The Pumping Lemma for Context-Free Languages

The pumping lemma for CFLs is stated in the text as follows: (Note that I like my naturals with a zero: $\mathbb{N} = \{0, 1, 2, ...\}$.) "If L is a CFL, then

$$\big(\exists n \in \mathbb{N}\big)\big(\forall z \in \Sigma^* \,\big|\, |z| \geq n\big)\big(\exists uvwxy \in \Sigma^* \,\big|\, uvwxy = z \wedge |vwx| \leq n \wedge vx \neq \epsilon\big)\big(\forall i \in \mathbb{N}\,(uv^iwx^iy \in L)\big).$$"

Now that's all well and good, but it's not what we use to prove that languages aren't regular. What we need is its contrapositive, i.e. "If

$$\neg\big(\exists n \in \mathbb{N}\big)\big(\forall z \in \Sigma^* \,\big|\, |z| \geq n\big)\big(\exists uvwxy \in \Sigma^* \,\big|\, uvwxy = z \wedge |vwx| \leq n \wedge vx \neq \epsilon\big)\big(\forall i \in \mathbb{N}\,(uv^iwx^iy \in L)\big),$$

then $\neg(L$ is a CFL)." The negation causes all the quantifiers to 'flip':

$$\big(\forall n \in \mathbb{N}\big)\big(\exists z \in \Sigma^* \,\big|\, |z| \geq n\big)\big(\forall uvwxy \in \Sigma^* \,\big|\, uvwxy = z \wedge |vwx| \leq n \wedge vx \neq \epsilon\big)\big(\exists i \in \mathbb{N}\,(uv^iwx^iy \notin L)\big).$$

Let's break this up to make it easier to swallow:

A language L is not a CFL if

$\forall n \in \mathbb{N}$	(1)
$\exists z \in \Sigma^*$ such that $\|z\| \geq n$	(2)
$\forall uvwxy \in \Sigma^*$ where $uvwxy = z \wedge \|vwx\| \leq n \wedge vx \neq \epsilon$	(3)
$\exists i \in \mathbb{N}\,(uv^iwx^iy \notin L)$.	(4)

Note that this is an "if" statement, and not an "if-and-only-if" — so you can't use the pumping lemma to prove that languages **are** context-free. Here are some rules to help understand how to use this:

- (1) tells you to treat n as a variable that can be any natural number.
- By (2), you **choose** a particular z based on n, but make sure that $|z| \geq n$.
- (3) says that $uvwxy$ represents any possible way to split up z subject to the constraints.
- Now you have to choose an i to "get out of L" if you can by (4), and we're done.

The most common mistakes on Test 2 were these: you have to exlicitly choose z based on n, and you need to deal with **all** the possible cases of $uvwxy$ in some suitable way.