Properties of Context-free Languages

We give the Pumping Lemma.

Proving Languages Not Context-Free

Some languages cannot be recognized by PDAs. But to prove this we need the Pumping Lemma.

> **Pumping Lemma.** *Let A be a context-free language. Then there is a constant k such that, for every string $z \in A$ of length at least k, one can split z as $uvwxy$ where:*
>
> - *vx is nonempty,*
> - *$|vwx| \leq k$, and*
> - *uv^iwx^iy is in A for all $i \geq 0$.*

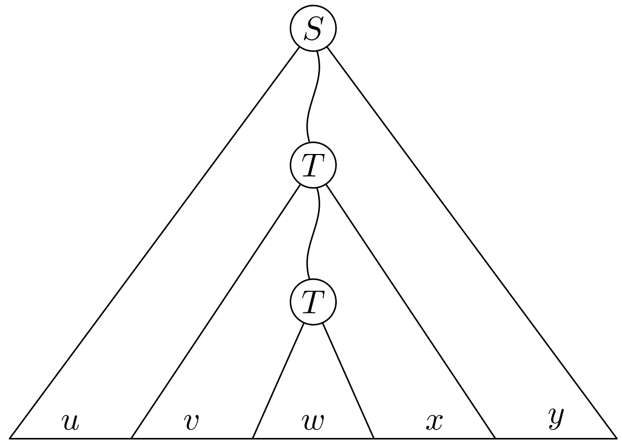
We use notation $z^{(i)} = uv^iwx^iy$.

Proof of Pumping Lemma

Assume A is generated by CFG. Consider long string $z \in A$. Any derivation tree for z has $|z|$ leaves. As there is a bound on the number of children a node can have, if we make z long enough, there must be a path from S to a leaf that contains the same variable twice, say $T \ldots$

Split Based on Repeated Variable

Split the leaves as follows:



This means that $T \stackrel{*}{\Longrightarrow} w$, $T \stackrel{*}{\Longrightarrow} vTx$, and $S \stackrel{*}{\Longrightarrow} uTy$.

Thus $T \stackrel{*}{\Longrightarrow} v^i w x^i$, and so $S \stackrel{*}{\Longrightarrow} uv^i wx^i y$. That is, $z^{(i)} \in A$ for all $i \geq 0$, the main conclusion of the lemma. . .

What's the k?

To get a bound on $|vwx|$, take T to be the lowest variable that is repeated.

The question is: how large must k be? Note that k must work for *all* input.

We should start the proof by putting CFG into Chomsky Normal Form. Then every internal node in derivation tree (except one directly above a leaf) has two children. It follows that $k = 2^{n_A+1}$ works, where n_A is the number of variables in the Chomsky Normal Form grammar.

Example: $\mathtt{0}^n\mathtt{1}^n\mathtt{2}^n$

The classic example is $\{\, \mathtt{0}^n\mathtt{1}^n\mathtt{2}^n : n \geq 0 \,\}$. This language is not context-free.

Let k be the constant of Pumping Lemma. Choose the string $z = \mathtt{0}^k\mathtt{1}^k\mathtt{2}^k$. Consider the split of z into $uvwxy$. Since $|vwx| \leq k$, the string vx cannot contain both $\mathtt{0}$'s and $\mathtt{2}$'s. This means that $z^{(0)} = uwy$ cannot have equal numbers of $\mathtt{0}$'s, $\mathtt{1}$'s and $\mathtt{2}$'s, and is therefore not in the language. A contradiction.

Example: $x\#x$

Consider the language $\{ x\#x : x \in \{0, 1\}^* \}$. This language is not context-free.

Let k be the constant of Pumping Lemma. Choose the string $z = 0^k 1^k \# 0^k 1^k$. Consider the split of z into $uvwxy$. Since $z^{(0)} = uwy$ is in the language, it must have a $\#$; so v occurs before the $\#$ and x after. But since $|vwx| \leq k$, this means that v is a string of 1's and x is a string of 0's, and so $z^{(0)}$ is not in the language after all. A contradiction.

Example. Not.

But note that we do not run into a contradiction with $\{ \texttt{0}^n\texttt{1}^n : n \geq 0 \}$.

For, if we take $z = \texttt{0}^k\texttt{1}^k$, then we may write $z = uvwxy$ where v is the last $\texttt{0}$, x is the first $\texttt{1}$, and w is empty. With such a choice, $z^{(i)} = uv^iwx^iy$ is in the language.

More About Grammars

The exercises in Chapter 9 discuss some more properties of regular languages. These include:

- Context-free languages are not closed under complementation.
- Every context-free language over a unary alphabet is regular.
- There is a special CFG called a linear grammar.
- There is another standard form called Greibach Normal Form.

Practice

Show that the language $\{\, \mathtt{a}^i \mathtt{b}^j \mathtt{c}^i \mathtt{d}^j \; : \; i, j \; > \; 0 \,\}$ is not context-free.

Solution to Practice

Suppose the language is context-free. Let k be the constant of Pumping Lemma. Choose the string $z = \mathtt{a}^k\mathtt{b}^k\mathtt{c}^k\mathtt{d}^k$. Consider the split of z into $uvwxy$. Since vx is nonempty, it contains some symbol. However, since $|vwx| \leq k$, the string vx cannot contain both $\mathtt{a}$'s and $\mathtt{c}$'s, nor can it contain both $\mathtt{b}$'s and $\mathtt{d}$'s. Thus $z^{(0)}$ is not in the language. A contradiction.

Summary

Context-free languages are closed under the Kleene operations but not under intersection or complementation. The Pumping Lemma can be used to prove that a language is not context-free.