Middle East Technical University Department of Computer Engineering

Formal Languages and Abstract Machines
Take Home Exam 2

Ugur Duzel
2171569

1 Context-Free Grammars (10 pts)

a) Give the rules of the Context-Free Grammars to recognize strings in the given languages where $\Sigma = \{a, b\}$ and S is the start symbol.

$L(G) = \{w \mid \ w \in \Sigma^*;\ |w| \geq 3;$ (2/10 pts)
the first and the second from the last symbols of w are the same$\}$

$R = \{S \rightarrow Xaa|Xbb,\ X \rightarrow Ya|Yb,\ Y \rightarrow Ya|Yb|e\}$ where $X, Y \in V - \Sigma$

$L(G) = \{w \mid \ w \in \Sigma^*;$ the length of w is odd$\}$ (2/10 pts)

$R = \{S \rightarrow aX|bX,\ X \rightarrow aS|bS|e\}$ where $X \in V - \Sigma$

$L(G) = \{w \mid \ w \in \Sigma^*;\ n(w, a) = 2 \cdot n(w, b)\}$ where $n(w, x)$ is the number of x symbols in w (3/10 pts)

3 possible orderings we accept, (aab, aba, baa) and also e. We need to permute them in every possible way.

$$\begin{aligned} R = \{S &\rightarrow Saab|aSab|aaSb|X|Y|e, \\ X &\rightarrow Xaba|aXba|abXa|S|Y, \\ Y &\rightarrow Ybaa|bYaa|baYa|S|X\} \quad \text{where } X, Y \in V - \Sigma \end{aligned} \tag{1}$$

b) Find the set of strings recognized by the CFG rules given below: (3/10 pts)

$S \to X \mid Y$
$X \to aXb \mid A \mid B$
$A \to aA \mid a$
$B \to Bb \mid b$
$Y \to CbaC$
$C \to CC \mid a \mid b \mid \varepsilon$

$$\begin{aligned} L(G) &= \{a^n b^m \mid n \neq m,\ 0 < n+m;\ n,m \in \mathbb{N}\} \cup \{a^k b a^{k+1} \mid 0 \leq k,\ k \in \mathbb{N}\} \cup \{b^{k+1} a b^k \mid 0 \leq k,\ k \in \mathbb{N}\} \\ &= \{a^n b^m \mid n \neq m,\ 0 < n+m;\ n,m \in \mathbb{N}\} \cup \{w \mid w \in \Sigma^*;\ \text{ w contains ba }\} \end{aligned}$$

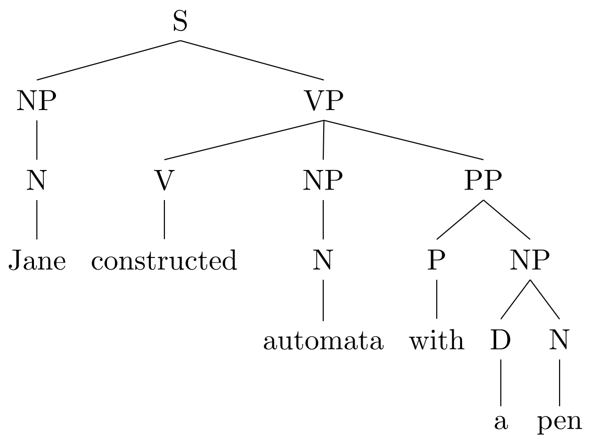
(2)

2 Parse Trees and Derivations (20 pts)

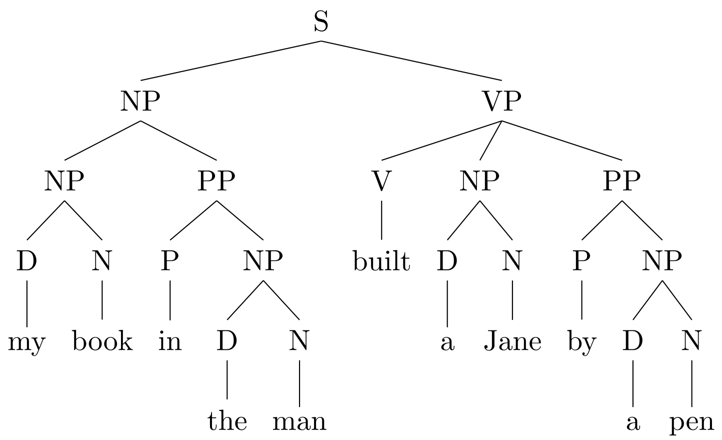
Given the CFG below, provide parse trees for given sentences in **a** and **b**.

```
S   → NP VP
VP  → V NP | V NP PP
PP  → P NP
NP  → N | D N | NP PP
V   → wrote | built | constructed
D   → a | an | the | my
N   → John | Mary | Jane | man | book | automata | pen | class
P   → in | on | by | with
```

a) Jane constructed automata with a pen (4/20 pts)



b) my book in the man built a Jane by a pen (4/20 pts)



Given the CFG below, answer **c**, **d** and **e**

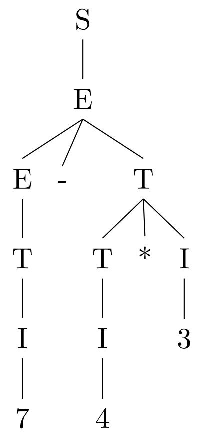
```
S  → E
E  → E + T | E - T | T
T  → T * I | T / I | I
I  → 0 | 1 | 2 | 3 | 4 | 6 | 7 | 8 | 9
```

c) Provide the left-most derivation of 7 - 4 * 3 step-by-step and plot the final parse tree matching that derivation (4/20 pts)

$S \Rightarrow E \Rightarrow E - T \Rightarrow T - T \Rightarrow I - T \Rightarrow 7 - T \Rightarrow 7 - T * I \Rightarrow 7 - I * I \Rightarrow 7 - 4 * I \Rightarrow 7 - 4 * 3$
! The parse tree is identical to the parse tree in part d. (I was not able to paste it here as well due to some tcolorbox error I couldn't solve)

d) Provide the right-most derivation of 7 - 4 * 3 step-by-step and plot the final parse tree matching that derivation (4/20 pts)

$S \Rightarrow E \Rightarrow E - T \Rightarrow E - T * I \Rightarrow E - T * 3 \Rightarrow E - I * 3 \Rightarrow E - 4 * 3 \Rightarrow T - 4 * 3 \Rightarrow I - 4 * 3 \Rightarrow 7 - 4 * 3$

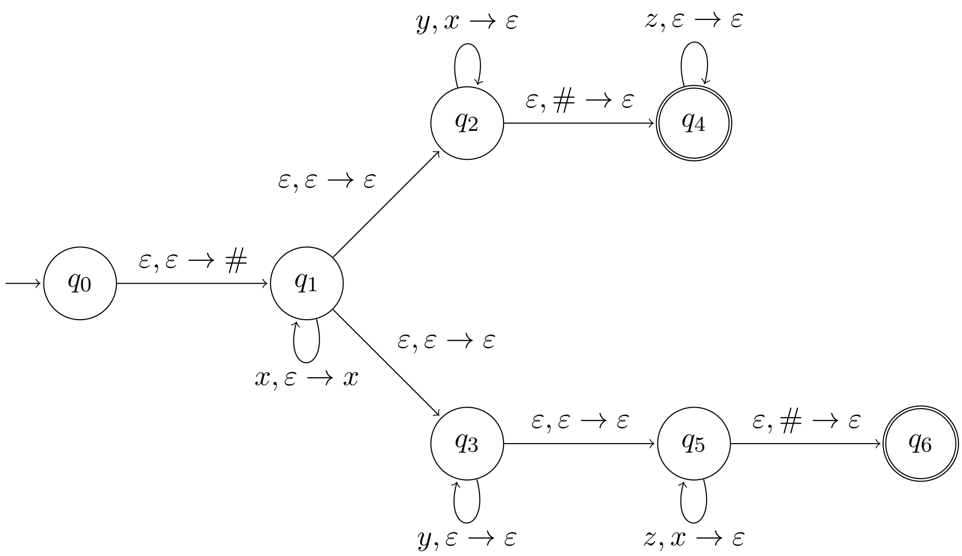


e) Are the derivations in **c** and **d** in the same similarity class? (4/20 pts)

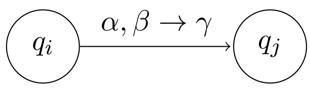
Yes. If we have to elaborate, we can go from c to d or from d to c by repeatedly following either a $\prec$, or an inverted $\prec$.
Also, we can say that if two derivations have the same parse tree then they are in the same similarity class. In this case both c and d have the same parse tree.

3 Pushdown Automata (30 pts)

a) Find the language recognized by the PDA given below (5/30 pts)

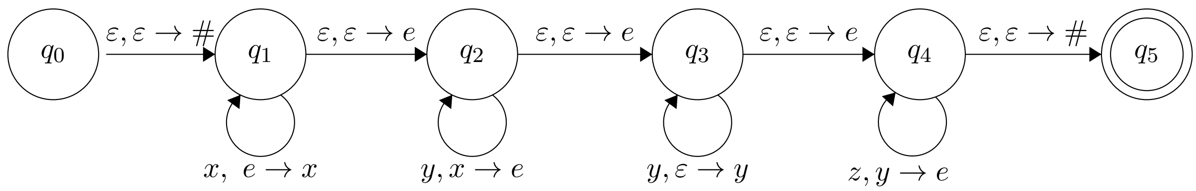


where the transition $((q_i, \alpha, \beta), (q_j, \gamma))$ is represented as:



$L = \{x^n y^n z^m \mid n, m \geq 0;\ n, m \in \mathbb{N}\} \cup \{x^n y^m z^n \mid n, m \geq 0;\ n, m \in \mathbb{N}\}$

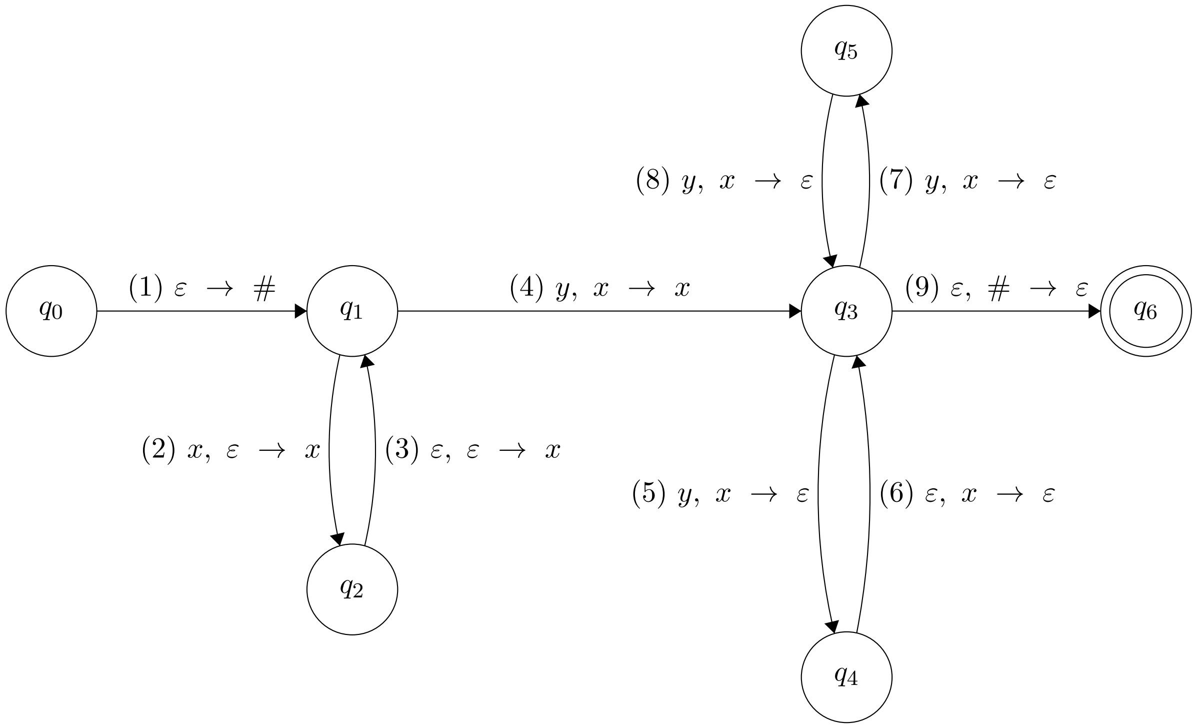
b) Design a PDA to recognize language $L = \{x^n y^{m+n} x^m \mid\ n, m \geq 0;\ n, m \in \mathbb{N}\}$ (5/30 pts)



c) Design a PDA to recognize language $L = \{x^n y^m \mid n < m \leq 2n;\ n, m \in \mathbb{N}^+\}$ (10/30 pts)
Do not use multi-symbol push/pop operations in your transitions.
Simulate the PDA on strings xxy (with only one rejecting derivation) and $xxyyyy$ (accepting derivation) with transition tables.

$M = (K, \Sigma, \Gamma, \Delta, q_0, \{q_6\})$ where $K = \{q_0, q_1, q_2, q_3, q_4, q_5, q_6\}$ and $\Sigma = \{a, b\}$

$$\begin{aligned}
\Delta = \{&((q_0, \varepsilon, \varepsilon), (q_1, \#)), \quad (1)\\
&((q_1, x, \varepsilon), (q_2, x)), \quad (2)\\
&((q_2, \varepsilon, \varepsilon), (q_1, x)), \quad (3)\\
&((q_1, y, x), (q_3, x)), \quad (4)\\
&((q_3, y, x), (q_4, \varepsilon)), \quad (5)\\
&((q_4, \varepsilon, x), (q_3, \varepsilon)), \quad (6)\\
&((q_3, y, x), (q_5, \varepsilon)), \quad (7)\\
&((q_5, y, x), (q_3, \varepsilon)), \quad (8)\\
&((q_3, \varepsilon, \#), (q_6, \varepsilon))\} \quad (9)
\end{aligned} \tag{3}$$



State	Unread Input	Stack	Transition
q_0	xxy	e	-
q_1	xxy	#	1
q_2	xy	x#	2
q_1	xy	xx#	3
q_2	y	xxx#	2
q_1	y	xxxx#	3
q_3	e	xxxx#	4

Input xxy is not accepted.

State	Unread Input	Stack	Transition
q_0	xxyyyy	e	-
q_1	xxyyyy	#	1
q_2	xyyyy	x#	2
q_1	xyyyy	xx#	3
q_2	yyyy	xxx#	2
q_1	yyyy	xxxx#	3
q_3	yyy	xxxx#	4
q_4	yy	xxx#	5
q_3	yy	xx#	6
q_5	y	x#	7
q_3	e	#	8
q_6	e	e	9

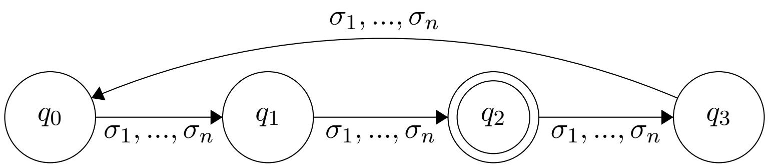
Input xxyyyy is accecpted.

d) Given two languages L' and L as $L' = \{w \mid \ w \in L;\ |w| = 4n + 2\ for\ n \in \mathbb{N}\}$ (10/30 pts)
If L is a CFL, show that L' is also a CFL by constructing an automaton for L' in terms of another automaton that recognizes L.

The given language $L' = L \cap L''$ where $L'' = \{w \mid |w| = 4n + 2\ for\ n \in \mathbb{N}\}$.
Lets say that $L = L(M_1)$ for some pushdown automaton $M_1 = (K_1, \Sigma, \Gamma_1, \Delta_1, s_1, F_1)$.
Furthermore, we can easily show that L'' is a regular language by constructing a deterministic finite automaton as follows:
Let $M_2 = (K_2, \Sigma, \delta, s_2, F_2)$ where $K_2 = \{q_0, q_1, q_2, q_3\},\ \Sigma = \{\sigma_1,\ ...,\ \sigma_n\},\ s_2 = q_0,\ F_2 = \{q_2\}$

$$\begin{aligned}
\delta(q_0, \sigma_i) &= q_1 \\
\delta(q_1, \sigma_i) &= q_2 \\
\delta(q_2, \sigma_i) &= q_3 \\
\delta(q_3, \sigma_i) &= q_0 \ \text{ for any } \sigma_i \in \Sigma
\end{aligned} \tag{4}$$

and we say that $L'' = L(M_2)$. To illustrate,



We need to construct a pushdown automaton by combining these two automata (M_1, M_2) into a single pushdown automaton so that we can show L' is a context free language. Let this automaton be $M = (K, \Sigma, \Gamma, \Delta, s, F)$,

$$\begin{aligned}
K &= K_1 \text{x} K_2 \\
\Gamma &= \Gamma_1 \\
s &= (s_1, s_2) \\
F &= F_1 \text{x} F_2
\end{aligned} \tag{5}$$

For each transition of form $((q_x, \sigma, \beta), (p_x, \gamma)) \in \Delta_1$, and for each state $q_0, q_1, q_2, q_3 \in K_2$ we add the following transition to Δ:

$$(((q_x, q_i), \sigma, \beta), ((p_x, \delta(q_i, \sigma))), \gamma) \quad \text{for any } \sigma \in \Sigma \text{ and } \underline{\text{for each}}\ q_i \in K_2 = \{q_0, q_1, q_2, q_3\}$$

For each transition of form $((q_x, e, \beta), (p_x, \gamma)) \in \Delta_1$, and for each state $q_0, q_1, q_2, q_3 \in K_2$ we add the following transition to Δ:

$$(((q_x, q_i), e, \beta), ((p_x, q_i), \gamma) \quad \text{for any } \sigma \in \Sigma \text{ and } \underline{\text{for each}}\ q_i \in K_2 = \{q_0, q_1, q_2, q_3\}$$

$$\begin{aligned}
\Delta = \{&(((q_x, q_0), \sigma, \beta), ((p_x, q_1)), \gamma) \\
&((q_x, q_1), \sigma, \beta), ((p_x, q_2)), \gamma) \\
&((q_x, q_2), \sigma, \beta), ((p_x, q_3)), \gamma) \\
&((q_x, q_3), \sigma, \beta), ((p_x, q_0)), \gamma)\} \ \text{ for every } \beta, \gamma \in \Gamma_1 \\
&\ \text{and so on... (there may need to be additions to them depending on } M_1)
\end{aligned} \tag{6}$$

We now can easily see that $w \in L(M)$ if and only if $w \in L(M_1) \cap L(M_2)$.
Meaning, $L(M) = L \cap L''$ and $L' = L(M)$.

4 Closure Properties (20 pts)

Let L_1 and L_2 be context-free languages which are not regular, and let L_3 be a regular language. Determine whether the following languages are necessarily CFLs or not. If they need to be context-free, explain your reasoning. If not, give one example where the language is a CFL and a counter example where the language is not a CFL.

a) $L_4 = L_1 \cap (L_2 \setminus L_3)$ (10/20 pts)

L_4 is not necessarily context-free.
$(L_2 \setminus L_3)$ this part can be written as $(L_2 \cap L_3')$. Since L_3 is a regular language we know from the second chapter that regular languages are closed under complementation, meaning L_3' is also regular. That's why $(L_2 \cap L_3')$ can be interpreted as the intersection of a context-free language and a regular language. From the textbook Theorem 3.5.2 we know that intersection of a context-free language and a regular language is context-free language.

$$\begin{aligned} L_4 = &\ L_1 \cap (L_2 \cap L_3') \\ = &\ \text{Context-free} \cap (\text{Context-free} \cap \text{Regular}) \\ = &\ \text{Context-free} \cap \text{Context-free} \neq \text{Context-free} \ \ \text{(not necessarily)} \end{aligned} \tag{7}$$

Not necessarily context-free because the class of context-free languages is not closed under intersection.
CFL example :
$L_1 = a^*b^*$ is both regular and context-free because the class of regular languages is the proper subset of the class of context-free langauges. $L_2 = \{a^nb^n \mid n \geq 0,\ n \in \mathbb{N}\}$. $L_1 \cap L_2 = \{a^nb^n \mid n \geq 0,\ n \in \mathbb{N}\}$ which is a context-free language.
Counter example :
$L_1 = \{a^mb^nc^n \mid n,m \geq 0;\ n,m \in \mathbb{N}\}$ and $L_2 = \{a^nb^nc^m \mid n,m \geq 0;\ n,m \in \mathbb{N}\}$. $L_1 \cap L_2 = \{a^nb^nc^n \mid n \geq 0,\ n \in \mathbb{N}\}$ which is not a context free language.

Therefore L_4 is not necessarily context-free.

b) $L_5 = (L_1 \cap L_3)^*$ (10/20 pts)

L_5 is context-free.
From the textbook Theorem 3.5.2 we know that intersection of a context-free language and a regular language is context-free language.
In this question we see that $L_1 \cap L_3$ is context-free since L_1 is context-free and L_3 is regular. Context free languages are closed under Kleene star operation. Meaning,

$$\begin{aligned} L_5 = &\ (L_1 \cap L_3)^* \\ = &\ (\text{Context-free})^* \\ = &\ \text{Context-free} \end{aligned} \tag{8}$$

Therefore L_5 is indeed a context-free language.

5 Pumping Theorem (20 pts)

a) Show that $L = \{a^n m^n t^i \mid n \le i \le 2n\}$ is not a Context Free Language (10/20 pts) using Pumping Theorem for CFLs.

Assume that language L is a context-free language then the pumping theorem would apply. Given pumping constant $K \ge 1$,

$$w = a^K m^K t^K \in L, \quad |w| = 3K > K$$

Then by the pumping theorem w can be written as follows,

$$w = uvxyz \text{ such that } vy \ne e,\ w' = |vxy| \le K$$

Now we have two general cases,

<u>w' **contains only 1 symbol**</u>

$w' = a^j$

$w' = m^j$

$w' = t^j$ where $1 \le j \le K$

<u>w' **contains 2 symbols**</u>

$w' = a^j m^l$

$w' = m^j t^l$ where $1 \le j + l \le K$

<u>w' **cannot contain** at **or** amt **because** $|w'| \le K$</u>

By the pumping theorem we can say that $uv^n xy^n z \in L$ must be true for all $n \ge 0$,

<u>**Case 1 :**</u>

$w' = a^j$ where $1 \le j \le K$. Take n=0,

$$uxz = a^{K'} m^K t^K, \text{ where } K' < K \text{ therefore } uxz \notin L$$

$w' = m^j$ where $1 \le j \le K$. Take n=0,

$$uxz = a^K m^{K'} t^K, \text{ where } K' < K \text{ therefore } uxz \notin L$$

$w' = t^j$ where $1 \le j \le K$. Take n=0,

$$uxz = a^K m^K t^{K'}, \text{ where } K' < K \text{ therefore } uxz \notin L$$

<u>**Case 2 :**</u>

$w' = a^j m^l$ where $1 \le j + l \le K$. Take n=0,

$$uxz = a^{K'} m^{K''} t^K, \text{ where } K' < K \text{ and } K'' < K \text{ therefore } uxz \notin L$$

$w' = m^j t^l$ where $1 \le j + l \le K$. Take n=0,

$$uxz = a^K m^{K'} t^{K''}, \text{ where } K' < K \text{ and } K'' < K \text{ therefore } uxz \notin L$$

There exists a contradiction with the assumption we made.
Therefore, L is not context-free.

b) Show that $L = \{a^n b^{2n} a^n \mid n \in \mathbb{N}+\}$ is not a Context Free Language using Pumping Theorem for CFLs. (10/20 pts)

Assume that language L is a context-free language then the pumping theorem would apply. Given pumping constant $K \geq 1$,

$$w = a^K b^{2K} a^K \in L, \quad |w| = 4K > K$$

Then by the pumping theorem w can be written as follows,

$$w = uvxyz \text{ such that } vy \neq e, \; w' = |vxy| \leq K$$

Now we have two general cases,

<u>**w' contains only 1 symbol**</u>

$w' = a^j$

$w' = b^j$ where $1 \leq j \leq K$

<u>**w' contains 2 symbols**</u>

$w' = a^j b^l$

$w' = b^j a^l$ where $1 \leq j + l \leq K$

<u>**w' cannot contain aa or aba because $|w'| \leq K$**</u>

By the pumping theorem we can say that $uv^n xy^n z \in L$ must be true for all $n \geq 0$,

<u>**Case 1 :**</u>

$w' = a^j$ where $1 \leq j \leq K$. Take n=0,

$$uxz = a^{K'} b^{2K} a^K \text{ or } uxz = a^K b^{2K} a^{K'}, \text{ for both of the cases } K' < K \text{ therefore } uxz \notin L$$

$w' = b^j$ where $1 \leq j \leq K$. Take n=0,

$$uxz = a^K b^T a^K, \text{ where } T < 2K \text{ therefore } uxz \notin L$$

<u>**Case 2 :**</u>

$w' = a^j b^l$ where $1 \leq j + l \leq K$. Take n=0,

$$uxz = a^{K'} b^T a^K, \text{ where } K' < K \text{ and } T < 2K \text{ therefore } uxz \notin L$$

$w' = b^j a^l$ where $1 \leq j + l \leq K$. Take n=0,

$$uxz = a^K b^T a^{K'}, \text{ where } K' < K \text{ and } T < 2K \text{ therefore } uxz \notin L$$

There exists a contradiction with the assumption we made.
Therefore, L is not context-free.