon, y = \varepsilon$
D. More than one choice above works.
E. None of the choices above work.

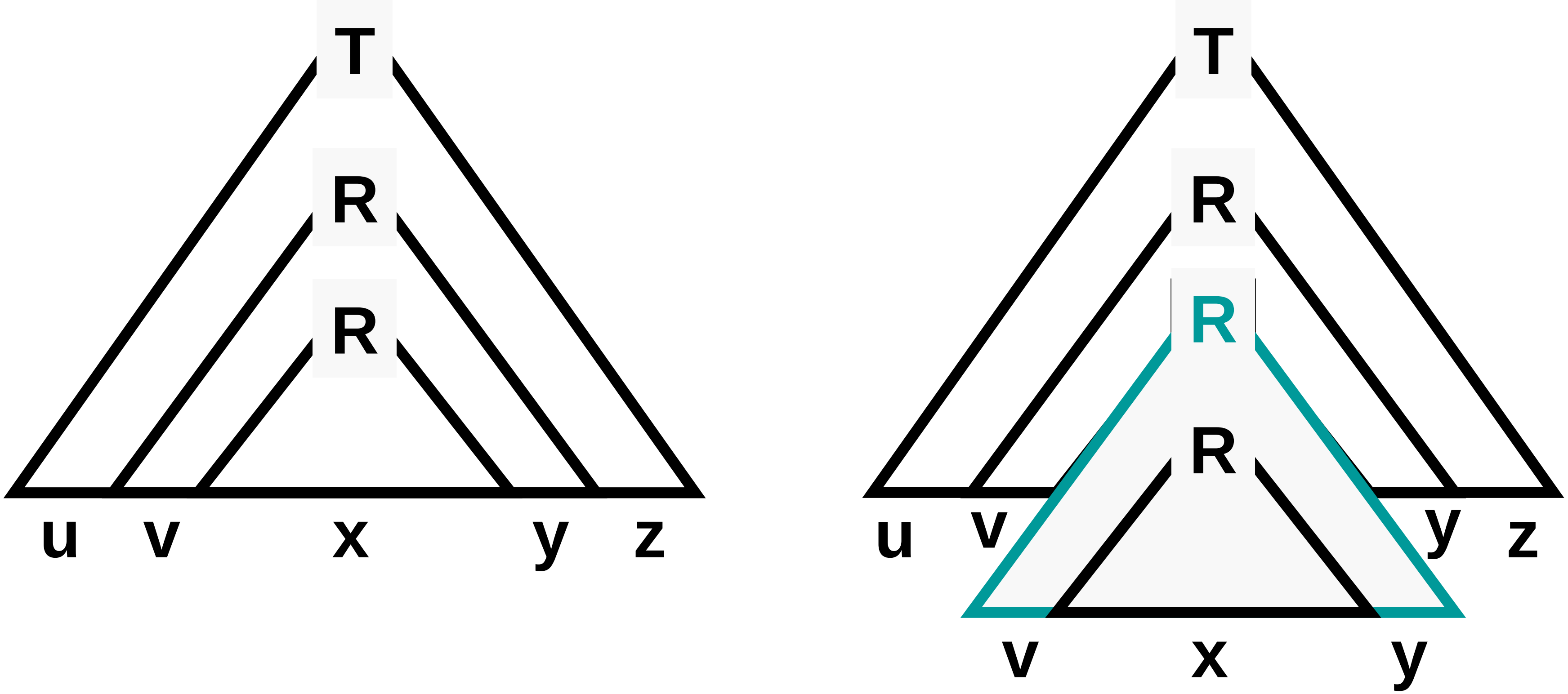
Recall: parse trees



$$A \Rightarrow 0A1 \Rightarrow 00A11 \Rightarrow 00B11 \Rightarrow 00\#11$$

Pumping lemma: proof idea

If string w is long enough, then every parse tree for w must have a path that contains a variable more than once.



Pumping lemma: proof

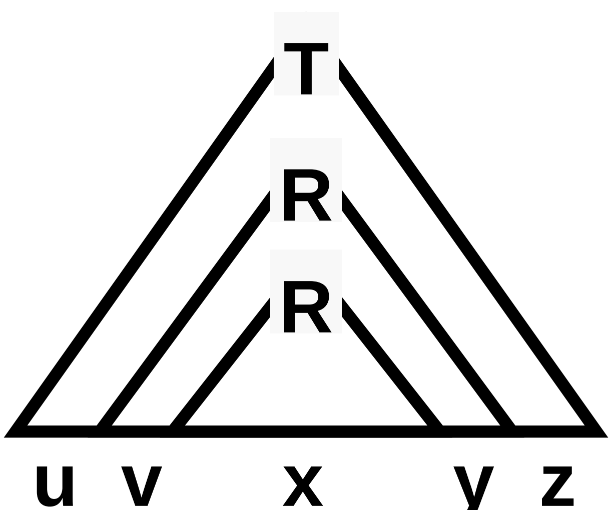
- Let b be the maximum number of symbols on the right-hand side of a rule.
- If the height of a parse tree is h, the length of the string generated is at most: b^h
- Let |V| be the number of variables in G.
- Define $p = b^{|V|+1}$.
- Let w be a string of length at least p.
- Let T be the parse tree for w with the smallest number of nodes.
- T must have height at least |V|+1.

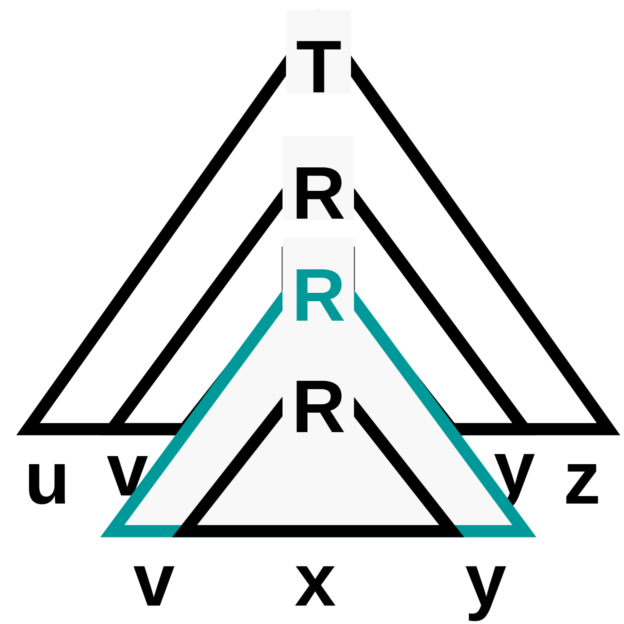
Pumping lemma: proof

The longest path in T must have ≥ |V|+1 variables

Select R to be the variable that repeats among the lowest |V|+1 variables

1. **$|vy| > 0$**
2. **$|vxy| \leq P$**





CS 332

Using the pumping lemma

Prove $L = \{ww \mid w \in \{0,1\}^*\}$ is not context-free

Assume for contradiction that L is context-free.

Then there is a pumping length P.

CS 332

Pumping lemma as a game

1. **YOU pick the language *L* to be proved not CFL.**
2. **ADVERSARY picks p, but doesn't reveal to YOU what p is; YOU must devise a play for all possible p's.**
3. **YOU pick w ∈ L, which should depend on p and which must be of length at least p.**
4. **ADVERSARY divides w into u, v, x, y, z, obeying PL conditions: vy is not empty and vxy has length $\leq p$. Again, ADVERSARY does not tell YOU what u, v, x, y, z are.**
5. **YOU win by picking i, which may be a function of p, u, v, x, y, z, such that uv^ixy^iz is not in L.**

Exercise

Prove {ww | w ∈ {0,1}*} is not context free.

Proof: Assume … pumping length p.

What string can you choose in your next move?

A. 0101

B. 0^p1^p

C. $(01)^p$

D. $(01)^{2p}$

E. $0^p1^p0^p1^p$

F. More than one choice above works.

Using the pumping lemma

Prove L = {ww | w ∈ {0,1}*} is not context-free

Assume for contradiction L is context-free.

Then there is a pumping length P.

No matter what P is, the string s = $0^P1^P0^P1^P$ has $|s| \geq P$ and s ∈ L.

So there should be uvxyz=s with:

1) |vy| > 0, 2) |vxy| ≤ P, 3) ∀ i, uv^ixy^iz ∈ L.

P P P P

s = 00...0011...1100...0011...11

{ww | w∈{0,1}*} is not a CFL: proof

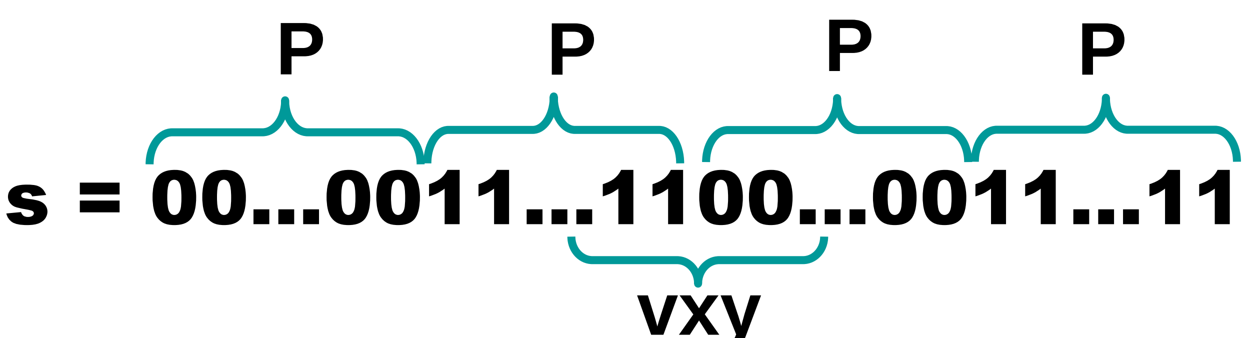
No matter what P is, the string $s = 0^P1^P0^P1^P$ has $|s| \geq P$ and $s \in L$

Then pumping down must remove at least one 1 or one 0, e.g. $uxz = 0^P1^i0^j1^P$ where $i < P$ or $j < P$

vxy cannot be only in the first half since pumping down would move a 0 to the end of the first half.

vxy cannot be only in the last half since pumping up would move a 0 to the end of the first half

Both... So there's no way to split s = uvxyz with 3) $|vy| > 0$, 2) $|vxy| \leq P$, 1) $\forall i \geq 0$, $uv^ixy^iz \in L$, however they are chosen, so L cannot be context-free!



Exercise

To prove that a language L is not context free, we can

A. argue that a PDA cannot remember enough information to recognize L;

B. use pumping lemma;

C. give a CFG and show that it does not generate L.

D. None of the above.

E. More than one choice above works.

CS 332

Using the pumping lemma

Prove L = $\{w\#w^R \mid w \in \{0,1\}^* \text{ and } \#1s = \#0s\}$ is not context-free

Assume L is context-free.

Then there is a pumping length P.

No matter what P is, the string $s = 0^P1^P\#1^P0^P$ has $|s| \geq P$ and $s \in L$.

So there should be uvxyz=s with:

1) $|vy| > 0$, 2) $|vxy| \leq P$, 3) $\forall i$, $uv^ixy^iz \in L$.

$$s = \overbrace{00\ldots00}^{P}\overbrace{11\ldots11}^{P}\#\overbrace{11\ldots11}^{P}\overbrace{00\ldots00}^{P}$$

$\{w\#w^R \mid w \in \{0,1\}^* \text{ and } \#1s = \#0s\}$ is not a CFL: proof

Assume L is context-free.

Then there is a pumping length P.

No matter what P is, the string $s = 0^P1^P\#1^P0^P$ has $|s| \geq P$ and $s \in L$.

So there should be uvxyz=s with:

1) $|vy| > 0$, 2) $|vxy| \leq P$, 3) $\forall$ i, $uv^ixy^iz \in L$.

vxy cannot be only in either half, since pumping would make one side of # longer than the other.

 # cannot be in v or y since pumping would add too many #s. So it must be in x.

Since # is in x and $|vxy| \leq P$, neither v nor y can have 0s, and at least one must have a 1.

Then uv^2xy^2z has more 1s than 0s and is not in L.

P P P P

s = 00...0011...11#11...1100...00

vxy

Context-free or not?

NOT $L_1 = \{ xy \mid x,y \in \{0,1\}^* \text{ and } x=y\}$

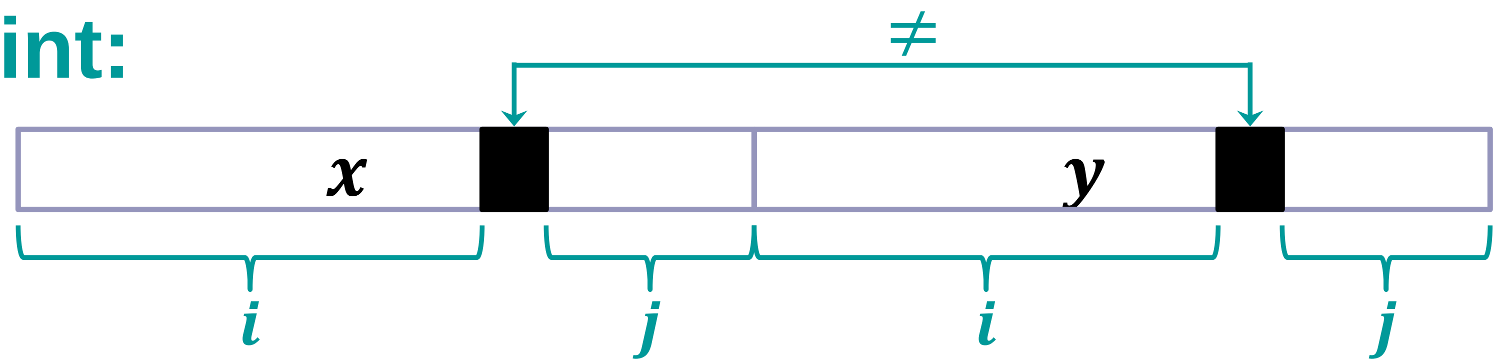
YES $L_2 = \{xy \mid x,y \in \{0,1\}^*, |x|=|y| \text{ and } x \neq y\}$

CS 332

Give an algorithmic description of a PDA for

$\{xy \mid x, y \in \{0,1\}^*$ **and** $|x| = |y|$ **but** $x \neq y\}$.

Hint:



Give an algorithmic description of a PDA for

$\{xy \mid x, y \in \{0,1\}^*$ **and** $|x| = |y|$ **but** $x \neq y\}$.

Hint: