CS 462: Winter 2018
The unary CFL proof, revised

Theorem. If $L \subseteq a^*$ is a CFL, then it is regular.

Proof. Let n be the constant for L in the pumping lemma for CFL's. The idea is to write L as union of (a) a finite set and (b) at most $n(n+1)/2$ regular languages of the form $a^i(a^j)^*$. Since the regular languages are closed under union, the result will follow.

We start by writing $L = S \cup T$, where $S = L \cap \{\epsilon, a\}^{n-1}$ and $T = L \cap a^n a^*$. Here S consists of those words in L of length $< n$, while T consists of those words in L of length $\geq n$. Now S is finite, and hence regular, so it suffices to show that T is regular.

We can now apply the pumping lemma for CFL's to every word in T. If $a^r \in T$, this gives an expression $a^r = uvwxy$ with $|vwx| \leq n$ and $k := |vx| \geq 1$. The pumping lemma then states that $a^{r+ik} \in L$ for all $i \geq 0$. In other words, the pumping lemma states that $a^r(a^k)^* \subseteq L$.

Now notice that if $r \equiv s \pmod{k}$, and $s \geq r$, then $a^s(a^k)^* \subseteq a^r(a^k)^*$. So if we are taking the union of many different languages of the form $a^r(a^k)^*$, the only important things to know are (a) the residue class of $r \pmod{k}$, and (b) the smallest r in that residue class.

Let's start by writing down all the possible residue classes. There are $n(n+1)/2$ of them: one for each pair (j, k) with $1 \leq k \leq n$ and $0 \leq j < k$. Not all of these, however, will correspond to pumping lemma decompositions, so let's write down the ones that do:

Define

$$M = \{(j,k)\ :\ \exists\ r \equiv j \pmod{k} \text{ such that } a^r \in T \text{ and there is}$$
$$\text{a pumping lemma decomposition } a^r = uvwxy \text{ with } |vx| = k\ \}.$$

Now that we have the possible residue classes, we can find the length of the shortest word in T with a decomposition corresponding to each residue class. We define a function $f : M \to \mathbb{N}$ as follows:

$$f(j,k) = \min\{r\ :\ a^r \in T \text{ and } r \equiv j \pmod{k} \text{ and there is}$$
$$\text{a pumping lemma decomposition } a^r = uvwxy \text{ with } |vx| = k\ \}.$$

Then we claim that $T = T'$, where

$$T' = \bigcup_{(j,k)\in M} a^{f(j,k)}(a^k)^*.$$

Proof:

$T \subseteq T'$: if $a^i \in T$, then $i \geq n$ and there is a pumping lemma decomposition $a^i = uvwxy$ with $|vx| = k \leq n$. Then $i \equiv j \pmod{k}$ for some j, and so $(j,k) \in M$ and $i \geq f(j,k)$. So $a^i \in a^{f(j,k)}(a^k)^*$.

$T' \subseteq T$: if $a^s \in T'$, then $s = f(j,k) + ik$ for some $i \geq 0$. Then, letting $r = f(j,k)$, we know that $a^r \in T$ and the pumping lemma applied to a^r gives a decomposition $uvwxy$ with $|vx| = k$. So $uv^{i+1}wx^{i+1}y = a^{r+ik} \in T$ for all $i \geq 0$ and hence $a^s \in T$.

Now observe that T' is the finite union of regular languages and hence regular. Hence T is regular. Hence L is regular.

This completes the proof.