CS310 : Automata Theory 2019

Lecture 20: Pumping lemma for CFLs

Instructor: Ashutosh Gupta

IITB, India

Compile date: 2019-02-21

Yield size

Theorem 20.1

Let a parse tree be according to a Chomsky-Normal-Form grammar, and the yield of the tree is w. If the length of the longest path is n , then $|w| \leq 2^{n-1}$.

Exercise 20.1

Prove the above theorem via an induction on n.

Pumping lemma for CFLs

Theorem 20.2

Let L be a CFL. Then there is a constant n such that if $z \in L$ such that $|z| \geq n$, then we can write

$$z = uvwxy,$$

subject to the following conditions:

1. $|vwx| \leq n$,
2. $|vx| > 0$, *and*
3. *for each* $i \geq 0$, $uv^iwx^iy \in L$.

Called "tandem" pumping

Exercise 20.2

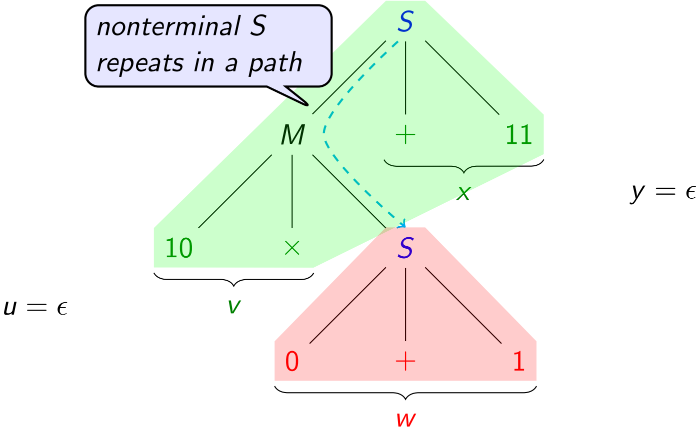
Can the following strings be empty?

- u
- v
- w
- x

Example: tandem pumping

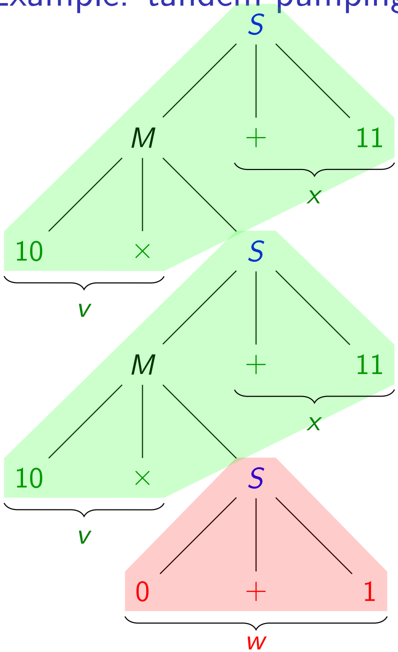
Example 20.1

Consider a parse tree for word $10 \times 0 + 1 + 11$ *due to some CFG*



We can use repeated S *on a path to generate tandem pumping.*

Example: tandem pumping (contd.)



$\underbrace{10\times}_{v}\ \underbrace{10\times}_{v}\ \underbrace{0+1}_{w}\ \underbrace{+11}_{x}\ \underbrace{+11}_{x}$

If parse tree is large enough, we will repeat some nonterminal in a path.

Therefore, tandem pumping.

Pumping lemma for CFLs

Proof.

Let $G = (N, T, P, S)$ be a CNF grammar for L-$\{\epsilon\}$.(why?) Let $|N| = m$.

We cannot find such a grammar if L is $\emptyset$ or $\{\epsilon\}$.
However, in both the cases the theorem trivially holds.(why?)
We need not worry of ϵ word, since we can always choose $n > 0$.

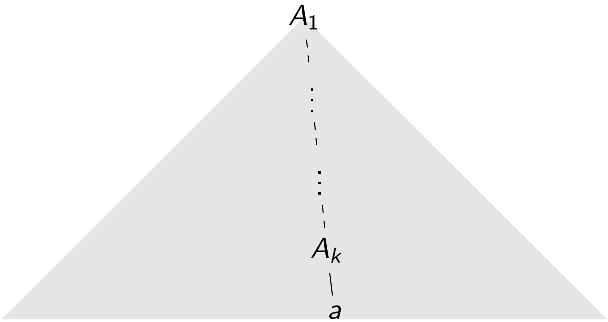
Let $n = 2^m$. Let us choose $z \in L$ such that $|z| \geq n$.
Let us consider a parse tree for z. ...

Pumping lemma for CFLs II

Proof(Contd.).

Due to theorem 20.1, if largest path is a parse tree is m, then the largest yield is $2^{m-1} = n/2$.

Therefore, the parse tree of z has a path longer than m.
Consider annotations on the path be $A_1 \ldots . A_k a$ where $k > m$.



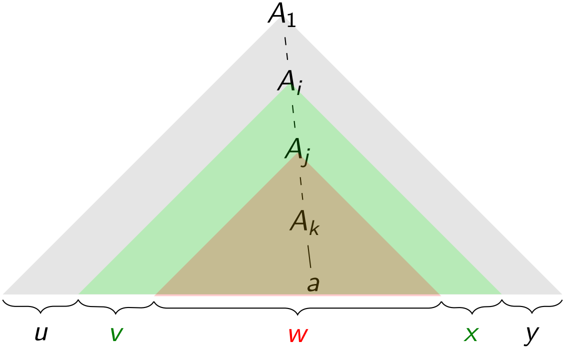
...

Pumping lemma for CFLs III

Proof(Contd.).

There must be i and j such that $A_i = A_j$ and

$$\underbrace{k - m \leq i < j \leq k}_{\text{must be a repeat in } m+1 \text{ nodes}} \quad .$$



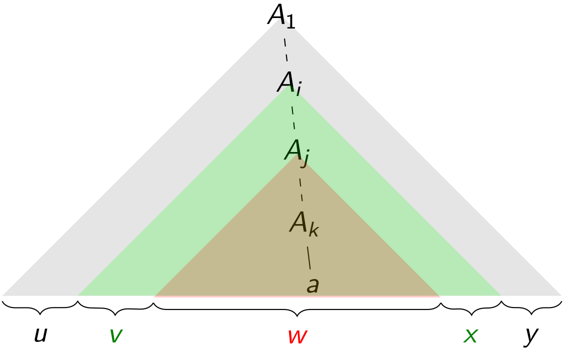
z is broken down to $uvwxy$ according the scheme in the figure. ...

Pumping lemma for CFLs IV

Proof(Contd.).

claim: $|vwx| \leq n$

All paths in subtree from A_i are at most $m+1$. vwx is the yield of the subtree.



Due to theorem 20.1, $|vwx| \leq 2^{(m+1)-1} = n$.

...

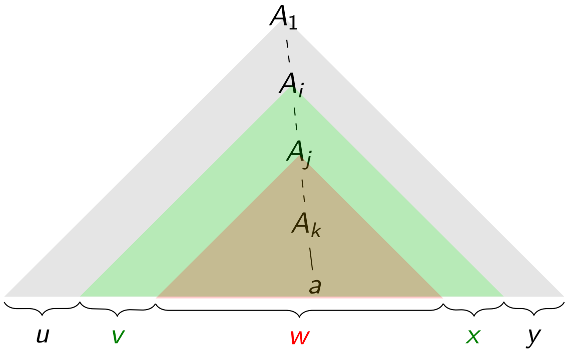
Pumping lemma for CFLs V

Proof(Contd.).

claim: $|vx| > 0$

Since the grammar G is in in CNF, no unit or epsilon productions.

Therefore, one side of the green zone is not empty.(why?)



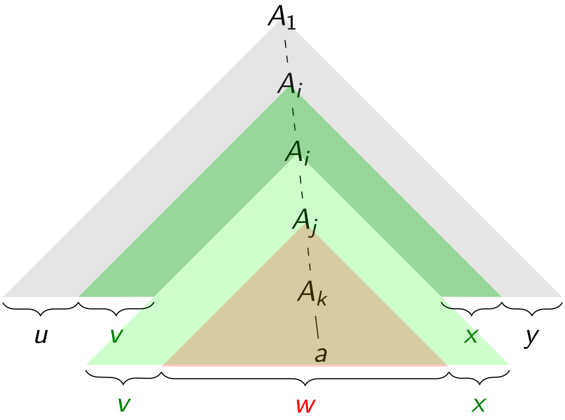
...

Pumping lemma for CFLs VI

Proof(Contd.).

claim: for each $i \geq 0$, $uv^iwx^iy \in L$

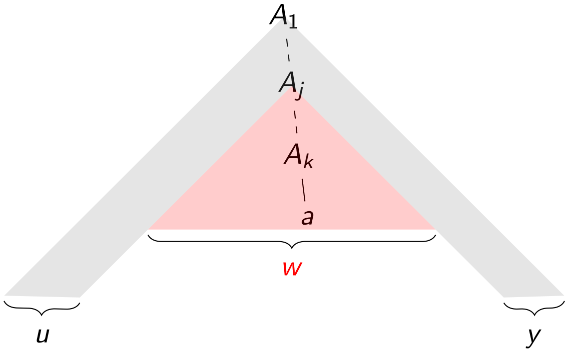
$uvvwxxy \in L$ because of the following parse tree.



Pumping lemma for CFLs VII

Proof(Contd.).

$uwy \in L$ because of the following parse tree.



Following the above examples, we can construct a parse tree for each i. □

Contrapositive of the pumping lemma for CFLs

Theorem 20.3

A language L is not a CFL, if for each n there is *a $z \in L$ such that $|z| \geq n$, and* for each *breakup of $z = uvwxy$, if*

1. $|vwx| \leq n$ *and*
2. $|vx| > 0$,

then there is a $k \geq 0$ such that $uv^kwx^ky \notin L$.

Very similar structure as RL pumping lemma.

We use the theorem to show that languages are not CFLs.

Example 1: using pumping lemma

Example 20.2

Consider language $L = \{0^n1^n2^n | n \geq 1\}$.

- *For each* n, *choose* $z = uvwxy = 0^n1^n2^n \in L$.
- *consider all the subwords* vwx *of* $0^n1^n2^n$ *such that* $|vwx| \leq n$.
- vwx *can not have both* 0 *or* 2.*(why?)* *Wlog, assume* vwx *has no* 2.

$$\underbrace{0\ldots}_{u}\underbrace{\ldots01\ldots}_{vwx}\underbrace{\ldots12\ldots\ldots2}_{y}$$

- *consider all splits of* vwx *such that* $|vx| > 0$.
- *In all splits, the length of either* 0*s or* 1*s will not be* n *in* uwy.*(why?)*
- *Therefore,* $uwy \notin L$.*(why?)*

Therefore, L *is not CFL.*

Example 2: using pumping lemma

Example 20.3

Consider language $L = \{0^n1^m2^n3^m | n \geq 1\}$.

- *For each n, choose $z = uvwxy = 0^n1^n2^n3^n \in L$.*
- *consider all the subwords vwx of $0^n1^n2^n3^n$ such that $|vwx| \leq n$.*
- *vwx can not have more than two symbols.(why?)*
- *There are three cases*

$$\underbrace{0\ldots}_{u}\underbrace{\ldots01\ldots}_{vwx}\underbrace{\ldots12\ldots\ldots23\ldots\ldots3}_{y}$$

$$\underbrace{0\ldots\ldots01\ldots}_{u}\underbrace{\ldots12\ldots}_{vwx}\underbrace{\ldots23\ldots\ldots3}_{y}$$

$$\underbrace{0\ldots\ldots01\ldots\ldots12\ldots}_{u}\underbrace{\ldots23\ldots}_{vwx}\underbrace{\ldots3}_{y}$$

Example 2: using pumping lemma II

- Now consider all splits of vwx such that $|vx| > 0$.
- In all splits, the length of one of

  0s, 1s, 2s, or 3s

  will not be n in uwy and the length of the counterpart

  2s, 3s, 0s, or 1s

  will be n respectively.(why?)

  Length should match with he counterpart
- Therefore, $uv^0wx^0y \notin L$.(why?)

Therefore, L is not CFL.

Exercise 20.3

Is $L = \{0^n1^m2^m3^n | n \geq 1\}$ *CFL?*

Example 3: using pumping lemma

Example 20.4

Consider language $L = \{ww | w \in \{0,1\}^*\}$.

- *For each n, choose $z = uvwxy = 0^n1^n0^n1^n \in L$.*
- *consider subwords vwx of $0^n1^n0^n1^n$ such that $|vwx| \leq n$ and $|vx| > 0$*

v and x can only be from same block or neighboring blocks.

cases	uv^0wx^0y $(n \geq k_1 + k_2 > 0)$	
$\underbrace{0\ldots}_{u}\underbrace{\ldots01\ldots}_{vwx}\underbrace{\ldots10\ldots\ldots01\ldots\ldots1}_{y}$	$0^{n-k_1}1^{n-k_2}0^n1^n$	$\notin L$
$\underbrace{0\ldots\ldots01\ldots}_{u}\underbrace{\ldots10\ldots}_{vwx}\underbrace{\ldots01\ldots\ldots1}_{y}$	$0^n1^{n-k_1}0^{n-k_2}1^n$	$\notin L$
$\underbrace{0\ldots\ldots01\ldots\ldots10\ldots}_{u}\underbrace{\ldots01\ldots}_{vwx}\underbrace{\ldots1}_{y}$	$0^n1^n0^{n-k_1}1^{n-k_2}$	$\notin L$

Therefore, L is not CFL.

End of Lecture 20