Formal Languages and Automata

Pumping Lemma For CFL

Ryan Stansifer

Computer Sciences
Florida Institute of Technology
Melbourne, Florida USA 32901

http://www.cs.fit.edu/~ryan/

3 March 2024

Most formal languages we encounter naturally are context-free grammar. Hence the great utility of CFG's and PDA's. For computation is clearly more powerful than PDA's.

Like before with regular languages we can easily find languages which are not context free. And like before we can easily prove this using a pumping theorem very much like the pumping theorem for regular languages.

Finite languages are little challenge.

The power of NFA's came from the looping necessary for infinite languages. This looping combined with the finite states led to languages which were very regular in structure.

Recursion makes the context-free grammars with their finite number of productions even more powerful than NFA's. Without recursion context-free grammars generate only finite sets of strings.

A context-free grammar with recursion is going to have some part to this effect:

$$S \rightarrow uRy$$
$$R \rightarrow vRx \mid w$$

Notice the either v or x must be non-empty. Otherwise a production like $R \rightarrow R$ is useless.

This grammar allows us to derive infinitely many strings of a certain pattern.

$$S \overset{1}{\Rightarrow} uRy \overset{1}{\Rightarrow} uwy$$
$$S \overset{1}{\Rightarrow} uRy \overset{1}{\Rightarrow} uuRyy \overset{1}{\Rightarrow} uuwyy$$
$$S \overset{1}{\Rightarrow} uRy \overset{1}{\Rightarrow} uuRyy \overset{1}{\Rightarrow} u^3Ry^3 \overset{1}{\Rightarrow} u^3wy^3$$
$$S \overset{1}{\Rightarrow} uRy \overset{1}{\Rightarrow} uuRyy \overset{1}{\Rightarrow} u^3Ry^3 \overset{1}{\Rightarrow} u^4Ry^4 \overset{1}{\Rightarrow} u^4wy^4$$

We see that all strings of the form u^kwy^k are derivable.

Although the argument is intuitively clear, we can prove it carefully and glean some additional information.

Recall . . .

Definition (Chomsky Normal Form)

A grammar is said to be in Chomsky Normal Form (CNF) if all its productions are in one of two forms: $A \to BC$ or $A \to a$.

Theorem (Chomsky Normal Form)

All (epsilon-free) context-free grammars can be put in Chomsky Normal Form.

This has interesting implications about the size of derivations and the number of productions used.

Theorem

In a grammar in Chomsky normal form, it takes exactly $2k - 1$ steps to derive a string of length k.

So, exactly $2k - 1$ productions (not necessary unique) are used in the derivation of string of length k.

For large k, k will be greater than the total number of productions in the grammar. So, some production would have to have been used twice in the derivation of a string of length k (by the pigeon-hole principle).

However, we will need more than this to prove the pumping lemma.

Here is an example grammar in CNF with five productions and three nonterminals:

$$
\begin{aligned}
S &\to XY \\
X &\to XX \mid a \\
Y &\to YY \mid b
\end{aligned}
$$

The grammar has two leftmost derivations of *aabbb*:

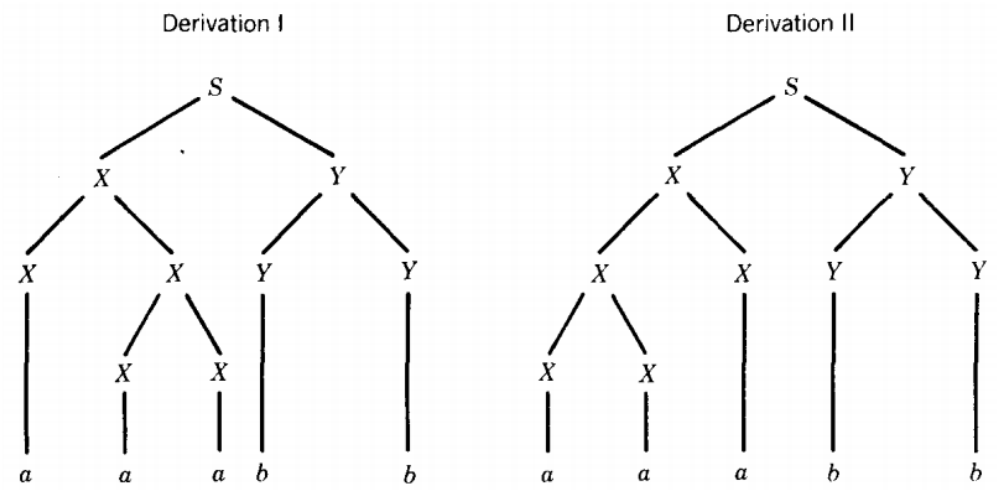
$$S \overset{1}{\Rightarrow} XY \overset{1}{\Rightarrow} XXY \overset{1}{\Rightarrow} aXY \overset{1}{\Rightarrow} aXXY \overset{1}{\Rightarrow} aaXY \overset{1}{\Rightarrow} aaaY \overset{1}{\Rightarrow} aaaYY \overset{1}{\Rightarrow} aaabY \overset{1}{\Rightarrow} aaabb$$

$$S \overset{1}{\Rightarrow} XY \overset{1}{\Rightarrow} XXY \overset{1}{\Rightarrow} XXXY \overset{1}{\Rightarrow} aXXY \overset{1}{\Rightarrow} aaXY \overset{1}{\Rightarrow} aaaY \overset{1}{\Rightarrow} aaaYY \overset{1}{\Rightarrow} aaabY \overset{1}{\Rightarrow} aaabb$$

(Thus, the grammar is ambiguous, but that is not relevant.) The length of the string *aaabb* is five; and the number of steps in either derivation is $2 \times 5 - 1 = 9$.

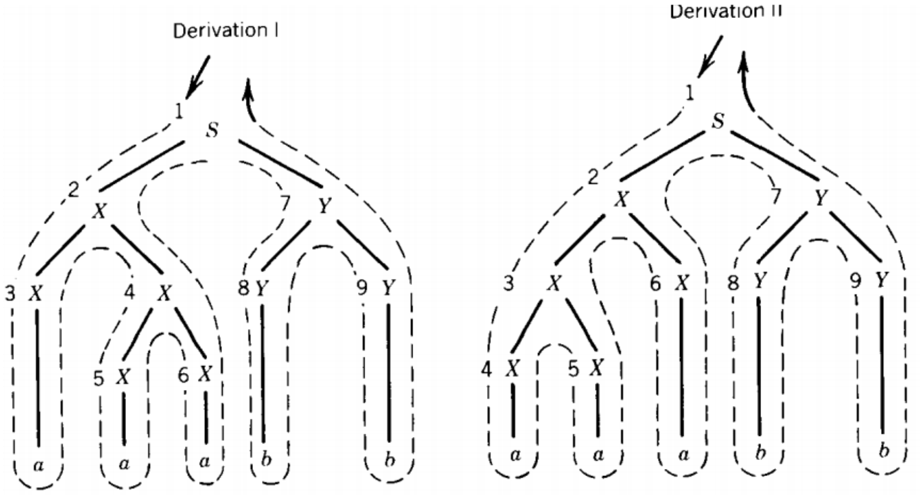
Next consider the parse tree for each derivation. Each interior node in the derivation tree is a step in the derivation.

•


Derivation I
Derivation II
S
X
Y
X
X
Y
Y
X
X
a
a
a
b
b
S
X
Y
X
X
Y
Y
X
X
a
a
a
b
b

Leftmost Derivation and Derivation Tree

•



Theorem

A binary (0, 1 or two children) tree with more than 2^k leaves, has a height of at least k.

Theorem

A binary (0, 1 or two children) tree with more than 2^k leaves, has at least one path from a leaf to the root with more than k interior nodes.

Let G be a context-free grammar in Chomsky normal form with k nonterminals. Then, in a derivation of a string w of length more than 2^k, there must be a least one nonterminal used at least twice.

WLOG we can assume a grammar is in CNF. Or, even, we can merely eliminate ϵ-productions. For we have the following:

Theorem

A d-ary tree with more than d^k leaves, has at least one path from a leaf to the root with more than k interior nodes.

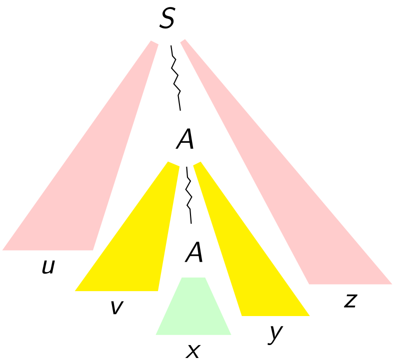
Either way we conclude for any (epsilon-free) CFL there is a grammar in which sufficiently long strings have parse trees in which there is path from leaf to root with repeated nonterminals.

The Pumping Lemma (Context-Free Languages)

Linz 6th, section 8.1, theorem 8.1, page 214.
HMU 3rd, section 7.2, theorem 7.18, page 281.
Kozen, theorem 22.1, page 148.
Sipser 3rd,
Floyd,
Sudkamp 3rd,
Pumping lemma for context-free languages

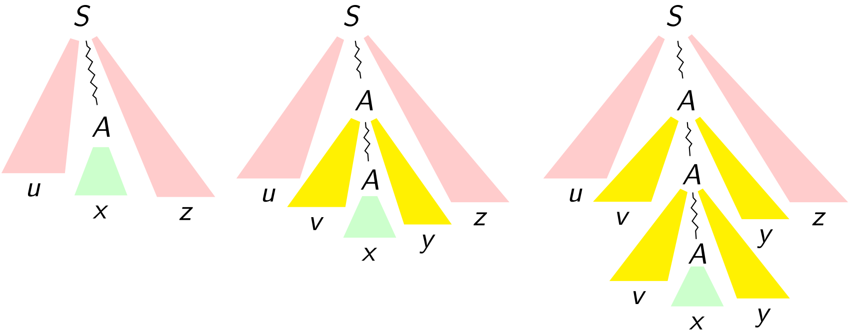
The proof of the pumping lemma for context-free languages depends on understanding the consequences of a repeated use of a nonterminal in the derivation of a long string.

The situation is shown in the following diagram.



Show above is a parse tree for a derivation with a repetition. From which we see three important facts.

$$A \Rightarrow^* x \qquad A \Rightarrow^* vAyS \Rightarrow^* uAz \Rightarrow^* uvAyz \Rightarrow^* uvxyz$$



We conclude for all $i \geq 0$ there is a distinct derivation of $S \overset{*}{\Rightarrow} uv^ixy^iz$.

$$S \Rightarrow^* uAz \Rightarrow^* uxz$$

$$S \Rightarrow^* uAz \Rightarrow^* uvAyz \Rightarrow^* uvxyz$$

$$S \Rightarrow^* uAz \Rightarrow^* uvAyz \Rightarrow^* uvAyz \Rightarrow^* uvvxyyz$$

$$S \Rightarrow^* uv^iAy^iz \Rightarrow^* uv^ixy^iz$$

For a sufficiently long string w in a CFL, there is a derivation for w with at least one nonterminal, say A, repeated.

$$S \stackrel{*}{\Rightarrow} uAz \stackrel{*}{\Rightarrow} uvAyz \stackrel{*}{\Rightarrow} uvxyz = w$$

Without loss of generality we will assume that no other nonterminals repeat in the derivation:

$$A \stackrel{*}{\Rightarrow} vAy \stackrel{*}{\Rightarrow} vxy$$

We have then that len(vxy) is less than 2^k (for otherwise there would be another repetition).
Without loss of generality we can also assume that the grammar has no unit productions. This means A does not derive the sentential form A. As a consequence, v and y above are not both ϵ. So uv^ixy^iz is in the language for all i and the strings are all different.

Pumping Lemma in English

For all languages, if the language is context-free, then

Pumping Lemma in English

For all languages, if the language is context-free, then there is a positive number m such that

Pumping Lemma in English

For all languages, if the language is context-free, then there is a positive number m such that for all sufficiently long strings w in the language

Pumping Lemma in English

For all languages, if the language is context-free, then there is a positive number m such that for all sufficiently long strings w in the language there is a partition $uvxyz$ of w with $\text{len}(vxy) \leq m$, and v and y are not both ϵ such that

Pumping Lemma in English

For all languages, if the language is context-free, then there is a positive number m such that for all sufficiently long strings w in the language there is a partition $uvxyz$ of w with $\mathrm{len}(vxy) \leq m$, and v and y are not both ϵ such that uv^ixy^iz is in the language for all i.

The context-free language pumping lemma:

$$
\begin{aligned}
&\forall L \subseteq \Sigma^* \ \Big(\mathrm{CFL}(L) \Rightarrow \\
&\quad \exists\, m \in \mathbb{N}\ \Big[m > 0 \text{ and} \\
&\qquad \forall\, w \in \Sigma^* \ \Big(w \in L \text{ and } |w| > m \Rightarrow \\
&\qquad\quad \exists\, u, v, x, y, z \in \Sigma^* \ \big[(w = uvxyz \text{ and } |vxy| \leq m \text{ and } |vy| \geq 1) \text{ and} \\
&\qquad\qquad \forall\, i \in \mathbb{N}\ (uv^i xy^i z \in L) \big] \Big) \Big] \Big)
\end{aligned}
$$

Notice that the logical structure of the pumping lemma for

1. regular languages,
2. context-free languages, and
3. linear languages

is the same, but the details of “pumping” are different.

Notice that the logical structure of the pumping lemma for

1. regular languages,
2. context-free languages, and
3. linear languages

is the same, but the details of "pumping" are different.

1. $w = xyz$ and $\text{len}(xy) \leq m$; length of prefix is bounded
2. $w = uvxyz$ and $|vxy| \leq m$; length of middle is bounded
3. $w = uvxyz$ and $|uvyz| \leq m$; length of shoulders are bounded

Proving a Language is Not Context Free

- ($\forall$ intro.) Let m be an arbitrary integer such that $m > 0$.
- ($\exists$ intro.) We pick a string w_m. We show $w_m \in L$ and $\text{len}(w_m) \geq m$.
- ($\forall$ intro.) Let u, v, x, y, z be arbitrary strings such that $uvxyz = w_m$, $\text{len}(vxy) \leq m$, and $0 < \text{len}(vy)$.
- ($\exists$ intro.) We pick a number $i_0 \in \mathbb{N}$. We show that $uv^{i_0}xy^{i_0}z$ is not in L_0.

Proving a Language is Not Context Free

- ($\forall$ intro.) Let m be an arbitrary integer such that $m > 0$.
- ($\exists$ intro.) We pick a string w_m. We show $w_m \in L$ and $\text{len}(w_m) \geq m$.
- ($\forall$ intro.) Let u, v, x, y, z be arbitrary strings such that $uvxyz = w_m$, $\text{len}(vxy) \leq m$, and $0 < \text{len}(vy)$.
- ($\exists$ intro.) We pick a number $i_0 \in \mathbb{N}$. We show that $uv^{i_0}xy^{i_0}z$ is not in L_0.

A key to understanding how one meets one's proof obligations is to think of arbitrary values (for-all introduction) as having been designed by a malevolent opponent to make it as difficult as possible to complete the proof.

Proving a Language is Not Context Free

- (∀ intro.) Let m be an arbitrary integer such that $m > 0$.
- (∃ intro.) We pick a string w_m. We show $w_m \in L$ and $\text{len}(w_m) \geq m$.
- (∀ intro.) Let u, v, x, y, z be arbitrary strings such that $uvxyz = w_m$, $\text{len}(vxy) \leq m$, and $0 < \text{len}(vy)$.
- (∃ intro.) We pick a number $i_0 \in \mathbb{N}$. We show that $uv^{i_0}xy^{i_0}z$ is not in L_0.

A key to understanding how one meets one's proof obligations is to think of arbitrary values (for-all introduction) as having been designed by a malevolent opponent to make it as difficult as possible to complete the proof.

It may be necessary to break into cases to cover all the arbitrary choices.

Proving a Language is Not Context-Free

- The adversary picks a number $m > 0$.
- We pick a string in L with length greater than m.
- The adversary picks strings u, v, x, y, z such that $uvxyz = w$, $|vxy| \leq m$, and $|vy| \geq 1$.
- We pick a number $i \in \mathbb{N}$ such that uv^ixy^iz is not in L.
- We win, if we have a winning strategy; i.e., $uv^ixy^iz \notin L$ no matter what choices the adversary makes.

Applications

Linz 6th, Example 8.1, page 216. We can use the pumping lemma to show the following language is not context-free:

$$L = \{\, a^n b^n c^n \mid 0 \leq n \,\}$$

Pick $w_m = a^m b^m c^m$.

Applications

The language $L_0 = \{ a^n b^n c^n \mid 0 \leq n \}$ is not context free.
Proof:
We apply the pumping lemma to the language $L_0 \subseteq \Sigma^*$.
We assume that L_0 is a context-free language for purposes of obtaining a contradiction. From this assumption it follows that all sufficiently long strings in L_0 can be "pumped." We will prove that, in fact, some sufficiently long strings in L_0 cannot be "pumped."
Let m be an arbitrary integer such that $m > 0$.

We pick $w_m = a^m b^m c^m$. Clearly $w_m \in L_0$ and $\text{len}(w_m) = 3m > m$.
Let u, v, x, y, z **be arbitrary strings such that** $w_m = uvxyz$**,** $\text{len}(vxy) \leq m$**, and** $0 < \text{len}(vy)$**.**
There are two cases. Either $\#_a(vxy) = 0$ or $\#_a(vxy) > 0$.
In the first case, we pick $i = 0$, and $uv^0xy^0z = uxz$ will lose b's or c's or both, but not a's. We have $uv^0xy^0z = uzx \notin L_0$ because it was too many a's.
In the second case, we (again) pick $i = 0$, and $uv^0xy^0z = uxz$ will lose a's or b's or both, but not c's. We have $uv^0xy^0z = uxz \notin L_0$ because it was too many c's.
No matter how u, v, x, y, z are chosen, $uv^0xy^0z = uxz$ is not in L_0.
Hence, not all sufficiently long strings can be "pumped" in L_0**. The assumption has led to a contradiction.**

Therefore, the language L_0 **is not a context-free language. QED**

Applications

Linz 6th, Example 8.2, page 217. We can use the pumping lemma to show the following language is not context-free:

$$L = \{\, ww \,|{\in}\, \Sigma^* \,\}$$

Choose $w = a^m b^m a^m b^m$. See Busch's notes.

Linz 6th, Example 8.4, page 218. We can use the pumping lemma to show the following language is not context-free:

$$L = \{\, a^n b^j \mid n = j^2 \,\}$$

Choose $w = a^{m^2} b^m$.

Challenge

1. Is $L = \{\, a^n w w^R a^n \mid 0 \leq n,\ w \in \Sigma^* \,\}$ context free?
2. Is $L = \{\, a^n b^j a^n b^j \mid 0 \leq n, j \,\}$ context free?

Linz 6th, Section 8.1, Exercise 21, page 222. The language $L = \{a^n b^n c^k \mid n \neq k\}$ is not context-free. It is not possible to prove this using the pumping lemma.

If the number of c's, say k, is larger than the number of a's and b's, say j, then pick $u = a^j b^j$, $v = \epsilon$, $x = \epsilon$, $y = c$, $z = c^{k-1}$ and pump just c's. For all $i \geq 0$ we have $a^j b^j c^i c^{k-1} \in L$. So we lose, the adversary wins, there is no contradiction to the pumping lemma.

If the number of c's, say k, is smaller than the number of a's and b's, say j, then pick $u = a^{j-1}$, $v = a$, $x = \epsilon$, $y = b$, $z = b^{j-1} c^k$ and pump equal number of a's and b's. For all $i \geq 0$ we have $a^{j-1} a^i b^i b^{j-1} c^k \in L$. In this case, too, we lose, the adversary wins, there is no contradiction to the pumping lemma.

Linear Languages

To bound the length of the shoulders we use a repetition closest to the root (not closest to the fringe).

The linear language pumping lemma:

$$\forall L \subseteq \Sigma^* \ \Big(L \text{ is linear} \Rightarrow$$

$$\exists m \in \mathbb{N} \ \Big[m > 0 \text{ and}$$

$$\forall w \in L \ \Big(|w| > m \Rightarrow$$

$$\exists u, v, x, y, z \in \Sigma^* \ \big[(w = uvxyz \text{ and } |uvyz| \leq m \text{ and } |vy| \geq 1) \text{ and}$$

$$\forall i \in \mathbb{N} \ (i \geq 0 \Rightarrow uv^i xy^i z \in L) \big] \Big) \Big] \Big)$$

The linear language pumping lemma:

$$\forall L \subseteq \Sigma^* \ \Big(L \text{ is linear} \Rightarrow$$

$$\exists m \in \mathbb{N} \ \Big[m > 0 \text{ and}$$

$$\forall w \in L \ \Big(|w| > m \Rightarrow$$

$$\exists u, v, x, y, z \in \Sigma^* \ \big[(w = uvxyz \text{ and } |uvyz| \leq m \text{ and } |vy| \geq 1) \text{ and}$$

$$\forall i \in \mathbb{N} \ (i \geq 0 \Rightarrow uv^i xy^i z \in L) \big] \Big) \Big] \Big)$$

Note difference

The linear language pumping lemma:

$$\forall L \subseteq \Sigma^* \; \Big(L \text{ is linear} \Rightarrow$$

$$\exists m \in \mathbb{N} \; \Big[m > 0 \text{ and}$$

$$\forall w \in L \; \Big(|w| > m \Rightarrow$$

$$\exists u, v, x, y, z \in \Sigma^* \; \big[(w = uvxyz \text{ and } |uvyz| \leq m \text{ and } |vy| \geq 1) \text{ and}$$

$$\forall i \in \mathbb{N} \; (i \geq 0 \Rightarrow uv^ixy^iz \in L) \big] \Big) \Big] \Big)$$

Note difference

For CFG: $|vxy| \leq m$; for linear grammar: $|uvyz| \leq m$.

If $\text{len}(vy) < 1$, then $v = \epsilon$ and $v = \epsilon$. And, so, $A \overset{*}{\Rightarrow} A$ which is impossible because we assumed there were no unit productions. Therefore $\text{len}(vy) \geq 1$.

If $\text{len}(vxy) > m$, then there is a cycle

A repeated nonterminal in a derivation tree gives rise to a five-way partition.

Let us look at some examples. This does not help one apply the pumping lemma.
But it does help one understand the proof of the pumping lemma and "pumping."

The language $L = \{a^k \mid 1 \leq k\} \cdot \{b^j c a^j \mid 0 \leq j\}$ is context-free and linear.

On the left is a context-free grammar for it.

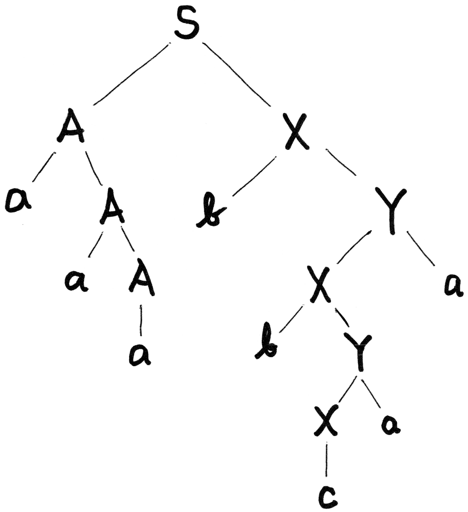
$$\begin{aligned} S &\to AX \\ A &\to aA \mid a \\ X &\to bY \mid c \\ Y &\to Xa \end{aligned} \qquad\qquad \begin{aligned} S &\to aS \mid aX \\ X &\to bY \mid c \\ Y &\to Xa \end{aligned}$$

The first grammar is evidence that L is context-free. (The grammar is not in Chomsky normal form, but pretty close.) The first grammar is not linear since the production $S \to AX$ has two nonterminals in it.

However, this is not evidence that the language is not linear. In fact, the language is linear as seen by the second grammar.

The context-free language $L = \{a^k \mid 1 \leq k\} \cdot \{b^j c a^j \mid 0 \leq j\}$ has the grammars mentioned before.
Consider the string *aaabbcaa* and this derivation in for it:

$$S \overset{1}{\Rightarrow} AX \overset{1}{\Rightarrow} aAX \overset{1}{\Rightarrow} aaAX \overset{1}{\Rightarrow} aaaX \overset{1}{\Rightarrow}$$
$$aaabY \overset{1}{\Rightarrow} aaabXa \overset{1}{\Rightarrow} aaabbY \overset{1}{\Rightarrow} aaabbXaa \overset{1}{\Rightarrow} aaabbcaa$$



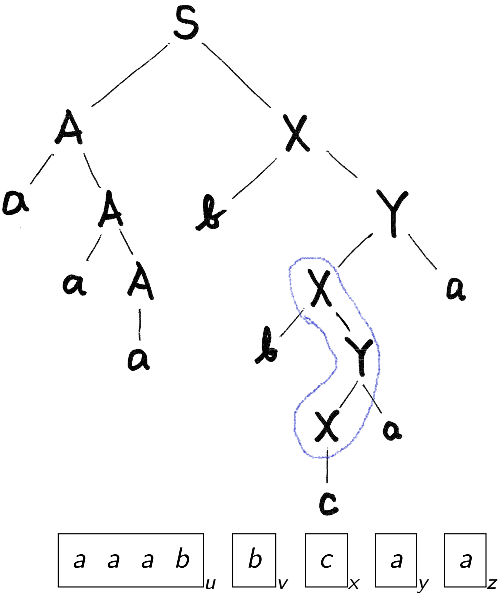
$$S \overset{1}{\Rightarrow} AX \overset{1}{\Rightarrow} aAX \overset{1}{\Rightarrow} aaAX \overset{1}{\Rightarrow} aaaX \overset{1}{\Rightarrow}$$

$$aaabY \overset{1}{\Rightarrow} aaabXa \overset{1}{\Rightarrow} aaabbY \overset{1}{\Rightarrow} aaabbXaa \overset{1}{\Rightarrow} aaabbcaa$$

There are several repeated nonterminals in this one deviation. Let us look at some of them and see how they partition the string $aaabbcaa$.

$$\begin{array}{ll} X \overset{*}{\Rightarrow} c & aaabca \in L \\ X \overset{*}{\Rightarrow} bXa & aaabbcaa \in L \\ X \overset{*}{\Rightarrow} bXa \overset{*}{\Rightarrow} bbXaa & aaabb^2ca^2a \in L \end{array}$$

NB. Given the subtree with the repetition has no other repetitions in it, the fringe of the subtree is bounded.

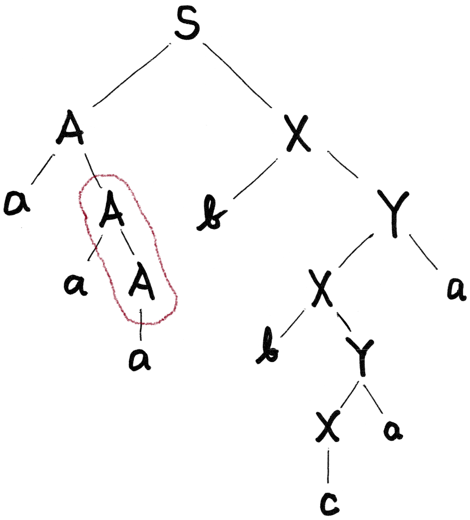


$$A \overset{*}{\Rightarrow} a \qquad aabbcaa \in L$$

$$A \overset{*}{\Rightarrow} aA \qquad aaabbcaa \in L$$

$$A \overset{*}{\Rightarrow} aA \overset{*}{\Rightarrow} aaA \quad aa^2a\epsilon^2bbcaa \in L$$

NB. As in the previous example, given the fact that the subtree with the repetition has no other repetitions in it, the fringe of the subtree is bounded.



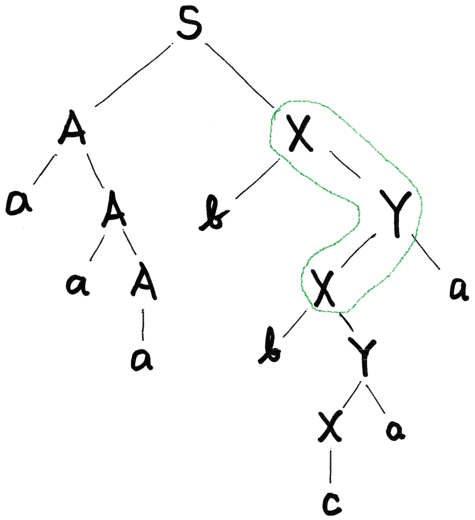
$\boxed{a}_u \ \boxed{a}_v \ \boxed{a}_x \ \boxed{\epsilon}_y \ \boxed{b\ b\ c\ a\ a}_z$

$$X \stackrel{*}{\Rightarrow} bca \qquad aaabca \in L$$

$$X \stackrel{*}{\Rightarrow} bXa \qquad aaabbcaa \in L$$

$$X \stackrel{*}{\Rightarrow} bXa \stackrel{*}{\Rightarrow} bbXaa \qquad aaab^2bcaa^2\epsilon \in L$$

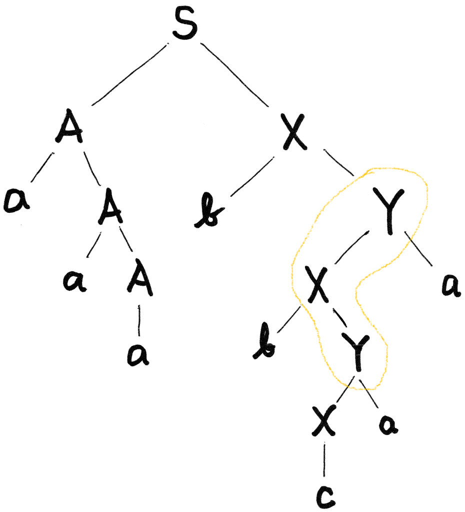
We can pump, but we are unsure how far apart v and y are. The extra information makes applying the pumping lemma easier.



$\boxed{a \quad a \quad a}_x \ \boxed{b}_v \ \boxed{bca}_x \ \boxed{a}_y \ \boxed{\epsilon}_z$

$$Y \overset{*}{\Rightarrow} ca \qquad aaabca \in L$$
$$Y \overset{*}{\Rightarrow} bYa \qquad aaabbcaa \in L$$
$$Y \overset{*}{\Rightarrow} bYa \overset{*}{\Rightarrow} bbYaa \quad aaabb^2caa^2\epsilon \in L$$

We can pump, but we are unsure how far apart v and y are. The extra information makes applying the pumping lemma easier.



$\boxed{a\ a\ a\ b}_x$ $\boxed{b}_v$ $\boxed{ca}_x$ $\boxed{a}_y$ $\boxed{\epsilon}_z$

Let us continue considering the language context-free language $L = \{a^k \mid 1 \leq k\} \cdot \{b^j c a^j \mid 0 \leq j\}$ and a derivation of the string *aaabbcaa*.

But this time consider a derivation in the linear grammar:

$$
\begin{aligned}
S &\rightarrow aS \mid aX \\
X &\rightarrow bY \mid c \\
Y &\rightarrow Xa
\end{aligned}
$$

$$S \overset{1}{\Rightarrow} aS \overset{1}{\Rightarrow} aaS \overset{1}{\Rightarrow} aaaX \overset{1}{\Rightarrow} aaabY \overset{1}{\Rightarrow}$$
$$aaabXa \overset{1}{\Rightarrow} aaabbYa \overset{1}{\Rightarrow} aaabbXaa \overset{1}{\Rightarrow} aaabbcaa$$

Repetition closest to the root

$$S \overset{*}{\Rightarrow} aabbcaa \qquad aabbcaa \in L$$
$$S \overset{*}{\Rightarrow} aS \qquad aaabbcaa \in L$$
$$S \overset{*}{\Rightarrow} aS \overset{*}{\Rightarrow} aaS \quad a^2 a\epsilon^2 bbcaa \in L$$

$$\boxed{\epsilon}_u \ \boxed{a}_v \ \boxed{a\ a\ b\ b\ c\ a\ a}_x \ \boxed{\epsilon}_y \ \boxed{\epsilon}_z$$

Because the repetition is close to the root, the shoulders (uv and yz) have bounded length leaving only part x out of control. Now there is more information with which to trip up the linear pumping lemma on non-linear languages.

Repetition closest to the leaves

$$\begin{array}{rr} X \overset{*}{\Rightarrow} c & aaabca \in L \\ X \overset{*}{\Rightarrow} bAa & aaabbcaa \in L \\ X \overset{*}{\Rightarrow} bAa \overset{*}{\Rightarrow} bbXaa & aaabb^2ca^2a \in L \end{array}$$

$$\boxed{a\ \ a\ \ a\ \ b}_u\ \boxed{b}_v\ \boxed{c}_x\ \boxed{a}_y\ \boxed{a}_z$$

Now the shoulders (uv and yx) escape control and we just the criteria of the context-free pumping lemma and not the little piece of more information of the linear language pumping lemmas.

Application of the Linear Language Pumping Lemma

Theorem

The language $L_0 = \{a^n b^n a^k b^k \mid 1 \le n, k\}$ is not linear.

Proof. **We apply the pumping lemma to the language $L_0 \subseteq \Sigma^*$. We assume that L_0 is a context-free language for purposes of obtaining a contradiction. From this assumption it follows that all sufficiently long strings in L_0 can be "pumped." We will prove that, in fact, some sufficiently long strings in L_0 cannot be "pumped."Let** m **be an arbitrary integer such that** $m > 0$**.**Pick the string $w_m = a^m b^m a^m b^m$. Clearly $w_m \in L_0$ and $\text{len}(w_m) > m$. **Let** u, v, x, y, z **be arbitrary strings such that** $w_m = uvxyz$**,** $\text{len}(vxy) \le m$**, and** $0 < \text{len}(vy)$**. We pick the integer** $i_0 = 0$ **and we will prove** $xy^{i_0}z \notin L_0$.

(The proof continues.)
Since len($uvyz$) $\leq m$ the length of the prefix len(uv) and the length of len(yz) must both be less than or equal to m. So u, and v are a's, and y, and z are all b's. Either v or y (or both) are not ϵ. Remove as few as a single a from the prefix or a single b from the suffix (or both), and the string uxz will not be in L_0, because either the a's no longer equal the b's in the prefix, or the b's no longer equal the a's in the suffix (or both.) **Hence, not all sufficiently long strings can be "pumped" in L_0. The assumption has led to a contradiction.**

Therefore, the language L_0 is not a context-free language. QED

Linz 6th, Section 8.1, Exercise 11, Page 222, Solution Page 427

Show that the language $L = \{a^n b^n a^m b^m \mid 0 \leq n, m\}$ is context-free, but not linear.

Here is a context-free grammar for it.

$$S \to XX$$
$$X \to aXb \mid \epsilon$$

The grammar is evidence that L is context-free. The grammar is not linear. However, the non-linear grammar is not evidence that the language is not linear.

Apply the linear pumping lemma. Given m, we pick $w = a^m b^m a^m b^m$ which is in L and longer than m. For any partition w we have v with a's in the prefix or y with b's in the suffix (or both). Change the number of a's in the prefix or the number of b's in the suffix (or both), yields strings not in L.

What goes wrong if you try to apply the context-free pumping lemma to $L = \{a^k b^k a^j b^j \mid 0 \leq k, j\}$? (It is context-free!)

There is no way to trap the adversary in the $b^k a^j$ middle region. The adversary can always pump the $a^k b^k$ or $a^j b^j$ boundary.

Variation of Exercise 11

The language $L = \{a^n cb^n a^m cb^m \mid 0 \leq n, m\}$ is context-free, but not linear.

Here is a context-free grammar for it.

$$\begin{aligned} S &\to XX \\ X &\to aY \mid c \\ Y &\to Xb \end{aligned}$$

The grammar is evidence that L is context-free. The grammar is not linear.
However, this is not evidence that the language is not linear.

[What L?]
What goes wrong if you try to apply the linear pumping lemma to L?

One may always partition the string $a^m b^m a^m$ such that $v = a^k$ and $y = a^0$. Then $xv^i xz \in L$. For $a^0 b^{m/2} a^{m/2}$ one may take the partition $v = b^k$ and $y = a^k$.

One can pump all long strings.