CSE 396 Lecture Thu 3/31 Spring 2016 ①

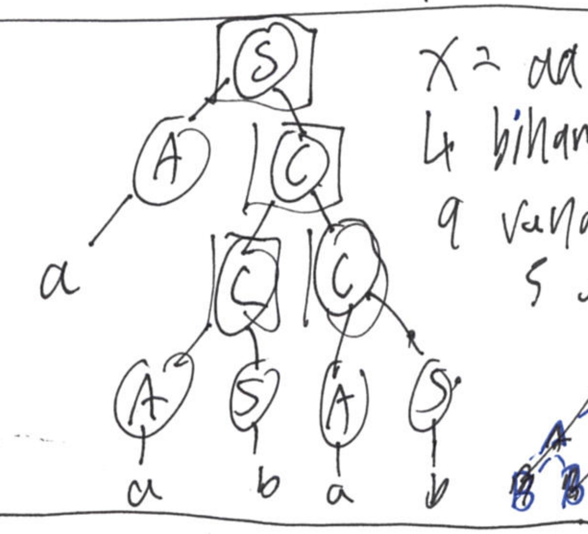
<u>Useful fact</u>: If G is in ChNF, then all parse trees in G are <u>binary</u>, and if $x \in \Sigma^n$ is in $L(G)$, then x is derivable in
- $n-1$ variable steps $A \to BC$
- n terminal steps $A \to c$ (or $S \to \epsilon$)

∴ Parse trees have $\underline{n-1}$ binary internal nodes. ($2n-1$ variable nodes)

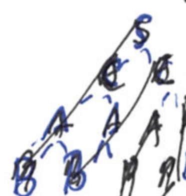
$$\boxed{\text{Some path from root to a terminal has length} \geq \log_2 n}$$

$S \to AC \mid b$
$A \to BS \mid a$
$B \to SC \mid AB$
$C \to AS \mid CC$

TEXT "b" is fixed to b = 2 thanks to ChNF



$x = aabab$
4 binary nodes
9 variable nodes,
5 unary.



Biggest tree with 4 vars and no repeats has 15 nodes

<u>Theorem</u>: If $n \geq 2^k$, then some path has $\geq k+1$ variable nodes.

∴ If k is the # of variables in the grammar, then some path must repeat a variable [including the non-binary variable at the end, which gives "slack" to the Thm]

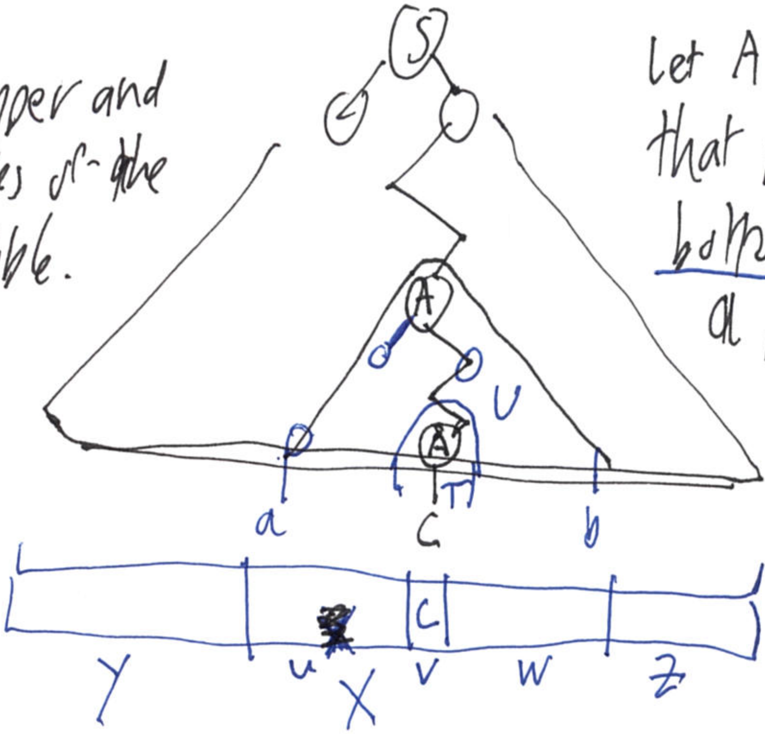
If $k = 4$ and $n \geq 16$ quite clearly some variable must repeat along some path.

Let us focus on the <u>lowest</u> two occurrences of a repeated variable along a path from the root to a terminal.

Focus on the upper and lower occurrences of the repeated variable.

Let U be the upper subtree

Let T be the lower subtree



Let A be a variable that repeats among the bottom $k+1$ levels of a parse tree for some $x \in L(G)$, $n =|x| \geq 2^k$

$|uvw| \leq 2^k$ because we took the lowest repeat.

Then x breaks as $x =: yuvwz$ such that $\begin{cases} v \text{ is the yield of } T \\ uvw \text{ is the yield of } U \\ yuvwz = x \end{cases}$

TEXT: S breaks as $S =: uvxyz$

Observe: • If we make the upper A do T instead of U, then u and w would become empty. The result is a parse tree for $x^{(0)} =_{def} yvz$.

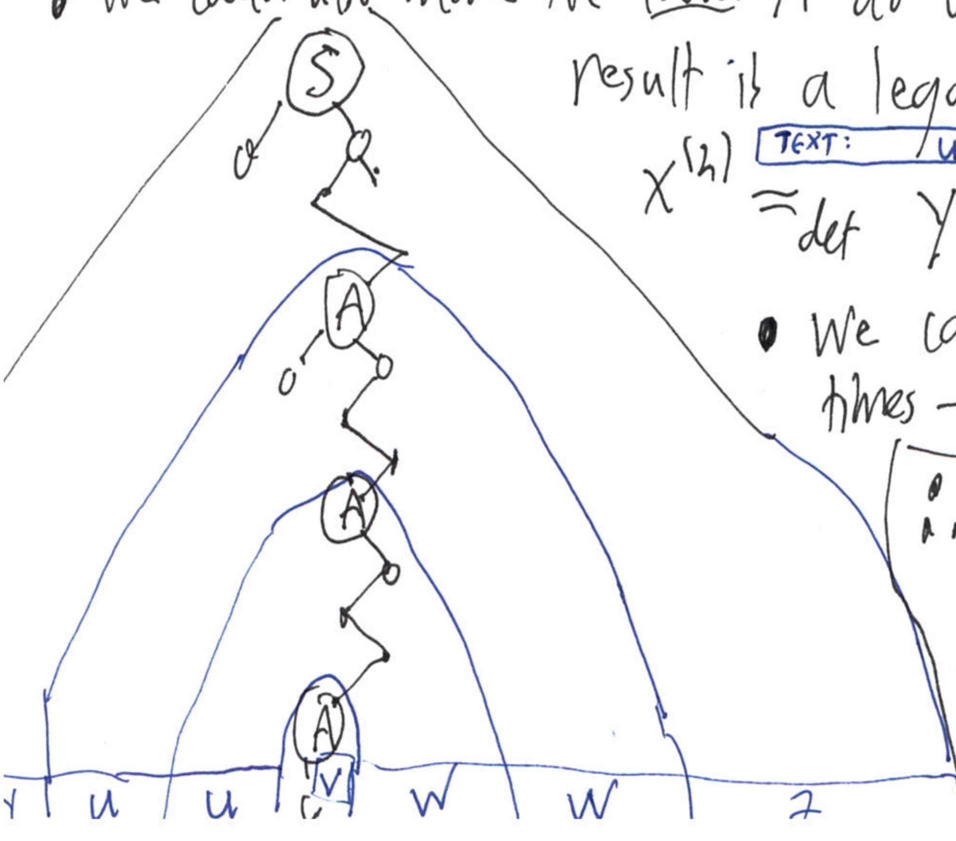
$\therefore x^{(0)} \in L(G)$

Here v is just a ~~char~~ c char

• We could also make the lower A do U instead of T. The result is a legal parse tree for the string

TEXT: $u\,v\,v\,x\,y\,y\,z$

$x^{(2)} =_{def} yuuvwwz$.



• We could expand U 3 or more times — say i times total.

$\therefore$ For each $i \geq 0$, the string $x^{(i)} =_{def} yu^ivw^iz$ belongs to $L(G)$

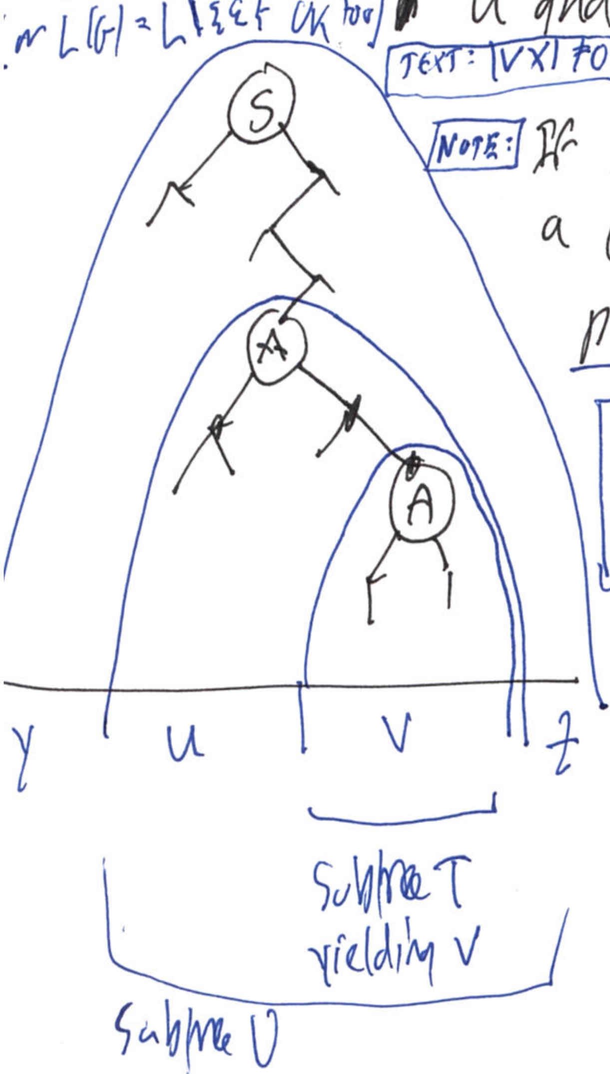
TEXT: SEL 3

∴ There exists p, $p \approx 2^{|V|}$, such that any given $x \in L$, $|x| \geq p$, can be broken as $x =: yuvwz$ such that

$|s| \geq p$ [TEXT: $|vxy| \leq p$] [TEXT: $s =: uv\cancel{w}yz$] [TEXT! $uv^ixy^iz \in L(G)$]

Recall: Every CFL L has a grammar G in ChNF st $L(G) = L$ (or $L(G) = L \setminus \{\epsilon\}$ OK too)

- $|uvw| \leq p$
- for all i, $yu^ivw^iz \in L(G)$ and
- u and w are not both ϵ. [TEXT: $|vx| \neq 0$]

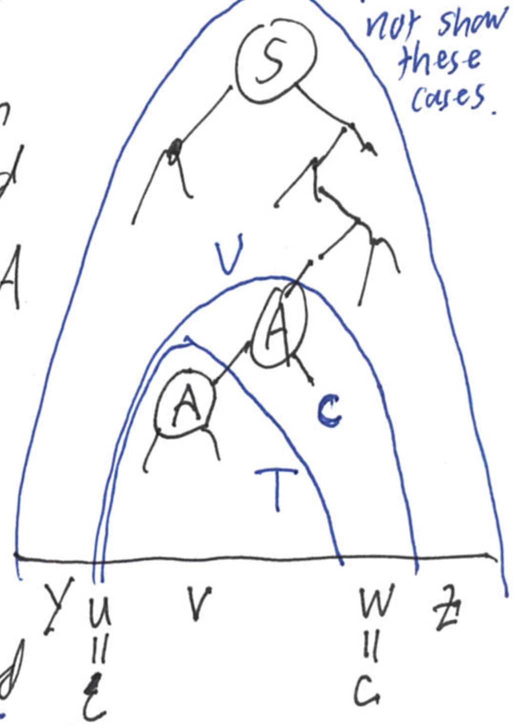


NOTE: If the repeated A is on a completely rightward path from the upper A

$\boxed{\text{Then } w = \epsilon \text{ but } u \neq \epsilon}$

If it is on a completely leftward path, eg a left child of the upper A, then $u = \epsilon$, $w \neq \epsilon$.

TEXT does not show these cases.



Here w must at least be a single char.

[The "pumping length" p can be $\geq 2^{|V|}$, where $|V|$ is the # of vars in a ChNF G for L

∴ The CFL Pumping Lemma: For any CFL L, there exists $p > 0$ such that for all $x \in L(G)$ with $|x| \geq p$, there exists a breakdown $x =: yuvwz$ with $|uvw| \leq p$, $uw \neq \epsilon$, such that for all $i \geq 0$, $x^{(i)} =_{def} yu^ivw^iz$ belongs to L.

[T: $s \in L(G)$, $|s| \geq p$, $s =: uvxyz$, $|vxy| \leq p$, $vy \neq \epsilon$]

i.e. If L is a CFL then [— Blah —]

Contrapositive: If $\neg$ Blah, then L is not a CFL.

$$\text{Blah} \equiv (\exists p > 0)(\forall x \in L(G), |x| \geq p)\underset{x =: yuvwz,\ |uvw| \leq p,\ uw \neq \varepsilon}{[\exists \text{ breakdown}]}(\forall i)\underset{x^{(i)} \in L}{[\text{Body}]}$$

$$\neg\text{Blah} \equiv (\forall p > 0)(\exists x \in L(G), |x| \geq p)\underset{x =: yuvwz \text{ st. } |uvw| \leq p \wedge uw \neq \varepsilon}{(\forall \text{ breakdowns})}(\exists i)\underset{x^{(i)} \notin L}{[\neg\text{Body}]}$$

$\therefore$ The CFL Pumping Lemma Contra:

Given any language L over an alphabet Σ, if

for all $p > 0$ there exists an $x \in L(G)$, $|x| \geq p$ (TEXT: $s \in L(G)$, $|s| \geq p$) such that

for each breakdown $x =: yuvwz$ st. $|uvw| \leq p \wedge uw \neq \varepsilon$ (Text: $s =: uvxyz$ and $|vxy| \leq p \wedge |vy| \neq 0$)

there exists $i \geq 0$ st. $x^{(i)} = yu^ivw^iz \notin L$. (TEXT: $uv^ixy^iz \notin L$)

then L is not a CFL. Proof Script for applying it:

Let any $p > 0$ be given. Take $x =$ ______ "Clearly $x \in L$."

very often $|x| \approx 3p$ or $4p$ or etc.

Let any breakdown $x =: yuvwz$ st. $|uvw| \leq p$, $uw \neq \varepsilon$, be given

Take $i =$ ______ (usually 0 or 2). Then $x^{(i)} =$ ______

which is not in L because ______. By CFLPL, L is not a CFL.

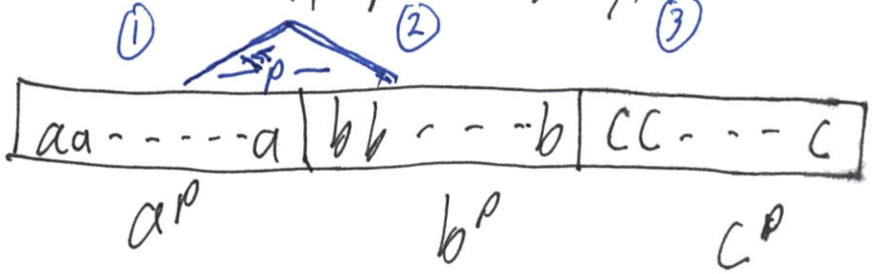
$\boxtimes$.

(2.36) 5

Example: $L = \{a^n b^n c^n : n \geq 1\}$. $\Sigma = \{a, b, c\}$.

Prove that L is not a CFL: Let any $p > 0$ be given.

Take $x = a^p b^p c^p$. Clearly $x \in L$.

$x =$ 

Let any breakdown $x =: yuvwz$ with $|uvw| \leq p$, $uw \neq \varepsilon$, be given.

<u>Take $i = 0$</u>. By $uw \neq \varepsilon$ this destroys at least one a, b, or c.

But by $|uvw| \leq p$, it can't destroy both an a or a c.

$\therefore$ In $x^{(0)}$, the a's b's and c's can't all be in balance.

$\therefore x^{(0)} \notin L$. <u>$\therefore L$ is not a CFL</u>.

<u>Added</u>: For next week, lec+rec together will cover at least one, intendedly both, of

<u>Example 2.37</u> (varied a bit) $L = \{a^i b^j c^k : i < j < k\}$ is not a CFL. Given p, take call it s or $x = \underline{a^p b^{p+1} c^{p+2}}$. Then $x \in L$. Let any breakdown...

<u>Example 2.38</u> done this way. Same proof works for $L = \{ww : w \in \{a,b\}^*\}$

Let: $L_1 = \{a^m b^n a^m b^n : m, n \geq 0\}$
$L_2 = \{a^m b^n a^n b^m : m, n \geq 0\}$

Then one of L_1, L_2 is a CFL and the other isn't. Which is which? <u>Answer</u>:

L_2 has "nicely nested" dependencies. Grammar: $S \to aSb \mid T$, $T \to bTa \mid \varepsilon$.

L_1 has "crossing dependencies". Given p, take $x = a^p b^p a^p b^p$ (regions ① ② ③ ④). Let $x =: yuvwz$ with $|uvw| \leq p$, $uw \neq \varepsilon$. The "$-p-$" hits at least one of regions ①, ②, ③, ④, but is too narrow to keep its other odd or even counterpart region balanced with it. So $x^{(0)} \notin L$.