Defn: A language $L$ is a Context-Free Language (CFL) if there is a Context-Free Grammar (CFG) $G$ s.t. $L = L(G)$.

Lemma: If $L$ is a CFL, then:

1. There exists a number $N > 0$ such that for all $x \in L$ with $|x| \geq N$:
2. There exists a breakdown $x = y u v w z$ such that $|u v w| \leq N$, $u$ and $w$ are not both $\varepsilon$, and:
3. For all $i \geq 0$, $x^{(i)} = y u^i v w z$ belongs to $L$.

Note: $x^{(0)} = y v z$

Proof: Since $L$ is a CFL, there is a grammar $G$ in Chomsky Normal Form such that $L(G) = L \setminus \varepsilon$. We have $G = (V, \Sigma, R, S)$. Take $k = |V|$, $N = 2^k$.

Let any $x \in L = L(G)$, $|x| > N$ be given.

By $x \in L(G)$, we can take a parse tree $T$ for $x$ in the CFG $G$. By binary counting, $T$ has at least $k+1$ edges in some path from the root $S$ to a leaf variable that derived a terminal in $x$.

Consider the bottom $k+1$ nodes of that path.

By the Pigeonhole Principle, some $B \in \Sigma$ occurs at least twice in those bottom $k+1$ nodes.
Let only path case.

Take $V$ to be the yield of the lower $B$.

Take $U$ to be everything to the left of $V$ that is derived by the upper $B$.

Take $W$ to be everything to the right of $V$ that is derived by the upper $B$.

Take $Y$ to be the rest of $X$ to the left of $U$, $Z$ the rest of $X$ to right of $W$.

Let any $i \geq 0$ be given. Then $x^{(i)} = y u v w z \in L(G)$ because we can splice or extend the parse tree between the two $B$'s. This results in a legal parse tree for $x^{(i)}$, so $x^{(i)} \in L(G) = L$.

Contrapositive: Suppose $L$ is any language such that

* for all $N > 0$
  * there exists an $x \in L$ with $|x| > N$ such that
    * for all productions $x = y u v w z \in L$ and $|u v w| \leq N$
      * there exists $i \geq 0$ s.t. $x^{(i)} = y u v w z$ is not in $L$.

Then $L$ is not a CFL.