CS 301

Lecture 14 – Non-context-free languages
Review of “pumpable” languages

Recall we call a language $L$ pumpable with pumping length $p$ if for all $w \in L$ with $|w| \geq p$, there exist strings $x, y, z \in \Sigma^*$ with $w = xyz$ such that

1. for all $i \geq 0$, $xy^iz \in L$;
2. $|y| > 0$; and
3. $|xy| \leq p$

Then we proved that regular languages are pumpable

This let us prove a language was not regular by showing it isn’t pumpable
CF-pumpability

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

1. for all $i \geq 0$, $uv^i xy^i z \in L$;
2. $|vy| > 0$; and
3. $|vxy| \leq p$

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

Two of the pieces ($v$ and $y$) are pumped together

Condition 2 tells us that at least one of $v$ or $y$ must not be $\varepsilon$
Example pumpable language

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

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

Every string in $w$ of length at least 3 has the form $w = sc#cs^R$ for $c \in \{a, b\}$ and $s \in \{a, b\}^*$. Note $\mid w \mid = 3 + 2\mid s\mid \geq 3$
Example pumpable language

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

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

Let $u = s$

$v = c$

$x = \#$

$y = c$

$z = s^R$
Example pumpable language

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

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

Let $u = s$

$v = c$

$x = \#$

$y = c$

$z = s^R$

for any $i \geq 0$, $uv^i xy^i z = sc^i #c^i s^R = (sc^i) #(sc^i)^R \in L$
Example pumpable language

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

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

Let $u = s$
$v = c$
$x = #$
$y = c$
$z = s^R$

1. for any $i \geq 0$, $uw^i xy^i z = sc^i c^i s^R = (sc^i)#(sc^i)^R \in L$
2. $|vy| = |cc| = 2 > 0$
Example pumpable language

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

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

Let $u = s$

$v = c$

$x = \#$

$y = c$

$z = s^R$

1. for any $i \geq 0$, $uw^i xy^i z = sc^i #c^i s^R = (sc^i)#(sc^i)^R \in L$

2. $|vy| = |cc| = 2 > 0$

3. $|vxy| = |c#c| = 3 \leq p$
Parse trees

CFG for \( A = \{w#w^R \mid w \in \{a, b\}^*\} \):
\[
S \rightarrow aSa \mid bSb \mid \#
\]
Consider a parse tree for \( w = aab#baa \)

- \( i = 1 \):

\[
\begin{align*}
S & \quad a \\
S & \quad S \\
a & \quad S \\
a & \quad a \\
a & \quad b \\
b & \quad b \\
S & \quad S \\
\# & 
\end{align*}
\]

- \( u = aa, \ v = b, \ x = \#, \ y = b, \ z = aa \)
  - Pumping down replaces the yellow trapezoid with the violet trapezoid
  - Pumping up replaces the violet trapezoid with the yellow trapezoid
Consider a parse tree for $w = aab#baa$

$i = 1$:

$i = 0$:

$u = aa$, $v = b$, $x = #$, $y = b$, $z = aa$

- 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 $w = aab#baa$

$i = 1$:

$i = 0$:

$i = 2$:

$u = aa$, $v = b$, $x = $, $y = b$, $z = aa$

- 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 $G = (V, \Sigma, R, S)$ in CNF
CFLs are CF-pumpable

Theorem (Pumping lemma for context-free languages)

*Context-free languages are CF-pumpable*

Proof idea.
Consider a CFG $G = (V, \Sigma, R, S)$ in CNF

Set $p$ large enough that any string of length at least $p$ repeats some variable in its derivation (it turns out $p = 2^{|V|} + 1$ works)
CFLs are CF-pumpable

**Theorem (Pumping lemma for context-free languages)**

*Context-free languages are CF-pumpable*

**Proof idea.**

Consider a CFG $G = (V, \Sigma, R, S)$ in CNF

Set $p$ large enough that any string of length at least $p$ repeats some variable in its derivation (it turns out $p = 2|V| + 1$ works)

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

Note that this means $R \Rightarrow^* vxy$ and $R \Rightarrow^* x$
Condition 1: \( \forall i \geq 0. \ uv^i \ xy^i \ z \in L \)

\[ i = 1: \]

\[ \text{Pumping down replaces the yellow triangle with the violet triangle} \]

\[ i = 0: \]

\[ \text{Pumping up replaces the violet triangle with the yellow triangle} \]

\[ i = 2: \]

\[ \text{We can pump up arbitrarily by repeating this process} \]

Thus we’ve satisfied the first condition:

1. for all \( i \geq 0 \), \( uv^i \ xy^i \ z \in L \)
Condition 2: $|vy| > 0$

To see that at least one of $v$ or $y$ is not $\varepsilon$, let's look at $R \Rightarrow vRy$

In either case, we've satisfied the second condition: $|vy| > 0$
Condition 2: $|vy| > 0$

To see that at least one of $v$ or $y$ is not $\varepsilon$, let's look at $R^* \Rightarrow vy$

Since $G$ is in CNF, we must have $R \Rightarrow AB \,^* \Rightarrow vRy$ for some variables $A$ and $B$

Two cases:
Condition 2: \(|vy| > 0\)

To see that at least one of \(v\) or \(y\) is not \(\varepsilon\), let’s look at 
\(R^* \Rightarrow vRy\)

Since \(G\) is in CNF, we must have \(R \Rightarrow AB \Rightarrow^* vRy\) for some variables \(A\) and \(B\)

Two cases:

- \(A \Rightarrow^* vRs\) and \(B \Rightarrow^* t\) where \(st = y\)
  - \(t\) (and thus \(y\)) cannot be \(\varepsilon\) because \(G\) is in CNF
Condition 2: $|vy| > 0$

To see that at least one of $v$ or $y$ is not $\varepsilon$, let's look at $R \ast vRy$

Since $G$ is in CNF, we must have $R \Rightarrow AB \Rightarrow vRy$ for some variables $A$ and $B$

Two cases:

- $A \ast vRs$ and $B \ast t$ where $st = y$
  $t$ (and thus $y$) cannot be $\varepsilon$ because $G$ is in CNF

- $A \ast s$ and $B \ast tRy$ where $st = v$
  $s$ (and thus $v$) cannot be $\varepsilon$ because $G$ is in CNF
Condition 2: \(|vy| > 0\)

To see that at least one of \(v\) or \(y\) is not \(\varepsilon\), let’s look at \(R \Rightarrow vRy\)

Since \(G\) is in CNF, we must have \(R \Rightarrow AB \Rightarrow vRy\) for some variables \(A\) and \(B\)

Two cases:

• \(A \Rightarrow vRs\) and \(B \Rightarrow t\) where \(st = y\)
  
  \(t\) (and thus \(y\)) cannot be \(\varepsilon\) because \(G\) is in CNF

• \(A \Rightarrow s\) and \(B \Rightarrow tRy\) where \(st = v\)
  
  \(s\) (and thus \(v\)) cannot be \(\varepsilon\) because \(G\) is in CNF

In either case, we’ve satisfied the second condition:

\[\text{2} \quad |vy| > 0\]
Condition 3: $|vxy| \leq p$

For strings with length at least $p = 2^{|V|} + 1$ we said there had to be a repeated variable.
Condition 3: $|vxy| \leq p$

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

Looking at all subtrees of height at most $|V| + 1$, there must be a repeated variable (pigeonhole principle), so pick one of those for $R$ that derives $vxy$.
Condition 3: $|vxy| \leq p$

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

Looking at all subtrees of height at most $|V| + 1$, there must be a repeated variable (pigeonhole principle), so pick one of those for $R$ that derives $vxy$.

Now since $R$ is at distance at most $|V| + 1$ from the leaves, we must have $|vxy| \leq 2^{|V|} \leq p$.

(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 3: $|vxy| \leq p$

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

Looking at all subtrees of height at most $|V| + 1$, there must be a repeated variable (pigeonhole principle), so pick one of those for $R$ that derives $vxy$.

Now since $R$ is at distance at most $|V| + 1$ from the leaves, we must have $|vxy| \leq 2^{|V|} \leq p$.

(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 final condition:

$3 \ |vxy| \leq p$
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.

Steps:
1. Assume the language, $L$, is context-free with some unspecified pumping length $p$.
2. Pick string $w \in L$ such that $|w| \geq p$
3. Consider every division of $w$ into $uvxyz = w$ such that $|vy| > 0$, and $|vxy| \leq p$
4. For each possible division, show that for some $i$, $uv^i xy^i z \notin L$
Example

\[ B = \{ a^n b^n c^n \mid n \geq 0 \} \text{ is not context-free} \]
Example

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

First, assume \( B \) is context-free with pumping length \( p \)
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 \( p \)

Select \( w = a^p b^p c^p \) which is in \( B \) and has length \( 3p \geq p \)
Example

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

First, assume \( B \) is context-free with pumping length \( p \)

Select \( w = a^p b^p c^p \) which is in \( B \) and has length \( 3p \geq p \)

Now consider all possible \( uvxyz = w \) with \( |vy| > 0 \) and \( |vxy| \leq p \)
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 $p$

Select $w = a^p b^p c^p$ which is in $B$ and has length $3p \geq p$

Now consider all possible $uvxyz = w$ with $|vy| > 0$ and $|vxy| \leq p$

- At least one of $v$ or $y$ contains two distinct symbols. Then $uv^2xy^2z$ contains symbols out of order so $uv^2xy^2z \notin B$
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 \( p \)

Select \( w = a^p b^p c^p \) which is in \( B \) and has length \( 3p \geq p \)

Now consider all possible \( uvxyz = w \) with \( |vy| > 0 \) and \( |vxy| \leq p \)

- At least one of \( v \) or \( y \) contains two distinct symbols. Then \( uv^2 y^2 z \) contains symbols out of order so \( uv^2 y^2 z \notin B \)
- Both \( v \) and \( y \) contain the same symbol (\( v = a^m \), \( y = a^n \); \( v = b^m \), \( y = b^n \); or \( v = c^m \), \( y = c^n \)). Then \( uxz \) doesn’t have the same number of as, bs, and cs, so \( uv^0 xy^0 z \notin B \)
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 \( p \)

Select \( w = a^p b^p c^p \) which is in \( B \) and has length \( 3p \geq p \)

Now consider all possible \( uvxyz = w \) with \(|vy| > 0 \) and \(|vxy| \leq p \)

- At least one of \( v \) or \( y \) contains two distinct symbols. Then \( uv^2xy^2z \) contains symbols out of order so \( uv^2xy^2z \notin B \)
- Both \( v \) and \( y \) contain the same symbol (\( v = a^m, y = a^n; v = b^m, y = b^n; \) or \( v = c^m, y = c^n \)). Then \( uxz \) doesn’t have the same number of as, bs, and cs, so \( uv^0xy^0z \notin B \)
- \( v \) and \( y \) contain different symbols, but only a single type each (\( v = a^m, y = b^n; v = a^m, y = c^n; \) or \( v = b^m, y = c^n \)). Again, \( uxz \) doesn’t have the same number of as, bs, and cs so \( uv^0xy^0z \notin B \)
Using closure properties

Using the pumping lemma for CFLs is a *pain*

We can prove that

\[ C' = \{ w \mid w \in \{a, b, c\}^* \text{ and } w \text{ has the same number of } a, b, \text{ and } c \} \]

is not context-free by intersecting it with a regular language

What language should we choose?
Using closure properties

Using the pumping lemma for CFLs is a pain

We can prove that

\[ C' = \{ w \mid w \in \{a, b, c\}^* \text{ and } w \text{ has the same number of as, bs, and cs} \} \]

is not context-free by intersecting it with a regular language

What language should we choose?

Intersect with \( a^*b^*c^* \):

\[ C' \cap a^*b^*c^* = B \]
Using closure properties

Using the pumping lemma for CFLs is a *pain*

We can prove that

\[ C' = \{ w \mid w \in \{a, b, c\}^* \text{ and } w \text{ has the same number of } a, b, \text{ and } c \} \]

is not context-free by intersecting it with a regular language

What language should we choose?

Intersect with \( a^*b^*c^* \):

\[ C' \cap a^*b^*c^* = B \]

Since context-free languages are closed under intersection with a regular language, if \( C' \) were context-free, then \( B \) would be context-free.

This is a contradiction so \( C' \) is not context-free.
Another example

\[ D = \{a^n b a^{2n} b a^{3n} \mid n \geq 0\} \] is not context-free
Another example

\[ D = \{ a^n b a^{2n} b a^{3n} \mid n \geq 0 \} \] is not context-free

Pick \( w = a^p b a^{2p} b a^{3p} \) and consider \( uvxyz = w \) such that \( |vy| > 0 \) and \( |vxy| \leq p \)
Another example

\[ D = \{ a^n b a^{2n} b a^{3n} \mid n \geq 0 \} \text{ is not context-free} \]

Pick \( w = a^p b a^{2p} b a^{3p} \) and consider \( uvxyz = w \) such that \( |vy| > 0 \) and \( |vxy| \leq p \)

- If \( v \) or \( y \) contains \( b \), then pumping down gives a string with too few bs
Another example

\[ D = \{ a^n b a^{2n} b a^{3n} \mid n \geq 0 \} \text{ is not context-free} \]

Pick \( w = a^p b a^{2p} b a^{3p} \) and consider \( uvxyz = w \) such that \( |vy| > 0 \) and \( |vxy| \leq p \)

- If \( v \) or \( y \) contains \( b \), then pumping down gives a string with too few \( b \)s
- If \( x \) doesn’t contain \( a \), then \( vxy = a^m \) is in the first, second, or third run of \( as \), for some \( m \). In any case, pumping down gives a string with \( as \) in the wrong ratio
Another example

\[ D = \{ a^n b a^{2n} b a^{3n} \mid n \geq 0 \} \text{ is not context-free} \]

Pick \( w = a^p b a^{2p} b a^{3p} \) and consider \( uvxyz = w \) such that \( |vy| > 0 \) and \( |vxy| \leq p \)

- If \( v \) or \( y \) contains \( b \), then pumping down gives a string with too few bs
- If \( x \) doesn't contain a b, then \( vxy = a^m \) is in the first, second, or third run of as, for some \( m \). In any case, pumping down gives a string with as in the wrong ratio
- If \( x \) contains a b, then either \( v = a^m \) is in the first run of as and \( y = a^n \) is in the second, or \( v \) is in the second and \( y \) is in the third. In either case, pumping down gives a string with as in the wrong ratio
Pumping lemma for CFLs is case analysis hell

Proofs using the pumping lemma always devolve to examining a bunch of cases

Try to use closure properties whenever possible!

If you cannot, here are some general hints

- Try to select $w$ that will lead to as few cases as possible
- Use the fact that $|vxy| \leq p$ to constrain the cases; e.g., if you need some $a$s followed by some $b$s followed by some $c$s, try to have at least $p$ of each so that $vxy$ cannot come from all three
- Try to cover as many similar cases at once as possible; e.g., if several cases are analogous, try to address them in one argument
Pumping lemma for CFLs is case analysis hell

Proofs using the pumping lemma always devolve to examining a bunch of cases

These proofs are inelegant and painful to read/write
Pumping lemma for CFLs is case analysis hell

Proofs using the pumping lemma always devolve to examining a bunch of cases

These proofs are inelegant and painful to read/write

Try to use closure properties whenever possible!
Pumping lemma for CFLs is case analysis hell

Proofs using the pumping lemma always devolve to examining a bunch of cases

These proofs are inelegant and painful to read/write

Try to use closure properties whenever possible!

If you cannot, here are some general hints
  • Try to select $w$ that will lead to as few cases as possible
Pumping lemma for CFLs is case analysis hell

Proofs using the pumping lemma always devolve to examining a bunch of cases

These proofs are inelegant and painful to read/write

Try to use closure properties whenever possible!

If you cannot, here are some general hints

- Try to select $w$ that will lead to as few cases as possible
- Use the fact that $|vxy| \leq p$ to constrain the cases; e.g., if you need some as followed by some bs followed by some cs, try to have at least $p$ of each so that $vxy$ cannot come from all three
Pumping lemma for CFLs is case analysis hell

Proofs using the pumping lemma always devolve to examining a bunch of cases

These proofs are inelegant and painful to read/write

Try to use closure properties whenever possible!

If you cannot, here are some general hints

• Try to select $w$ that will lead to as few cases as possible

• Use the fact that $|vxy| \leq p$ to constrain the cases; e.g., if you need some $a$s followed by some $b$s followed by some $c$s, try to have at least $p$ of each so that $vxy$ cannot come from all three

• Try to cover as many similar cases at once as possible; e.g., if several cases are analogous, try to address them in one argument
Intersection of CFLs

We know that the intersection of a CFL and a regular language is context-free.

Is the intersection of two CFLs necessarily context-free?
Intersection of CFLs

We know that the intersection of a CFL and a regular language is context-free.

Is the intersection of two CFLs necessarily context-free?

No!

What are two context-free languages whose intersection is not context-free?

\[ E = \{ a^m b^m c^n \mid m, n \geq 0 \} \]
\[ F = \{ a^m b^n c^n \mid m, n \geq 0 \} \]
\[ E \cap F = \{ a^n b^n c^n \mid n \geq 0 \} \]
Intersection of CFLs

We know that the intersection of a CFL and a regular language is context-free.

Is the intersection of two CFLs necessarily context-free?
No!

What are two context-free languages whose intersection is not context-free?

\[ E = \{ a^m b^m c^n \mid m, n \geq 0 \} \]
\[ F = \{ a^m b^n c^n \mid m, n \geq 0 \} \]
\[ E \cap F = \{ a^n b^n c^n \mid n \geq 0 \} \]