Middle East Technical University Department of Computer Engineering

Formal Languages and Abstract Machines
Take Home Exam 2

Beyazıt Yalçınkaya
2172138

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.

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

$S \rightarrow aTaa \mid aTab \mid bTbb \mid bTba$
$T \rightarrow aT \mid bT \mid \varepsilon$

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

$S \rightarrow aSa \mid aSb \mid bSa \mid bSb \mid a \mid b$

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

$S \rightarrow TaTaTbT \mid TaTbTaT \mid TbTaTaT$
$T \rightarrow TaTaTbT \mid TaTbTaT \mid TbTaTaT \mid \varepsilon$

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

$S \rightarrow X \mid Y$
$X \rightarrow aXb \mid A \mid B$

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

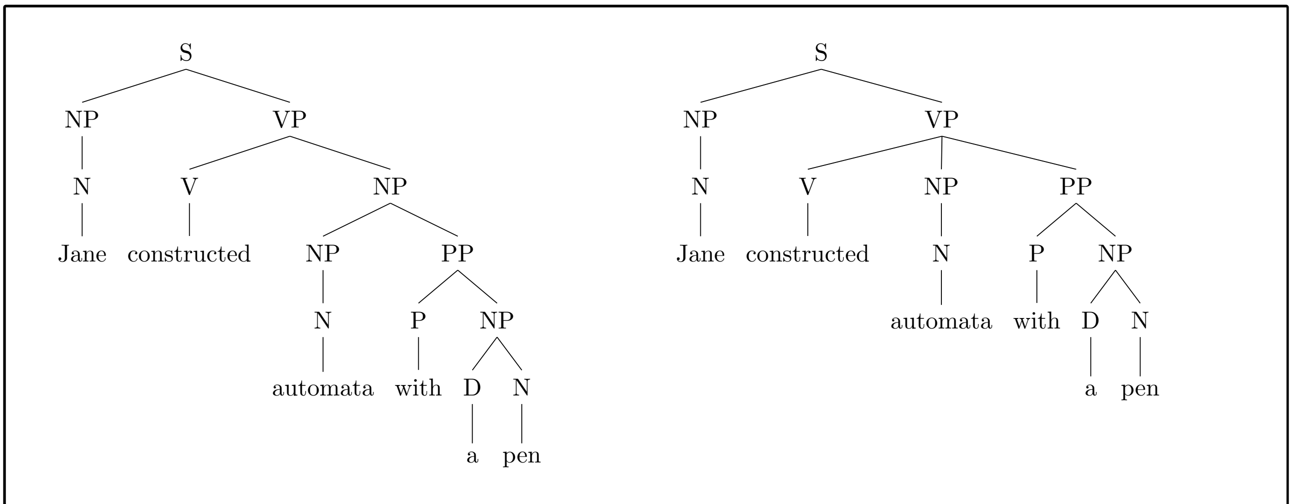
$$\boxed{L = \{a^n b^m \mid n \neq m,\ n, m \geq 1,\ n, m \in \mathbb{N}^+\} \cup \{w \in \{a,b\}^* \mid w \text{ contains } ba\}}$$

2 Parse Trees and Derivations (20 pts)

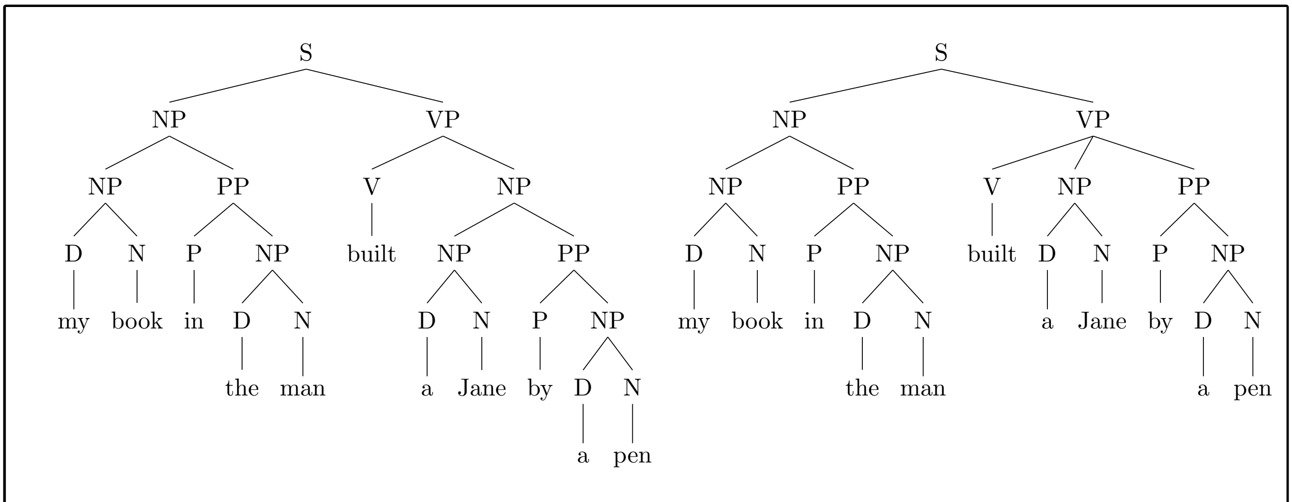
Given the CFG below, provide parse trees for given sentences in **a** and **b**.

```
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)

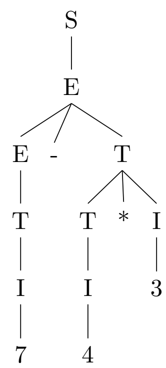


Given the CFG below, answer **c**, **d** and **e**

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

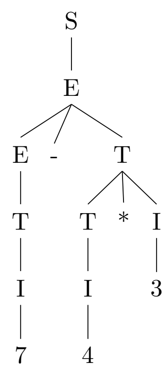
c) Provide the left-most derivation of 7 - 4 * 3 step-by-step and plot the final parse tree matching that derivation (4/20 pts)

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



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)

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

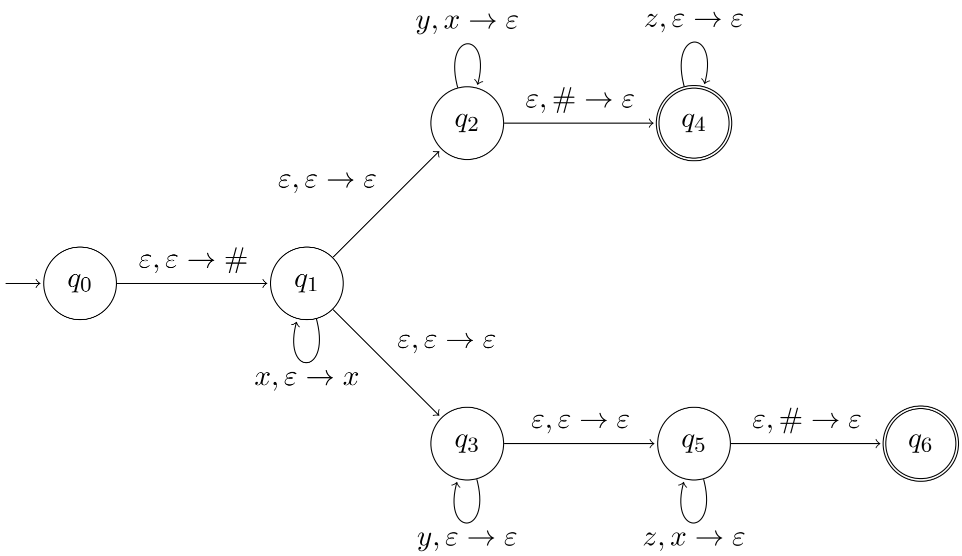


e) Are the derivations in **c** and **d** in the same similarity class? (4/20 pts)

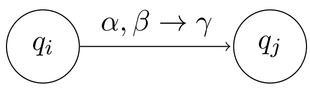
Yes, since (D_L, D_R) belongs in the reflexive, symmetric, transitive closure of $\prec$.

3 Pushdown Automata (30 pts)

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

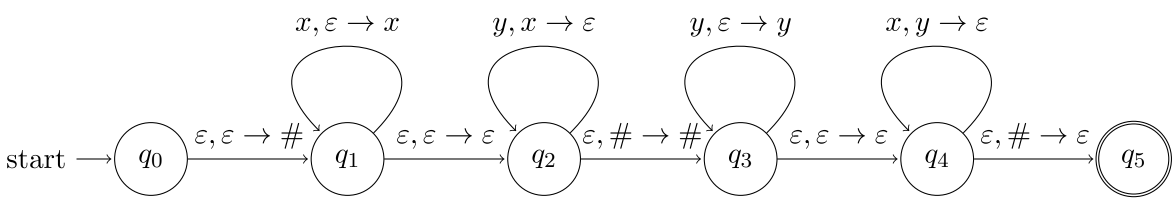


where the transition $((q_i, \alpha, \beta), (q_j, \gamma))$ is represented as:



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

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.

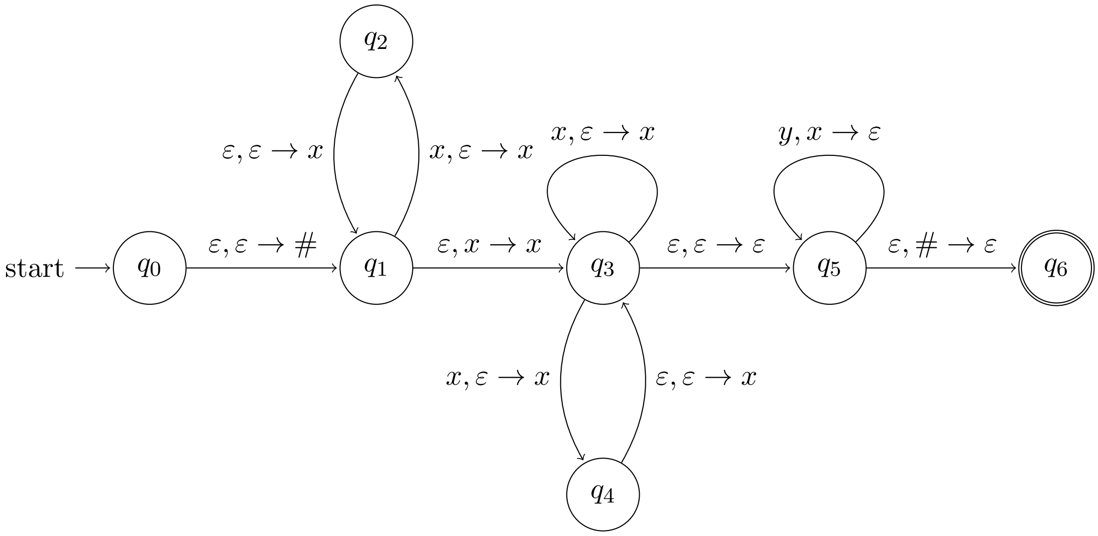


Table for xxy,

State	Unread Input	Stack	Transition Used
q_0	xxy	ε	$-$
q_1	xxy	$\#$	$((q_0, \varepsilon, \varepsilon), (q_1, \#))$
q_2	xy	$x\#$	$((q_1, x, \varepsilon), (q_2, x))$
q_1	xy	$xx\#$	$((q_2, \varepsilon, \varepsilon), (q_1, x))$
q_3	xy	$xx\#$	$((q_1, \varepsilon, x), (q_3, x))$
q_3	y	$xxx\#$	$((q_3, x, \varepsilon), (q_3, x))$
q_5	y	$xxx\#$	$((q_3, \varepsilon, \varepsilon), (q_5, \varepsilon))$
q_5	ε	$xx\#$	$((q_5, y, x), (q_5, \varepsilon))$

Table for $xxyyyy$,

State	Unread Input	Stack	Transition Used
q_0	$xxyyyy$	ε	$-$
q_1	$xxyyyy$	$\#$	$((q_0, \varepsilon, \varepsilon), (q_1, \#))$
q_2	$xyyyy$	$x\#$	$((q_1, x, \varepsilon), (q_2, x))$
q_1	$xyyyy$	$xx\#$	$((q_2, \varepsilon, \varepsilon), (q_1, x))$
q_3	$xyyyy$	$xx\#$	$((q_1, \varepsilon, x), (q_3, x))$
q_4	$yyyy$	$xxx\#$	$((q_3, x, \varepsilon), (q_4, x))$
q_3	$yyyy$	$xxxx\#$	$((q_4, \varepsilon, \varepsilon), (q_3, x))$
q_5	$yyyy$	$xxxx\#$	$((q_3, \varepsilon, \varepsilon), (q_5, \varepsilon))$
q_5	yyy	$xxx\#$	$((q_5, y, x), (q_5, \varepsilon))$
q_5	yy	$xx\#$	$((q_5, y, x), (q_5, \varepsilon))$
q_5	y	$x\#$	$((q_5, y, x), (q_5, \varepsilon))$
q_5	ε	$\#$	$((q_5, y, x), (q_5, \varepsilon))$
q_6	ε	ε	$((q_5, \varepsilon, \#), (q_6, \varepsilon))$

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.

Say alphabet of the language L is Σ and $L'' = \{w \in \Sigma^* \mid |w| = 4n + 2 \text{ for } n \in \mathbb{N}\}$. Then the language L' can be interpreted as $L' = L \cap L''$. Note that L'' is a regular language so by the Theorem 3.5.2 (Elements of Theory of Computation, p.144), we know that L' is a CFL. Now, we will formally construct an automaton for L' to prove that it is, indeed, a CFL. Name the PDA that recognizes L as $M_1 = (K_1, \Sigma, \Gamma_1, \Delta_1, s_1, F_1)$ and the DFA that recognizes L'' as $M_2 = (K_2, \Sigma, \delta, s_2, F_2)$. The PDA that recognizes L' is defined $M = (K, \Sigma, \Gamma, \Delta, s, F)$ where,

- $K = K_1 \times K_2$,
- $\Gamma = \Gamma_1$,
- $s = (s_1, s_2)$,
- $F = F_1 \times F_2$, and
- Δ is defined as follows, for each $((q_1, a, \beta), (q_1', \gamma)) \in \Delta_1$ and for each $q_2 \in K_2$, add $(((q_1, q_2), a, \beta), ((q_1', \delta(q_2, a)), \gamma))$ to Δ and for each $((q_1, e, \beta), (q_1', \gamma)) \in \Delta_1$ and for each $q_2 \in K_2$, add $(((q_1, q_2), e, \beta), ((q_1', q_2), \gamma))$ to Δ.

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)

Example for L_4 is not a CFL,
$L_1 = \{a^n b^n c^m \mid n, m \geq 0,\ n, m \in \mathbb{N}\}$
$L_2 = \{a^n b^m c^m \mid n, m \geq 0,\ n, m \in \mathbb{N}\}$
$L_3 = \{\}$
$L_4 = L_1 \cap (L_2 \setminus L_3) = \{a^n b^n c^n \mid n \geq 0,\ n \in \mathbb{N}\}$

Example for L_4 is a CFL,
$L_1 = \{a^n b^n c^m \mid n, m \geq 0,\ n, m \in \mathbb{N}\}$
$L_2 = \{a^n b^n \mid n \geq 0,\ n \in \mathbb{N}\}$
$L_3 = \{\}$
$L_4 = L_1 \cap (L_2 \setminus L_3) = \{a^n b^n \mid n \geq 0,\ n \in \mathbb{N}\}$

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

By the Theorem 3.5.2 (Elements of Theory of Computation, p.144), the intersection of a CFL and a regular language is a CFL. Thus, $L_1 \cap L_3$ is a CFL. By the Theorem 3.5.1 (Elements of Theory of Computation, p.143), the CFLs are closed under union, concatenation, and Kleene star. Hence, L_5 is a CFL.

5 Pumping Theorem (20 pts)

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

Assume that L is a CFL and let K be the pumping length. Then, by the Pumping Theorem, for the string $w = a^K m^K t^K \in L$ there must be a split $uvxyz$ such that $|vy| \geq 1$, $|vxy| \leq K$, and $uv^i xy^i z \in L$ for every $i \geq 0$. There are five cases to consider, listed as follows,

1. $vxy = a^j$ for some $1 \leq j \leq K$. vy consists of a symbols, pumping i any value other than $i = 1$ would result into a string that is not in L, since number of a symbols will be different than K whereas number of m symbols will stay as K.

2. $vxy = m^j$ for some $1 \leq j \leq K$. vy consists of m symbols, pumping i any value other than $i = 1$ would result into a string that is not in L, since number of m symbols will be different than K whereas number of a symbols will stay as K.

3. $vxy = t^j$ for some $1 \leq j \leq K$. vy consists of t symbols, pumping i to $i = 3K$ would result into a string that is not in L. Since vy contains at least 1 and at most K t symbols, after pumping w with $i = 3K$, number of t symbols would be at least $3K > 2K$ and at most $4K - 1 > 2K$ which results into a string that is not in L.

4. $vxy = a^j m^p$ for some $1 \leq j + p \leq K$. vy consists of a and m symbols. Firstly, consider the case that v contains a symbols and y contains m symbols. In this case if $|v| = |y|$, then after pumping i to $i = 2K$, the string will become at least $a^{2K} m^{2K} t^K$ and at most $a^{3K-1} m^{3K-1} t^K$ which is not in L. If $|v| \neq |y|$, then after pumping i any value other than $i = 1$ would result into a string that has unequal number of a and m symbols which is not in L. Secondly, consider either v contains both a and m symbols or y contains both a and m symbols, then in this case pumping i any value other than $i = 1$ would break the order of symbols in the string and result into a string that is not in L.

5. $vxy = m^j t^p$ for some $1 \leq j + p \leq K$. vy consists of m and t symbols. Firstly, consider the case that v contains m symbols and y contains t symbols. In this case pumping i any value other than $i = 1$ would result into a string that has unequal number of a and m symbols which is not in L. Secondly, consider either v contains both m and t symbols or y contains both m and t symbols, then in this case pumping i any value other than $i = 1$ would break the order of symbols in the string and result into a string that is not in L.

Hence, in all cases it has been shown that pumping i results into a string that is not in the language, this contradicts with the assumption. Assumption is discharged, L is not a CFL.

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 L is a CFL and let K be the pumping length. Then, by the Pumping Theorem, for the string $w = a^K b^{2K} a^K \in L$ there must be a split $uvxyz$ such that $|vy| \geq 1$, $|vxy| \leq K$ and $uv^ixy^iz \in L$ for every $i \geq 0$. There are four cases to consider, listed as follows,

1. $vxy = a^j$ for some $1 \leq j \leq K$. vy consists of a symbols coming from the first part of the string, pumping i any value other than $i = 1$ would result into a string that is not in L, since number of a symbols in the first part will be different than K whereas number of b symbols will stay as $2K$ and number of a symbols at the end of the string will stay as K. The same argument goes for picking vxy from the end of the string.
2. $vxy = b^j$ for some $1 \leq j \leq K$. vy consists of b symbols, pumping i any value other than $i = 1$ would result into a string that is not in L, since number of b symbols will be different than $2K$ whereas number of a symbols at the beginning will stay as K and number of a symbols at the end will also stay as K.
3. $vxy = a^j b^p$ for some $1 \leq j + p \leq K$. vy consists of a and b symbols respectively, pumping i any value other than $i = 1$ would result into a string that is not in L, since either number of a and b symbols will be different than K and $2K$, respectively, whereas number of a symbols at the end will stay as K or the order of symbols in the string will be broken.
4. $vxy = b^j a^p$ for some $1 \leq j + p \leq K$. vy consists of b and a symbols respectively, pumping i any value other than $i = 1$ would result into a string that is not in L, since either number of b and a symbols will be different than $2K$ and K, respectively, whereas number of a symbols at the beginning will stay as K or the order of symbols in the string will be broken.

Hence, in all cases it has been shown that pumping i results into a string that is not in the language, this contradicts with the assumption. Assumption is discharged, L is not a CFL.

6 CNF and CYK (not graded)

a) Convert the given context-free grammar to Chomsky Normal Form.

$S \rightarrow XSX \mid xY$
$X \rightarrow Y \mid S$
$Y \rightarrow z \mid \varepsilon$

answer here ...

b) Use the grammar below to parse the given sentence using Cocke–Younger–Kasami algorithm. Plot the parse trees.

S → NP VP
S → X1 VP
X1 → Aux NP
S → book | include | prefer
S → Verb NP
S → X2 PP
S → Verb PP
S → VP PP
NP → I | she | me | Houston
NP → Det Nom
Nom → book | flight | meal | money
Nom → Nom Noun
Nom → Nom PP

VP → book | include | prefer
VP → Verb NP
VP → X2 PP
X2 → Verb NP
VP → Verb PP
VP → VP PP
PP → Prep NP
Det → that | this | the | a
Noun → book | flight | meal | money
Verb → book | include | prefer
Aux → does
Prep → from | to | on | near | through

book the flight through Houston

Empty parse table:

1:5 → 1:1 2:5 1:5 → 1:2 3:5 1:5 → 1:3 4:5 1:5 → 1:4 5:5				
1:4 → 1:1 2:4 1:4 → 1:2 3:4 1:4 → 1:3 4:4	2:5 → 2:2 3:5 2:5 → 2:3 4:5 2:5 → 2:4 5:5			
1:3 → 1:1 2:3 1:3 → 1:2 3:3	2:4 → 2:2 3:4 2:4 → 2:3 4:4	3:5 → 3:3 4:5 3:5 → 3:4 5:5		
1:2 → 1:1 2:2	2:3 → 2:2 3:3	3:4 → 3:3 4:4	4:5 → 4:4 5:5	
1:1	2:2	3:3	4:4	5:5
book	the	flight	through	Houston

rest of the answer here ...

7 Deterministic Pushdown Automata (not graded)

Provide a DPDA to recognize the given languages, the DPDA must read its entire input and finish with an empty stack.

a) $a^*bc \cup a^n b^n c$

answer here ...

b) $(aa)^*c \cup a^n b^n c$

answer here ...