

CS 383

Pumping lemma for context-free languages

Stephen Checkoway

Fall 2026

Recall: Pumping lemma for regular languages

If A is a regular language, then there exists $k \geq 0$ (the pumping length) such that for all strings x, y, z with $xyz \in A$, and $|y| \geq k$, there exist strings u, v, w such that $y = uvw$, $|v| \geq 0$, and for all $i \geq 0$, $xuv^i wz \in A$

We used the contrapositive in the form of the “demon game” to prove that languages are not regular

CF-pumpability (a nonstandard name for this property)

A language L is **CF-pumpable** with pumping length k if for all $z \in L$ with $|z| \geq k$, there exist strings $u, v, w, x, y \in \Sigma^*$ with $z = uvwxy$ such that

- ① $|vx| > 0$; and
- ② $|vwx| \leq k$
- ③ for all $i \geq 0$, $uv^iwx^iy \in L$;

Rather than dividing the string into 3 pieces, we're dividing it into 5

Two of the pieces (v and x) are pumped together

Condition 1 tells us that at least one of v or x must not be ε

Example pumpable language

The language $A = \{w\#w^{\mathcal{R}} \mid w \in \{a, b\}^*\}$ is CF-pumpable with pumping length $k = 3$

Example pumpable language

The language $A = \{w\#w^{\mathcal{R}} \mid w \in \{a, b\}^*\}$ is CF-pumpable with pumping length $k = 3$

Every string in A of length at least 3 has the form $z = sc\#cs^{\mathcal{R}}$ for $c \in \{a, b\}$ and $s \in \{a, b\}^*$. Note $|z| = 3 + 2|s| \geq 3$

Example pumpable language

The language $A = \{w\#w^{\mathcal{R}} \mid w \in \{a, b\}^*\}$ is CF-pumpable with pumping length $k = 3$

Every string in A of length at least 3 has the form $z = sc\#cs^{\mathcal{R}}$ for $c \in \{a, b\}$ and $s \in \{a, b\}^*$. Note $|z| = 3 + 2|s| \geq 3$

Let $u = s$

$v = c$

$w = \#$

$x = c$

$y = s^{\mathcal{R}}$

Example pumpable language

The language $A = \{w\#w^{\mathcal{R}} \mid w \in \{a, b\}^*\}$ is CF-pumpable with pumping length $k = 3$

Every string in A of length at least 3 has the form $z = sc\#cs^{\mathcal{R}}$ for $c \in \{a, b\}$ and $s \in \{a, b\}^*$. Note $|z| = 3 + 2|s| \geq 3$

Let $u = s$

$v = c$

$w = \#$

$x = c$

$y = s^{\mathcal{R}}$

① $|vx| = |cc| = 2 > 0$

Example pumpable language

The language $A = \{w\#w^{\mathcal{R}} \mid w \in \{a, b\}^*\}$ is CF-pumpable with pumping length $k = 3$

Every string in A of length at least 3 has the form $z = sc\#cs^{\mathcal{R}}$ for $c \in \{a, b\}$ and $s \in \{a, b\}^*$. Note $|z| = 3 + 2|s| \geq 3$

Let $u = s$

$$v = c$$

$$w = \#$$

$$x = c$$

$$y = s^{\mathcal{R}}$$

$$\textcircled{1} \quad |vx| = |cc| = 2 > 0$$

$$\textcircled{2} \quad |vwx| = |c\#c| = 3 \leq k$$

Example pumpable language

The language $A = \{w\#w^{\mathcal{R}} \mid w \in \{a, b\}^*\}$ is CF-pumpable with pumping length $k = 3$

Every string in A of length at least 3 has the form $z = sc\#cs^{\mathcal{R}}$ for $c \in \{a, b\}$ and $s \in \{a, b\}^*$. Note $|z| = 3 + 2|s| \geq 3$

Let $u = s$

$v = c$

$w = \#$

$x = c$

$y = s^{\mathcal{R}}$

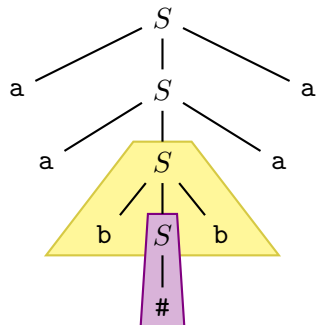
- ① $|vx| = |cc| = 2 > 0$
- ② $|vwx| = |c\#c| = 3 \leq k$
- ③ for any $i \geq 0$, $uv^iwx^iy = sc^i\#c^is^{\mathcal{R}} = (sc^i)\#(sc^i)^{\mathcal{R}} \in L$

Parse trees

CFG for $A = \{w\#w^R \mid w \in \{a, b\}^*\}$: $S \rightarrow aSa \mid bSb \mid \#$

Consider a parse tree for $w = aab\#baa$

$i = 1$:



$u = aa, v = b, w = \#, x = b, y = aa$

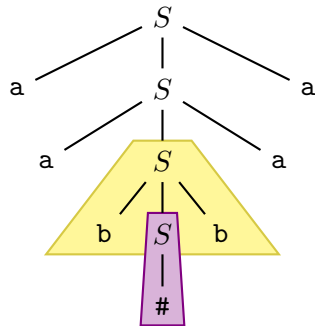
- Pumping down replaces the yellow trapezoid with the violet trapezoid
- Pumping up replaces the violet trapezoid with the yellow trapezoid

Parse trees

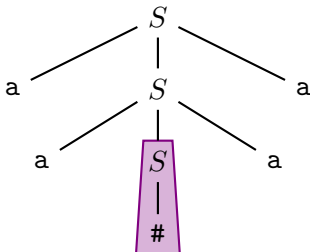
CFG for $A = \{w\#w^R \mid w \in \{a,b\}^*\}$: $S \rightarrow aSa \mid bSb \mid \#$

Consider a parse tree for $w = aab\#baa$

$i = 1$:



$i = 0$:



$u = aa$, $v = b$, $w = \#$, $x = b$, $y = aa$

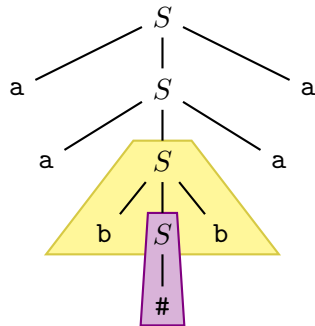
- Pumping down replaces the yellow trapezoid with the violet trapezoid
- Pumping up replaces the violet trapezoid with the yellow trapezoid

Parse trees

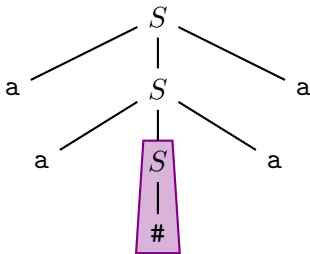
CFG for $A = \{w\#w^R \mid w \in \{a,b\}^*\}$: $S \rightarrow aSa \mid bSb \mid \#$

Consider a parse tree for $w = aab\#baa$

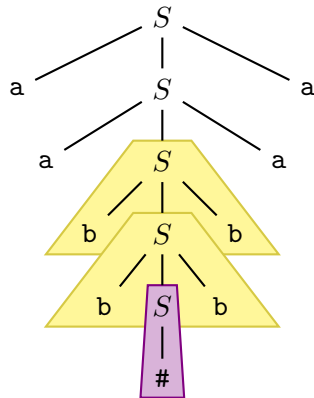
$i = 1$:



$i = 0$:



$i = 2$:



$u = aa$, $v = b$, $w = \#$, $x = b$, $y = aa$

- Pumping down replaces the yellow trapezoid with the violet trapezoid
- Pumping up replaces the violet trapezoid with the yellow trapezoid

CFLs are CF-pumpable

Theorem (Pumping lemma for context-free languages)

Context-free languages are CF-pumpable

CFLs are CF-pumpable

Theorem (Pumping lemma for context-free languages)

Context-free languages are CF-pumpable

Proof idea.

Consider a CFG $G = (N, \Sigma, P, S)$ in CNF

CFLs are CF-pumpable

Theorem (Pumping lemma for context-free languages)

Context-free languages are CF-pumpable

Proof idea.

Consider a CFG $G = (N, \Sigma, P, S)$ in CNF

Set k large enough that any string of length at least k repeats some nonterminal in its derivation (it turns out $k = 2^{|N|} + 1$ works)

CFLs are CF-pumpable

Theorem (Pumping lemma for context-free languages)

Context-free languages are CF-pumpable

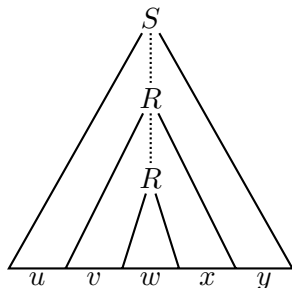
Proof idea.

Consider a CFG $G = (N, \Sigma, P, S)$ in CNF

Set k large enough that any string of length at least k repeats some nonterminal in its derivation (it turns out $k = 2^{|N|} + 1$ works)

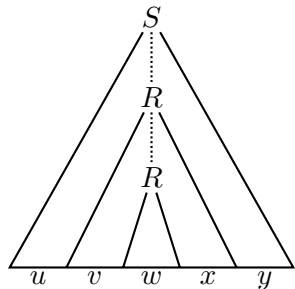
This repeated variable, call it R , will play the same role as the repeated state did in proving that regular languages are pumpable

The diagram to the right shows $S \xRightarrow{*} uRy$,
 $R \xRightarrow{*} vRx$, and $R \xRightarrow{*} w$ (and thus $R \xRightarrow{*} vwx$)



Condition 1: $|vx| > 0$

To see that at least one of v or x is not ε , let's look at
 $R \stackrel{*}{\Rightarrow} vRx$

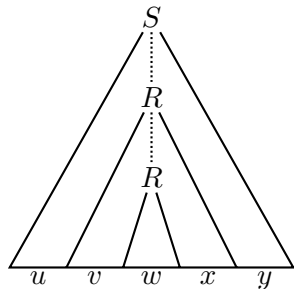


Condition 1: $|vx| > 0$

To see that at least one of v or x is not ε , let's look at $R \stackrel{*}{\Rightarrow} vRx$

Since G is in CNF, we must have $R \Rightarrow AB \stackrel{*}{\Rightarrow} vRx$ for some variables A and B

Two cases:



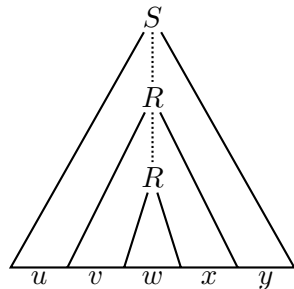
Condition 1: $|vx| > 0$

To see that at least one of v or x is not ε , let's look at $R \stackrel{*}{\Rightarrow} vRx$

Since G is in CNF, we must have $R \Rightarrow AB \stackrel{*}{\Rightarrow} vRx$ for some variables A and B

Two cases:

- $A \stackrel{*}{\Rightarrow} vRs$ and $B \stackrel{*}{\Rightarrow} t$ where $st = x$
 t (and thus x) cannot be ε because G is in CNF



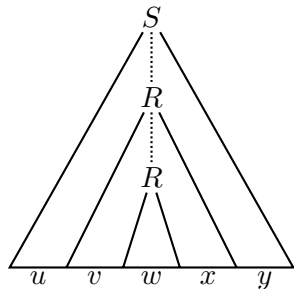
Condition 1: $|vx| > 0$

To see that at least one of v or x is not ε , let's look at $R \stackrel{*}{\Rightarrow} vRx$

Since G is in CNF, we must have $R \Rightarrow AB \stackrel{*}{\Rightarrow} vRx$ for some variables A and B

Two cases:

- $A \stackrel{*}{\Rightarrow} vRs$ and $B \stackrel{*}{\Rightarrow} t$ where $st = x$
 t (and thus x) cannot be ε because G is in CNF
- $A \stackrel{*}{\Rightarrow} s$ and $B \stackrel{*}{\Rightarrow} tRy$ where $st = v$
 s (and thus v) cannot be ε because G is in CNF



Condition 1: $|vx| > 0$

To see that at least one of v or x is not ε , let's look at $R \stackrel{*}{\Rightarrow} vRx$

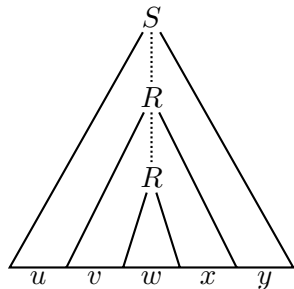
Since G is in CNF, we must have $R \Rightarrow AB \stackrel{*}{\Rightarrow} vRx$ for some variables A and B

Two cases:

- $A \stackrel{*}{\Rightarrow} vRs$ and $B \stackrel{*}{\Rightarrow} t$ where $st = x$
 t (and thus x) cannot be ε because G is in CNF
- $A \stackrel{*}{\Rightarrow} s$ and $B \stackrel{*}{\Rightarrow} tRy$ where $st = v$
 s (and thus v) cannot be ε because G is in CNF

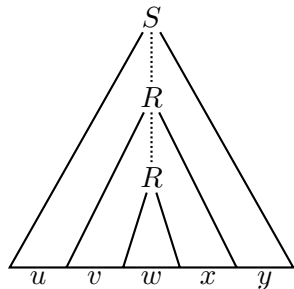
In either case, we've satisfied the first condition:

- 1 $|vx| > 0$



Condition 2: $|vwx| \leq k$

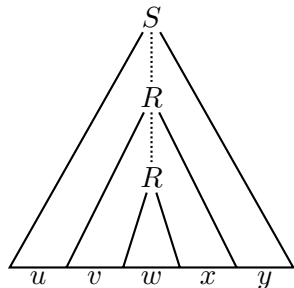
For strings with length at least $k = 2^{|N|} + 1$ we said there had to be a repeated variable



Condition 2: $|vwx| \leq k$

For strings with length at least $k = 2^{|N|} + 1$ we said there had to be a repeated variable

Looking at all paths of height at least $|N| + 1$, there must be a repeated variable (pigeonhole principle), so pick one of those in the lowest $|N| + 1$ for the R that derives $vw x$



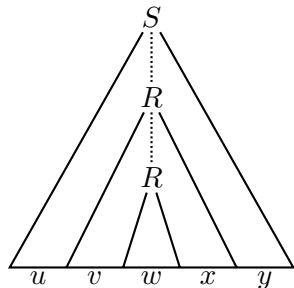
Condition 2: $|vwx| \leq k$

For strings with length at least $k = 2^{|N|} + 1$ we said there had to be a repeated variable

Looking at all paths of height at least $|N| + 1$, there must be a repeated variable (pigeonhole principle), so pick one of those in the lowest $|N| + 1$ for the R that derives vwx

Now since R is at distance at most $|N| + 1$ from the leaves, we must have $|vwx| \leq 2^{|N|} \leq k$

(A perfect binary tree of height h has 2^h leaves, but the last level of interior nodes in a parse tree for a grammar in CNF have a single child each)



Condition 2: $|vwx| \leq k$

For strings with length at least $k = 2^{|N|} + 1$ we said there had to be a repeated variable

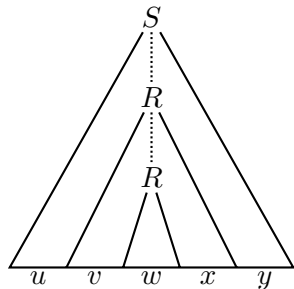
Looking at all paths of height at least $|N| + 1$, there must be a repeated variable (pigeonhole principle), so pick one of those in the lowest $|N| + 1$ for the R that derives vwx

Now since R is at distance at most $|N| + 1$ from the leaves, we must have $|vwx| \leq 2^{|N|} \leq k$

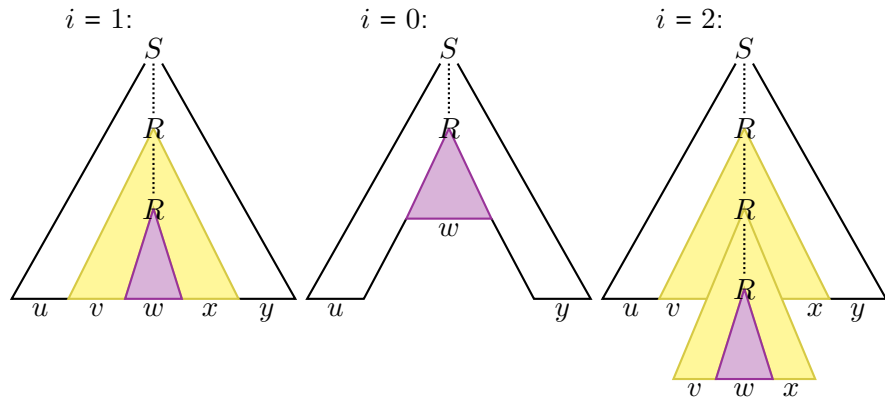
(A perfect binary tree of height h has 2^h leaves, but the last level of interior nodes in a parse tree for a grammar in CNF have a single child each)

Therefore, we've satisfied the second condition:

② $|vwx| \leq k$



Condition 3: $\forall i \geq 0. uv^iwx^iy \in L$



- Pumping down replaces the yellow triangle with the violet triangle
- Pumping up replaces the violet triangle with the yellow triangle
- We can pump up arbitrarily by repeating this process

Thus we've satisfied the final condition:

- ③ for all $i \geq 0$, $uv^iwx^iy \in L$

Showing that a language is not context-free

We can prove that a language is not context-free by showing that it violates the pumping lemma for context-free languages (proof by contradiction)

Steps:

- ➊ Assume the language, L , is context-free with some unspecified pumping length k
- ➋ Pick string $z \in L$ such that $|z| \geq k$
- ➌ Consider every division of z into $uvwxy = z$ such that $|vx| > 0$, and $|vwx| \leq k$
- ➍ For each possible division, show that for some i , $uv^iwx^iy \notin L$

Demon game version of the contrapositive

To prove a language A is not CF, play (and win!) the following game

- 1 The demon picks $k \geq 0$
- 2 You pick $z \in A$ of length at least k
- 3 The demon picks strings u, v, w, x, y such that $uvwxy = z$, $|vx| > 0$, and $|vwx| \leq k$
- 4 You pick $i \geq 0$. If $uv^iwx^iy \notin A$, then you win

Example

$B = \{a^n b^n c^n \mid n \geq 0\}$ is not context-free

Example

$B = \{a^n b^n c^n \mid n \geq 0\}$ is not context-free

First, assume B is context-free with pumping length k (Demon picks k)

Example

$B = \{a^n b^n c^n \mid n \geq 0\}$ is not context-free

First, assume B is context-free with pumping length k (Demon picks k)

Select $z = a^k b^k c^k$ which is in B and has length $3k \geq k$ (You pick z)

Example

$B = \{a^n b^n c^n \mid n \geq 0\}$ is not context-free

First, assume B is context-free with pumping length k (Demon picks k)

Select $z = a^k b^k c^k$ which is in B and has length $3k \geq k$ (You pick z)

Now consider all possible $uvwxy = z$ with $|vx| > 0$ and $|vwx| \leq k$ (Demon picks u, v, w, x, y)

Example

$B = \{a^n b^n c^n \mid n \geq 0\}$ is not context-free

First, assume B is context-free with pumping length k (Demon picks k)

Select $z = a^k b^k c^k$ which is in B and has length $3k \geq k$ (You pick z)

Now consider all possible $uvwx y = z$ with $|vx| > 0$ and $|vwx| \leq k$ (Demon picks u, v, w, x, y)

- At least one of v or x contains two distinct symbols. Then uv^2wx^2y contains symbols out of order so $uv^2wx^2y \notin B$ (You pick $i = 2$)

Example

$B = \{a^n b^n c^n \mid n \geq 0\}$ is not context-free

First, assume B is context-free with pumping length k (Demon picks k)

Select $z = a^k b^k c^k$ which is in B and has length $3k \geq k$ (You pick z)

Now consider all possible $uvwx y = z$ with $|vx| > 0$ and $|vwx| \leq k$ (Demon picks u, v, w, x, y)

- At least one of v or x contains two distinct symbols. Then uv^2wx^2y contains symbols out of order so $uv^2wx^2y \notin B$ (You pick $i = 2$)
- Both v and x contain the same symbol ($v = a^m, x = a^n$; $v = b^m, x = b^n$; or $v = c^m, x = c^n$). Then $uvw y$ doesn't have the same number of as, bs, and cs, so $uv^0wx^0y \notin B$ (You pick $i = 0$)

Example

$B = \{a^n b^n c^n \mid n \geq 0\}$ is not context-free

First, assume B is context-free with pumping length k (Demon picks k)

Select $z = a^k b^k c^k$ which is in B and has length $3k \geq k$ (You pick z)

Now consider all possible $uvwx y = z$ with $|vx| > 0$ and $|vwx| \leq k$ (Demon picks u, v, w, x, y)

- At least one of v or x contains two distinct symbols. Then uv^2wx^2y contains symbols out of order so $uv^2wx^2y \notin B$ (You pick $i = 2$)
- Both v and x contain the same symbol ($v = a^m, x = a^n$; $v = b^m, x = b^n$; or $v = c^m, x = c^n$). Then $uw y$ doesn't have the same number of as, bs, and cs, so $uv^0wx^0y \notin B$ (You pick $i = 0$)
- v and x contain different symbols, but only a single type each ($v = a^m, x = b^n$; $v = a^m, x = c^n$; or $v = b^m, x = c^n$). Again, $uw y$ doesn't have the same number of as, bs, and cs so $uv^0wx^0y \notin B$ (You pick $i = 0$)

Important point about the demon game

Notice that the order of steps matters!

We picked $i = 2$ for the first case but $i = 0$ for the second and third cases

In general, you can pick an i that works for the specific choice of $uvwxy = z$

You still have to consider all possible choices of $uvwxy = z$ though