Useful fact: If $G$ is in ChNF, then all parse trees in $G$ are binary, and if $X \in \Sigma^*$ is in $L(G)$, then $X$ is derivable in $n-1$ variable steps $A \to BC$.

2n-1 variable nodes

$n-1$ terminal steps $A \to C$

So parse trees have $n-1$ internal nodes.

Some path from root to a terminal has length $\geq \log_2 n$

\[
S \to AC1b \\
A \to BS1a \\
B \to SC1AB \\
C \to AS1CC
\]

Test "b" is fixed to $b = 2$. Thanks to ChNF.

\[
\begin{array}{c}
\text{X = dabab} \\
4 \text{ binary nodes} \\
9 \text{ variable nodes, 5 unary.} \\
\end{array}
\]

Biggest tree with 4 vars and no repeats has 15 nodes.

Theorem: If $N \geq 2^K$, then some path has $2K+1$ 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 tree).

If $N \geq 2^K$ quite clearly some variable must repeat along some path.

Let us focus on the lowest two occurrences of a repeated variable along a path from the root to a terminal.
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 largest repeat.

Then $X$ breaks as $X = yuvwz$ such that $V$ is the yield of $T$.

Observe: If we made the upper $A$ do $T$ instead of $V$, then $u$ and $w$ would become empty.

The result is a parse tree for $X^{(1)} = \text{det } YV_2$.

We could also make the lower $A$ do $V$ instead of $T$. The result is a legal parse tree for the string $X^{(2)} = \text{det } YUUUVWW_2$.

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

For each $i \geq 0$, the string $X^{(i)} = \text{det } YU_iVWW_2$ belongs to $L(G)$. 

Focus on the upper and lower occurrences of the repeated variable. Let $U$ be the upper subtree. Let $T$ be the lower subtree.
Theorem: 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 $|uv| \leq p$, $uw \neq \epsilon$, such that for all $i \geq 0$, $x(i) = x(yu^iuvwz)$ belongs to $L$.
If \( L \) is a CFL then \([-Blah-]\)

Contrapositive: If \([-Blah-]\), then \( L \) is not a CFL.
\[
\neg Blah \equiv (\exists p > 0)(\forall x \in L(p), x \xi p) \text{ [breakdown]} (\forall i)(\neg \text{Body})
\]
\[
\neg Blah \equiv (\forall p > 0)(\exists x \in L(p), x \xi p) \text{ [breakdown]} (\exists i)(\neg \text{Body})
\]

The CFL Pumping Lemma can be:

Given any language \( L \) over an alphabet \( \Sigma \), if

for all \( p > 0 \) there exists an \( x \in L(p), x \xi p \) such that

for each breakdown \( x \equiv \gamma \nu \nu \nu \xi \lambda, |\nu\nu| \leq p \cap \nu \lambda \neq 0 \)

then there exists \( i \geq 0 \) \( \lambda \equiv \gamma \nu \nu \nu \xi \lambda \in L \).

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

Let any \( p > 0 \) be given. Take \( x = \)

Very often \( 1x1 \xi 3p \) or \( 4p \) or etc.

Let any breakdown \( x = \gamma \nu \nu \nu \xi \lambda, |\nu\nu| \leq p \cap \nu \lambda \neq 0 \)

usually 0 or 2

Take \( i = \)

Then \( x^{(i)} = \)

which is not in \( L \) because \( \).

By CFLPL, \( L \) is not a CFL.
Example: \( L = \{ a^m b^n c^n : m, n \geq 1 \} \). \( \Sigma = \{ a, b, c \} \).

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

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

Let any breakdown \( x = uvwxyz \) with \( |uvw| \leq p \), \( uv \neq \epsilon \) be given.

Take \( i = 0 \). By \( uv \neq \epsilon \) this destroys at least one of \( a, b, c \).

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

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

\( \therefore x^{(0)} \notin L \). \( \therefore L \) is not a CFL.

---

Example 2.37 \( L = \{ a^i b^j c^k \} : 1 \leq j \leq k \} \) is not a CFL. Given \( p \), take

(mentioned a bit)
call it \( S \) or \( x = a^p b^{p+1} c^{p+2} \). Then \( x \in L \). Let any breakdown.

Example 2.38 Let \( L_1 = \{ a^m b^n a^m b^n : m, n \geq 0 \} \) one this way. Let \( L_2 = \{ a^m b^n a^m b^m : m, n \geq 0 \} \)

Then one of \( L_1, L_2 \) is a CFL and the other isn't. Which is which? Answer:

2, has "nicely nested" dependencies. Grammar: \( S \rightarrow aSb | T \), \( T \rightarrow bTa | \epsilon \).

1, has "crossing dependencies." Given \( \rho \), take \( x = a^p b^p a^p b^p \). Let \( x = uvwxyz \) with \( |uvw| \leq p \), \( uv \neq \epsilon \). The "\( \rho \)" hits at least one of regions 1, 2, 3, 4, one of regions 1, 2, 3, 4, but is now narrow to keep its other odd or even counterpart region balanced with it. So \( x^{(0)} \notin L_2 \).