Middle East Technical University Department of Computer Engineering

Formal Languages and Abstract Machines
Take Home Exam 2

Adil Kaan Akan
2171155

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 → aTaa | aTab | bTba | bTbb
T → Ta | Tb | e

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

S → aSa | aSb | bSa | bSb | a | 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)

T → S | e
S → Taab | aTab | aaTb | aabT | Taba | aTba | abTa | abaT | Tbaa | bTaa | baTa | baaT

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

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

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

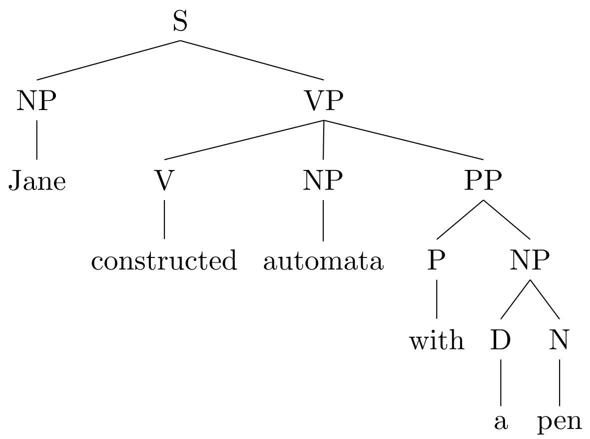
$$\boxed{\text{L(G)} = a^+ \cup b^+ \cup aa^+b \cup ab^+b \cup a^+a^+b^+ \cup a^+b^+b^+ \cup \{a,b\}^*ba\{a,b\}^*}$$

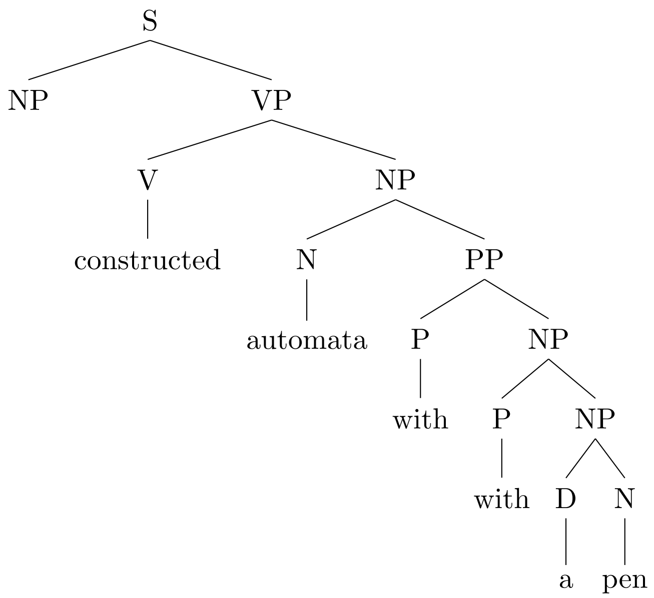
2 Parse Trees and Derivations (20 pts)

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

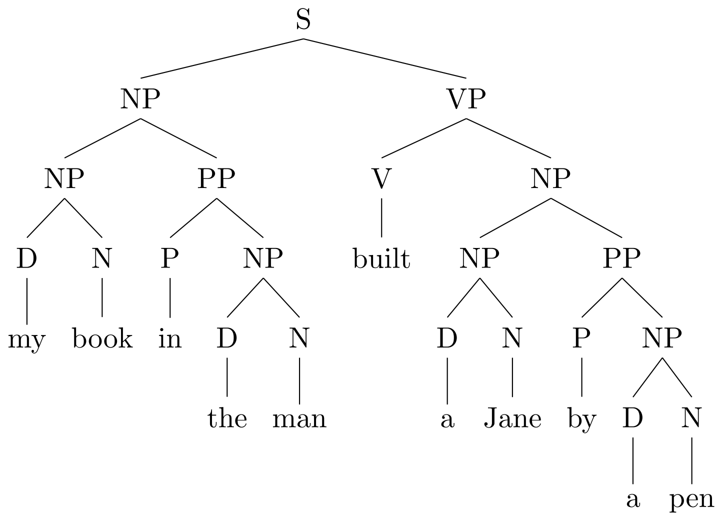
```
S   → NP VP
VP  → V NP | V NP PP
PP  → P NP
NP  → N | D N | NP PP
V   → wrote | built | constructed
D   → a | an | the | my
N   → John | Mary | Jane | man | book | automata | pen | class
P   → in | on | by | with
```

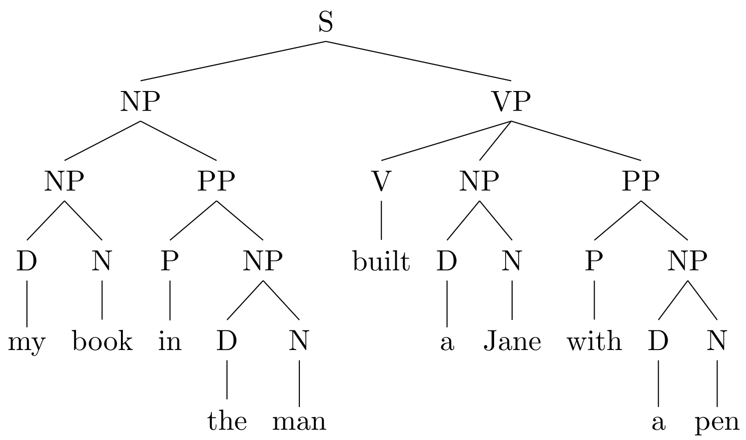
a) Jane constructed automata with a pen (4/20 pts)





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





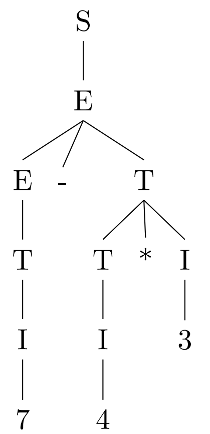
Given the CFG below, answer **c**, **d** and **e**

```
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)

The arrows means that the left most derivation. (I cannot put the L letter on to the top of the arrow.)

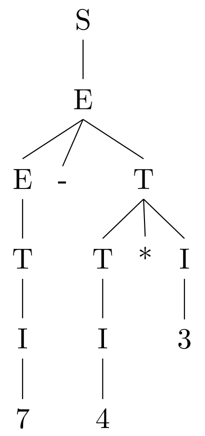
$S \to E \to E - T \to T - T \to I - T \to 7 - T \to 7 - T * I \to 7 - I * I \to 7 - 4 * I \to 7 - 4 * 3$



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)

The arrows means that the right most derivation. (I cannot put the R letter on to the top of the arrow.)

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

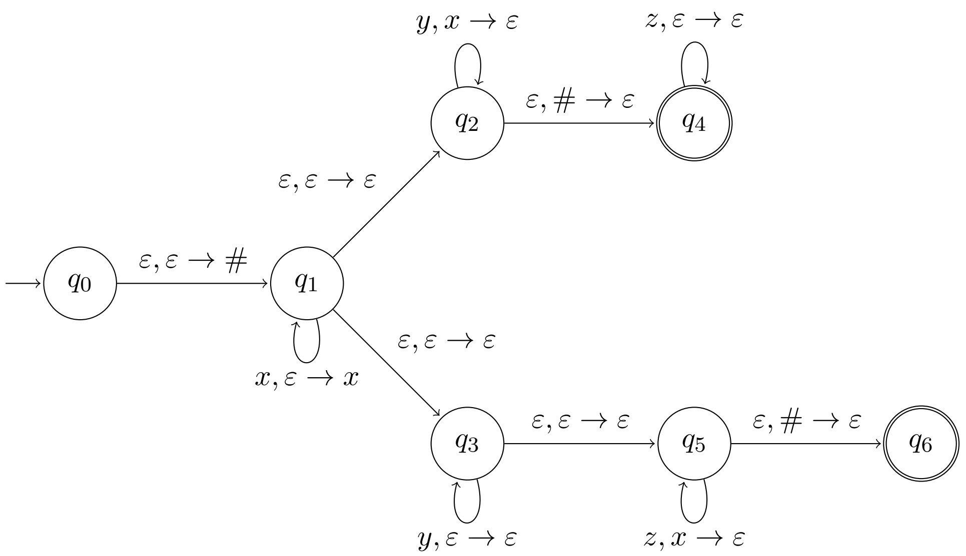


e) Are the derivations in **c** and **d** in the same similarity class? (4/20 pts)

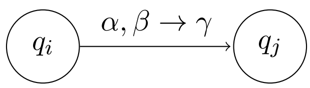
Since the parse trees at the end of the derivations are same, we can say that the both of the parse trees are in the same similarity class. The parse trees are same because we change only the preference of the rules of the derivation and rules of the derivation are same in the both of the derivation of the parse trees.

3 Pushdown Automata (30 pts)

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

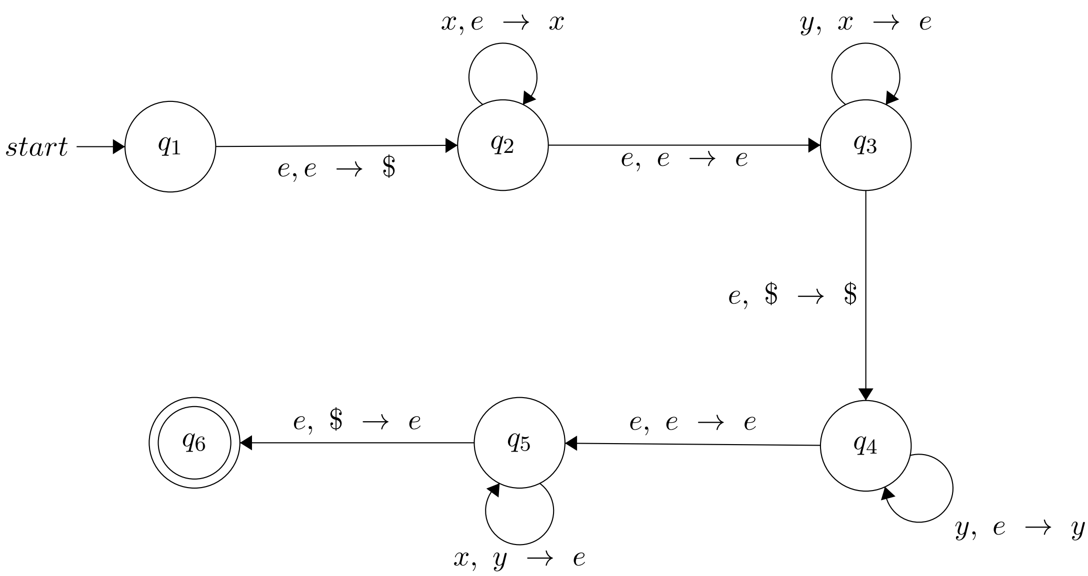


where the transition $((q_i, \alpha, \beta), (q_j, \gamma))$ is represented as:



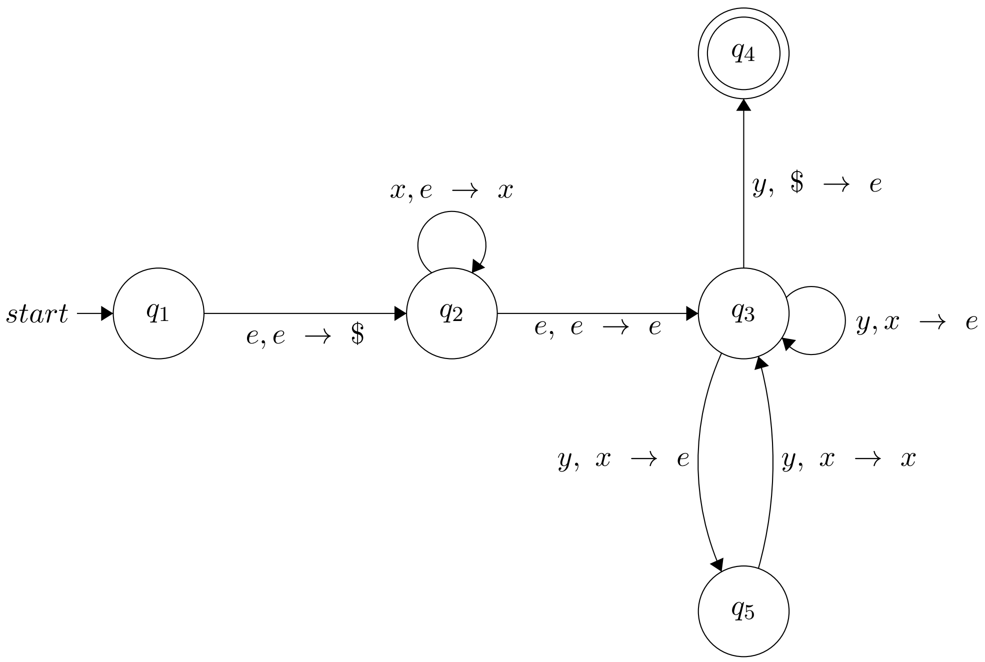
$\{x^i y^j z^k \mid$ i,j,k ≥ 0 and i = j or i=k $\}$

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.

CFG: $S \to xTyy$
$T \to xTy | xTyy | e$
PDA:



Simulation of the string xxy

Transition	Current State	Unread Input	Stack contents
1	q_1	xxy	e
2	q_2	xxy	\$
3	q_2	xy	x\$
4	q_2	y	xx\$
5	q_3	y	xx\$
6	q_3	e	x\$

Since at the end the stack is not empty and the state q_3 is not a accepting state, the string xxy will not be accepted.

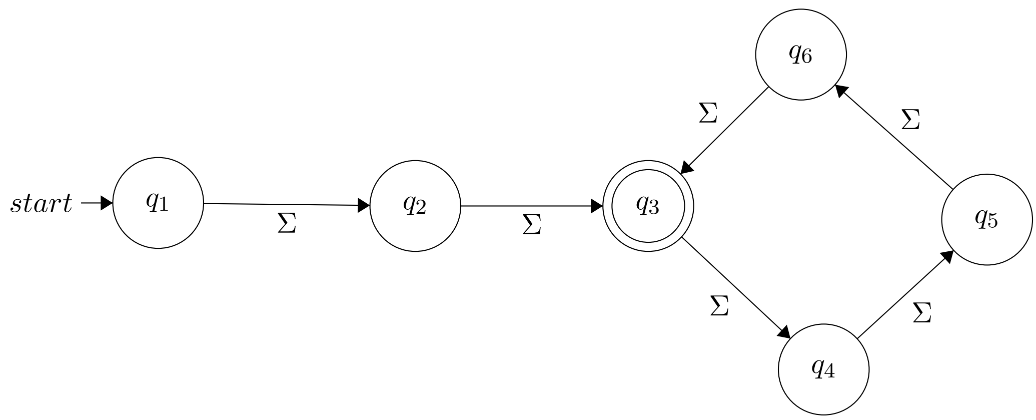
Simulation of the string xxyyyy

Transition	Current State	Unread Input	Stack contents
1	q_1	xxyyyy	e
2	q_2	xxyyyy	\$
3	q_2	xyyyy	x\$
4	q_2	yyyy	xx\$
5	q_3	yyyy	xx\$
6	q_5	yyy	x\$
7	q_3	yy	x\$
8	q_3	y	\$
9	q_4	e	e

Since at the end the stack is empty and q_4 is accepting state, the string xxyyyy will be accepted.

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.

Denote L as L $= \{K_L, \Sigma, \Gamma_L, \sigma_L, s_L, F_L\}$ where F is set of accepting states of the PDA L.
Define a DFA,say L_0, which recognizes the language $|w| = 4n + 2 \ for \ n \in \mathbb{N}$ as follows



Then, we can say that $|w| = 4n + 2 \ for \ n \in \mathbb{N}$,say L_0, is regular language. Since L' is intersection of the L and L_0, it is context free language. I will give the proof by using book's proof. We have two machines one is PDA L=$(K_L, \Sigma, \Gamma_L, \sigma_L, s_L, F_L)$ and the other one is DFA $L_0 = (K_0, \Sigma, \sigma_0, s_0, F_0)$. The idea is combining these two automatons into single pushdown automaton M that carries out computations by L and L_0 in parallel and accepts only if both of the automatons would accepted. We can specify M $= (K, \Sigma, \Gamma, \sigma, s, F)$ where
$K = K_L x K_0$
$\Gamma = \Gamma_L$

$s = (s_L x s_0)$
$F = F_L x F_0$

σ, the transition function of the automaton M, is defined as follows. For each transition of the pushdown automaton L $((q_1, a, \beta), (p_1, \gamma)) \in \sigma_L$, and for each state $q_2 \in K_0$, we add to σ the transition $(((q_1, q_2)a, \beta), ((p_1, \sigma_0(q_2, a)), \gamma))$; and for each transtion of the form $((q_1, e, \beta), (p_1, \gamma)) \in \sigma_L$ and each state $q_2 \in K_0$, we add the transition $(((q_1, q_2), e, \beta), ((q_1, q_2), \gamma))$. That means that M passes from state (q_1, q_2) to state (p_1, p_2) in the same way that M_1 passes from state q_1 to p_1, except that in addition M keeps track of the change in the state of L_0 caused by reading the same input.

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)

We can write $(L_2 \setminus L_3)$ as $L_2 \cap \overline{L_3}$ and $L_2 \cap \overline{L_3}$ is context free because $\overline{L_3}$ is regular since regular languages are closed under complementation and intersection of a regular language and context free language is context free by the theorem (3.5.2). Then, since both of the languages that L_1 and $L_2 \cap \overline{L_3}$ are both context free and context free language are not closed under the intersection with each other, L_4 is not necessarily context free language.
For example, take $L_1 = \{a^n b^n c^m | n \geq 0 and m \geq 0\}$ and $L_2 = \{a^m b^n c^n | n \geq 0 and m \geq 0\}$. Then, $L_3 = L_1 \cap L_2 = \{a^n b^n c^n | n \geq 0\}$ is not context free because L_1 says that the number of a's should be equal to the number of b's and L_2 says that the number b's should be equal to the number of c's. However, in their intersections L_3, both conditions need not to be true for satisfying the property of L_3. But a pushdown automata can compare only two. Therefore, L_3 is not context free language since it cannot be recognized by a pushdown automata.

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

The intersection of a context free language and regular language is a context free language. It is clear by the theorem from the book whose number is (3.5.2). Since $(L_1 \cap L_3)$ is context free language clarified with theorem 3.5.2 and every context free language are closed under Kleene star operation by theorem from the book whose number is (3.5.1), we can say that L_5 is context free language.

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 using Pumping Theorem for CFLs. (10/20 pts)

We state by using pumping lemma that for every context free language, there is an integer called pumping length such that for every string which is element of L, context free language, the length of the string is greater than or equal to the pumping length n and we can parse that string,say s, as s = uvxyz where $|vy| \geq 0$ which means that both of the v and y subtrings cannot be empty and $|vxy| \leq n$ and for every positive integer, say i=0,1,2,3,..., the string uv^ixy^iz must be in the language.
Take the string $a^nb^nt^n$. Since the property of the language is satisfied the string is in the language L.Since the length of the substring vxy is less than or equal to the pumping length n and and between a and t there are m letters of n times, it cannot contain more than two symbols. If we analyze the cases, there are four cases as follows:
a)vxy contains only m's, it cannot be in the language since it contains more m's than a's.
b)vxy contains only a's, it cannot be in the language since it contains more a's than m's.
c)vxy contains m's and t's, it cannot be in the language since it contains more m's than a's.
d)vxy contains a's and m's, it cannot be in the language since it contains less t's than a's and m's.

Since we examine all of the cases and one of the cases must be true eventually, the string cannot be in the language in any case. Therefore, the language $L = \{a^n m^n t^i \mid n \leq i \leq 2n\}$ is not context free language.

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)

We state by using pumping lemma that for every context free language, there is an integer called pumping length such that for every string which is element of L, context free language, the length of the string is greater than or equal to the pumping length n and we can parse that string,say s, as s = uvxyz where $|vy| \geq 0$ which means that both of the v and y subtrings cannot be empty and $|vxy| \leq n$ and for every positive integer, say i=0,1,2,3,..., the string uv^ixy^iz must be in the language.
Take the string $a^nb^{2n}a^n$. Since the property of the language is satisfied the string is in the language L. Since the length of the substring vxy is less than or equal to the pumping length n, vxy substring can only contain a's or b's or a's before b's or b's before a's.(a^p,b^{2p},a^p not from all three). If we analyze the cases, there are five cases as follows:
a)vxy contains only b's, it cannot be in the language since it does not contain b's exactly two times of a's from the first part of the string and a's from the last part of the string.
b)vxy contains only a's from the first part of the substring, it cannot be in the language since it contains more a's in the first part of the substring than last part of the substring.
c)vxy contains only a's from the last part of the substring, it cannot be in the language since it contains more a's in the last part of the substring than first part of the substring.
d)vxy contains b's and a's from the last part of the substring, it cannot be in the language

since it cannot be in the language since it contains more a's in the last part of the substring than first part of the substring.
e)vxy contains a's from the first part of the substring and b's , it cannot be in the language since it cannot be in the language since it contains more a's in the first part of the substring than last part of the substring.

Since we examine all of the cases and one of the cases must be true eventually, the string cannot be in the language in any case. Therefore, the language $L = \{a^n b^{2n} a^n \mid n \in \mathbb{N}+\}$ is not context free language.

6 CNF and CYK (not graded)

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

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

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) $(aa)^*c \cup a^nb^nc$

answer here ...