Theorem: Given any CFG $G_0 = (V, \Sigma, R_0, S)$ we can build a grammar $G'$ in Chomsky NF such that $L(G') = L(G_0) \setminus \delta e^*$. Proof: Skipped - part covered in Ch 4. Will be.

CFG Pumping Lemma: Let $L \subseteq \Sigma^*$ be any CFL. Then:

- There exists an $N > 0$ such that for all $w \in L$ with $|w| > N$:
  - There exists a breakdown $x = yuvz$ such that $|vy| \leq N$ and $|vz| > 0$.
  - $yuv^2z \in L$.
  - $uyv^2z \in L$.

Proof: By the Theorem, there is a CFG $G$ in Chomsky NF, $L(G) = L \setminus \delta e^*$. Take $N = 2^{1N}$. Let any $w \in L(G)$ with $|w| > N$, be given. By $w \in L(G)$, we can take a parse tree $T$ for $w$. Since $G$ is in ChNF, $T$ is a binary tree.

Hence some path from $S$ in the tree has a terminal leaf at least $1N + 1$ edges from $S$, thus (more than) $1N + 1$ variables.

By PHP, some variable appears twice if $T$ has $n > N = 2^h$ leaves, then it has height $\geq h + 1$ in $T$. Hence some path from the variable (here, the upper $O$ in red), has height $\leq 1N + 1$ in nodes, which is $\leq 1N$ + 1! Hence the path of the subtree of that variable has length at most $2^{1N} = N$. Look at $T$ now.
We can do two kinds of adjustment to T to derive new strings.

1. Make the lower variable imitate the upper multiple times until finally doing what it did.

2. Make the upper variable imitate what the lower did.

This immediately implies the contrapositive form:

**Original:** If \( L \) is a CFL, then \( \text{blab} \)

**Contradiction:** If \( \neg \text{blab} \), then \( L \) is not a CFL!
Contaposition: Let $L$ be any language. Suppose that for all $N > 0$
there exists $x \in L$ with $|x| > N$ such that for all breakdowns $X = yuuvwz$ subject to $|uvw| \leq N$ and $uv \neq \varepsilon$,
there exists $i > 20$ such that $x^{(i)} = yu^iuv^iz$ is not in $L$.

Then $L$ is not a CFL.

Example: $L = \{anbcn : n \geq 1\}$.

Let any $N > 0$ be given

Take $x = anbncn$.

Then $x \in L$. (clearly)

Let any breakdown $X = yuuvwz$ with $|uvw| \leq N$, $uv \neq \varepsilon$, be given.

Take $i = 0$.

In each case we get $x^{(i)} \notin L$, so $L$ is not a CFL by the CFL PL.

By $|uvw| \leq N$, we cannot have that $u$ and $w$ collectively include both an $a$ and a $c$. But by $uv \neq \varepsilon$, $uw$ must include at least one $a$, $b$, or $c$. Hence $x^{(10)} = yvz$ subtract at least one letter but cannot balance the counts of all three letters. Hence $x^{(i)} \notin L$ with $i = 0$, so $L$ is not a CFL.