Middle East Technical University Department of Computer Engineering

CENG 280

Formal Languages and Abstract Machines

Spring 2017-2018

Take Home Exam 2

Due date: April 22^{nd}, 23:55

Objectives

To familiarize with Context Free Languages, grammars for CFL and Pushdown Automata, parse trees and derivations, closure properties of CFL, Pumping Lemma for CFL, Chomsky Normal Form and Cocke-Younger-Kasami Algorithm for parsing, Deterministic PDA.

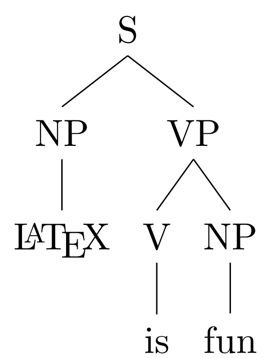
Specifications

- You must adhere to the notation conventions adopted in the textbook.
- Your solution should be delivered as a .tex file based on your modification of the provided template file. For convenience, a simple code for drawing a tree is included in the following. On the left-hand side you can see the code segment, and generated tree is placed on the right. You can also use the automata template given in THE1.

```
% preamble
\usepackage{tikz}
\usepackage{tikz-qtree}

% document
% use qtree
\Tree [.S [.NP $$\LaTeX$$ ] [.VP [.V is ] [.NP fun ] ] ]

% or tikz-qtree with possible tikz options
\begin{tikzpicture}[scale=1]
\Tree [.S [.NP $$\LaTeX$$ ] [.VP [.V is ] [.NP fun ] ] ]
\end{tikzpicture}
```



- The questions and submission regulations are included in subsequent sections. While designing your solutions to the tasks, explicitly state any assumptions you make and pay particular attention to the notation you use. Your proofs must be sound and complete. Grading will be heavily affected by the formalization of your solutions.

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$\}$

$S \rightarrow aXaa \mid aXab \mid bXba \mid bXbb$
$X \rightarrow aX \mid bX \mid \varepsilon$

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

$S \rightarrow a \mid b \mid aaS \mid abS \mid baS \mid bbS$

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

$S \rightarrow SaSaSbS \mid SaSbSaS \mid SbSaSaS \mid \varepsilon$

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

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

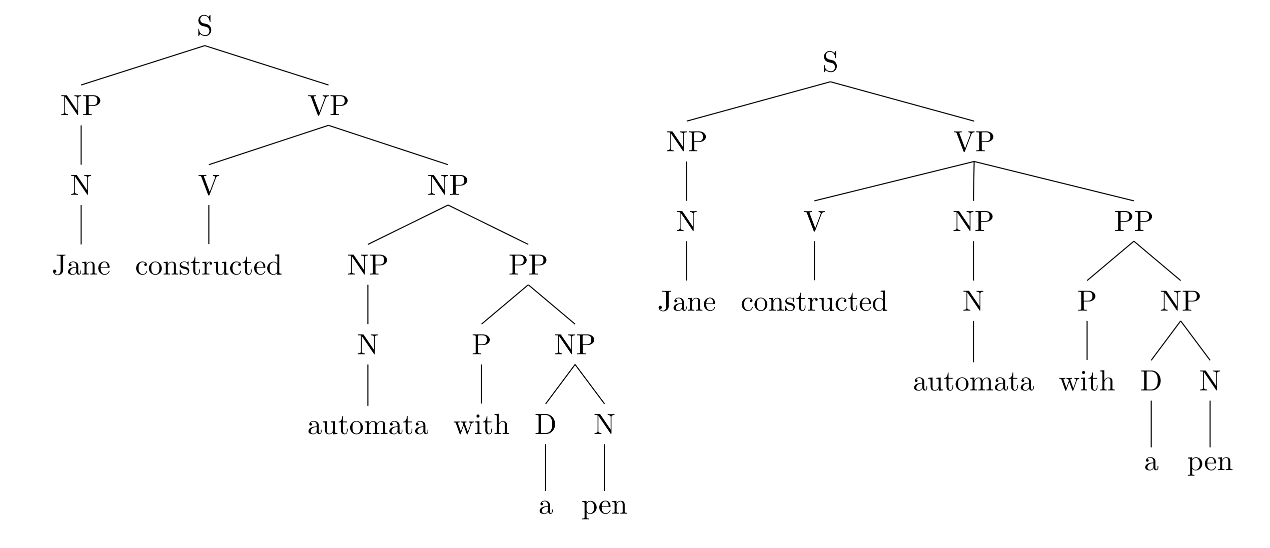
$L(G) = \Sigma^* - \{a^n b^n \mid \ n \geq 0\}$

2 Parse Trees and Derivations (20 pts)

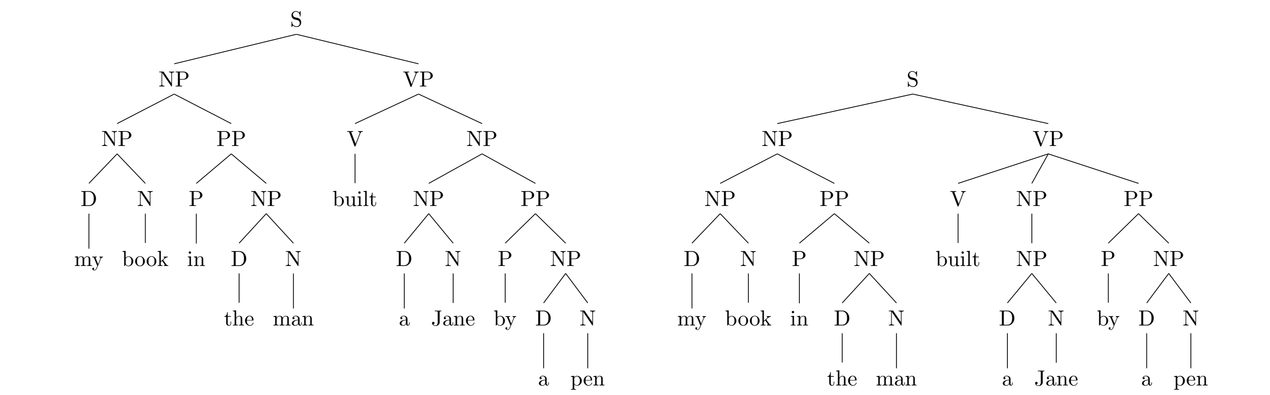
Given the CFG below, provide parse trees for given sentences in **a** and **b**.

```
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)

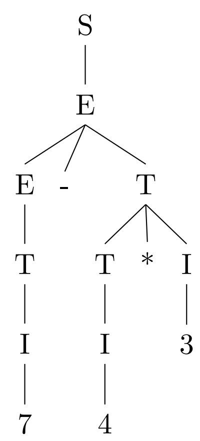


Given the CFG below, answer **c**, **d** and **e**

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

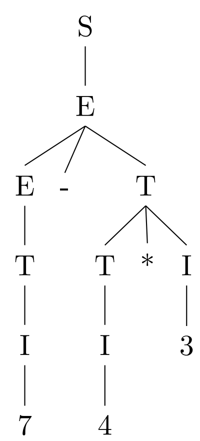
c) Provide the left-most derivation of 7 - 4 * 3 step-by-step and plot the final parse tree matching that derivation (4/20 pts)

$$
\begin{aligned}
S &\Rightarrow E \\
&\Rightarrow E - T \\
&\Rightarrow T - T \\
&\Rightarrow I - T \\
&\Rightarrow 7 - T \\
&\Rightarrow 7 - T * I \\
&\Rightarrow 7 - I * I \\
&\Rightarrow 7 - 4 * I \\
&\Rightarrow 7 - 4 * 3
\end{aligned}
$$



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)

$$
\begin{aligned}
S &\Rightarrow E \\
&\Rightarrow E - T \\
&\Rightarrow E - T * I \\
&\Rightarrow E - T * 3 \\
&\Rightarrow E - I * 3 \\
&\Rightarrow E - 4 * 3 \\
&\Rightarrow T - 4 * 3 \\
&\Rightarrow I - 4 * 3 \\
&\Rightarrow 7 - 4 * 3
\end{aligned}
$$

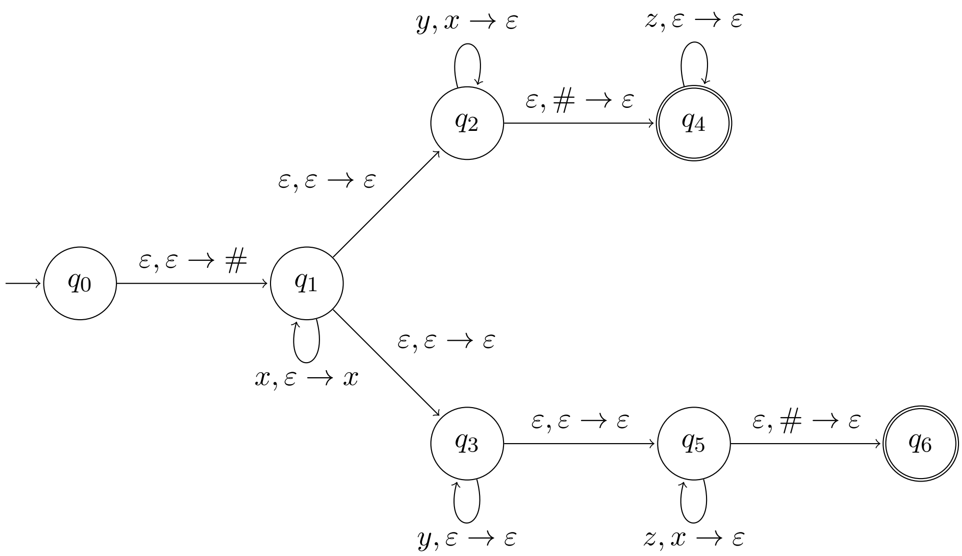


e) Are the derivations in **c** and **d** in the same similarity class? (4/20 pts)

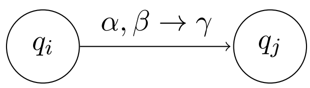
Yes, the same applications are used at the same positions, they only differ at the relative order of rule applications.

3 Pushdown Automata (30 pts)

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

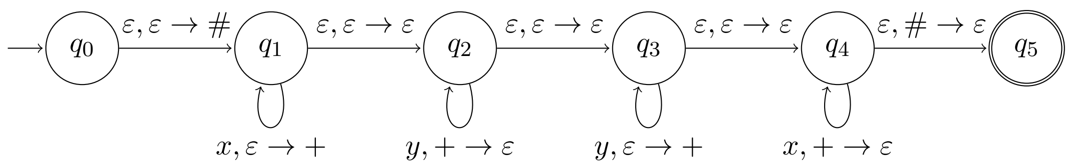


where the transition $((q_i, \alpha, \beta), (q_j, \gamma))$ is represented as:



$L = \{x^i y^j z^k \mid\ i = j \text{ or } i = k;\ i, j, k \in \mathbb{N}\}$

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.

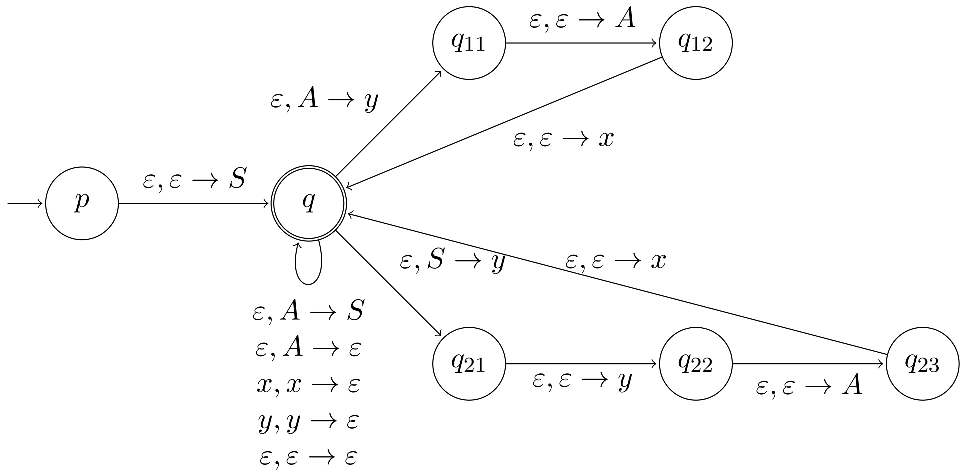
A grammar $G = (V, \Sigma, R, S)$ for the language:

$V = (S, A, x, y)$

$\Sigma = (x, y)$

$R = (S \to xAyy,\ A \to xAy \mid S \mid \varepsilon)$

A PDA constructed using the proof of lemma 3.4.1:



simulation on **xxy**

state	input	stack	transition
p	xxy		$\varepsilon, \varepsilon \to S$
q	xxy	S	$\varepsilon, S \to y$
q_{21}	xxy	y	$\varepsilon, S \to y$
q_{22}	xxy	yy	$\varepsilon, \varepsilon \to A$
q_{23}	xxy	Ayy	$\varepsilon, \varepsilon \to x$
q	xy	xAyy	$x, x \to \varepsilon$
q	xy	Ayy	$\varepsilon, A \to y$
q_{11}	xy	yyy	$\varepsilon, \varepsilon \to A$
q_{12}	xy	Ayyy	$\varepsilon, \varepsilon \to x$
q	xy	xAyy	$x, x \to \varepsilon$
q	y	Ayyy	$\varepsilon, A \to \varepsilon$
q	y	yyy	$y, y \to \varepsilon$
q		yy	rejected

simulation on **xxyyyy**

state	input	stack	transition
p	xxyyyy		$\varepsilon, \varepsilon \to S$
q	xxyyyy	S	$\varepsilon, S \to y$
q_{21}	xxyyyy	y	$\varepsilon, S \to y$
q_{22}	xxyyyy	yy	$\varepsilon, \varepsilon \to A$
q_{23}	xxyyyy	Ayy	$\varepsilon, \varepsilon \to x$
q	xyyyy	xAyy	$x, x \to \varepsilon$
q	xyyyy	Ayy	$\varepsilon, A \to S$
q	xyyyy	Syy	$\varepsilon, S \to y$
q_{21}	xyyyy	yyyy	$\varepsilon, S \to y$
q_{22}	xyyyy	yyyy	$\varepsilon, \varepsilon \to A$
q_{23}	xyyyy	Ayyyy	$\varepsilon, \varepsilon \to x$
q	xyyyy	xAyyyy	$x, x \to \varepsilon$
q	yyyy	Ayyyy	$\varepsilon, A \to \varepsilon$
q	yyyy	yyyy	$y, y \to \varepsilon$
q	yyy	yyy	$y, y \to \varepsilon$
q	yy	yy	$y, y \to \varepsilon$
q	y	y	$y, y \to \varepsilon$
q			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.

> Given a PDA $M = (Q, \Sigma, \Gamma, \Delta, s, F)$ for the language L,
> We can construct a PDA $M' = (Q', \Sigma', \Gamma', \Delta', s', F')$ for L' as follows:
>
> $\Sigma' = \Sigma$
> $\Gamma' = \Gamma$
> $Q' = \{(p, q) \mid \ p \in Q \ and \ q \in \{q_0, q_1, q_2, q_3\}\}$
> $s' = (s, q_0)$
> $F' = F \times \{q_2\}$
> $\Delta'((p, q_i), a, b) = ((r, q_j), c)$ if $\Delta(p, a, b) = (r, c)$ where
> $p, r \in Q$, $a \in \Sigma$, $b, c \in \Gamma$ and
> (q_i, q_j) is one of (q_0, q_1), (q_1, q_2), (q_2, q_3) or (q_3, q_0).

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 a CFL.

An example:
$L_1 = L_2 = \{a^n b^n \mid n \geq 0\}$ are non-regular CFLs. $L_3 = \emptyset$ is regular. In this case, $L_4 = L_1 \cap (L_1 \setminus \emptyset) = L_1 \cap L_1 = L_1 = \{a^n b^n \mid n \geq 0\}$ is a CFL.

A counter example:
$L_1 = \{a^n b^n c^m \mid n \geq 0, m \geq 0\}$ and $L_2 = \{a^n b^m c^m \mid n \geq 0, m \geq 0\}$ are non-regular CFLs. $L_3 = \emptyset$ is regular. In this case, $L_4 = L_1 \cap (L_2 \setminus \emptyset) = L_1 \cap L_2 = \{a^n b^n c^n \mid n \geq 0\}$ is not a CFL.

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

Since the intersection of a CFL and a regular language is a CFL, we know that $L_1 \cap L_3$ is a CFL. Since the Kleene star of a CFL is also a CFL, $L_5 = (L_1 \cap L_3)^*$ 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 using Pumping Theorem for CFLs. (10/20 pts)

Assume L is a context-free language. Let p be the pumping length.
The string $s = a^p m^p t^p$ is in L and the length of the string $|s| > p$.
The pumping lemma guarantees that the s can be divided into 5 as $s = uvxyz$ satisfying these 3 conditions below:

- $\forall i \geq 0,\ uv^ixy^iz \in L$
- $|vy| > 0 \implies$ at least one of v or y is not empty
- $|vxy| \leq p$

However;

1. If v or y contains more than one symbol each, any string uv^ixy^iz with $i > 1$ contains symbols in incorrect order.
2. If v and y contains only one symbol, whatever those symbols are, uv^ixy^iz (when $i \neq 1$) cannot contain equal numbers for all three of symbols.

Therefore, L is not a context-free language. $\square$

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 L is a context-free language. Let p be the pumping length.
The string $s = a^p b^{2p} a^p$ is in L and the length of the string $|s| > p$.
The pumping lemma guarantees that the s can be divided into 5 as $s = uvxyz$ satisfying these 3 conditions below:

- $\forall i \geq 0,\ uv^ixy^iz \in L$
- $|vy| > 0 \implies$ at least one of v or y is not empty
- $|vxy| \leq p$

However;

1. If v or y contains more than one symbol each, any string uv^ixy^iz with $i > 1$ contains symbols in incorrect order.
2. If v and y contains b's pumping these results in more b's than necessary.
3. If v contains a's and y contains b's or vice versa, pumping these cannot be balanced by appropriate number of b's versus a's.

Therefore, L is not a context-free language. $\square$

6 CNF and CYK (not graded)

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

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

1. Add a new start variable

$\boldsymbol{S_0 \to S}$
$S \to XSX \mid xY$
$X \to Y \mid S$
$Y \to z \mid \varepsilon$

2. Remove ε productions ($A \to \varepsilon$)

(a) Eliminate $Y \to \varepsilon$
$S_0 \to S$
$S \to XSX \mid xY \mid \boldsymbol{x}$
$X \to Y \mid S \mid \boldsymbol{\varepsilon}$
$Y \to z$

(b) Eliminate $X \to \varepsilon$
$S_0 \to S$
$S \to XSX \mid xY \mid x \mid \boldsymbol{S} \mid \boldsymbol{XS} \mid \boldsymbol{SX}$
$X \to Y \mid S$
$Y \to z$

3. Remove unit productions ($A \to B$)

(a) Eliminate $S \to S$
$S_0 \to S$
$S \to XSX \mid xY \mid x \mid XS \mid SX$
$X \to Y \mid S$
$Y \to z$

(b) Eliminate $S_0 \to S$
$S_0 \to \boldsymbol{XSX} \mid \boldsymbol{xY} \mid \boldsymbol{x} \mid \boldsymbol{XS} \mid \boldsymbol{SX}$
$S \to XSX \mid xY \mid x \mid XS \mid SX$
$X \to Y \mid S$
$Y \to z$

(c) Eliminate $X \to S$
$S_0 \to XSX \mid xY \mid x \mid XS \mid SX$
$S \to XSX \mid xY \mid x \mid XS \mid SX$
$X \to Y \mid \boldsymbol{XSX} \mid \boldsymbol{xY} \mid \boldsymbol{x} \mid \boldsymbol{XS} \mid \boldsymbol{SX}$
$Y \to z$

(d) Eliminate $X \to Y$
$S_0 \to XSX \mid xY \mid x \mid XS \mid SX$
$S \to XSX \mid xY \mid x \mid XS \mid SX$
$X \to \boldsymbol{z} \mid XSX \mid xY \mid x \mid XS \mid SX$
$Y \to z$

4. Replace long productions

(a) Eliminate XSX as XX_1 and SX
$S_0 \to \boldsymbol{XX_1} \mid xY \mid x \mid XS \mid SX$
$S \to \boldsymbol{XX_1} \mid xY \mid x \mid XS \mid SX$
$X \to 1 \mid \boldsymbol{XX_1} \mid xY \mid x \mid XS \mid SX$
$\boldsymbol{X_1 \to SX}$
$Y \to z$

5. Move terminals to unit productions

(a) Eliminate xY as ZY and $Z \to x$
$S_0 \to XX_1 \mid \boldsymbol{ZY} \mid x \mid XS \mid SX$
$S \to XX_1 \mid \boldsymbol{ZY} \mid x \mid XS \mid SX$
$X \to 1 \mid XX_1 \mid \boldsymbol{ZY} \mid x \mid XS \mid SX$
$X_1 \to SX$
$\boldsymbol{Z \to x}$
$Y \to z$

Final CNF grammar:

$S_0 \to XX_1 \mid XS \mid SX \mid ZY \mid x$
$S \to XX_1 \mid XS \mid SX \mid ZY \mid x$
$X \to XX_1 \mid XS \mid SX \mid ZY \mid x \mid z$
$X_1 \to SX$
$Z \to x$
$Y \to z$

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

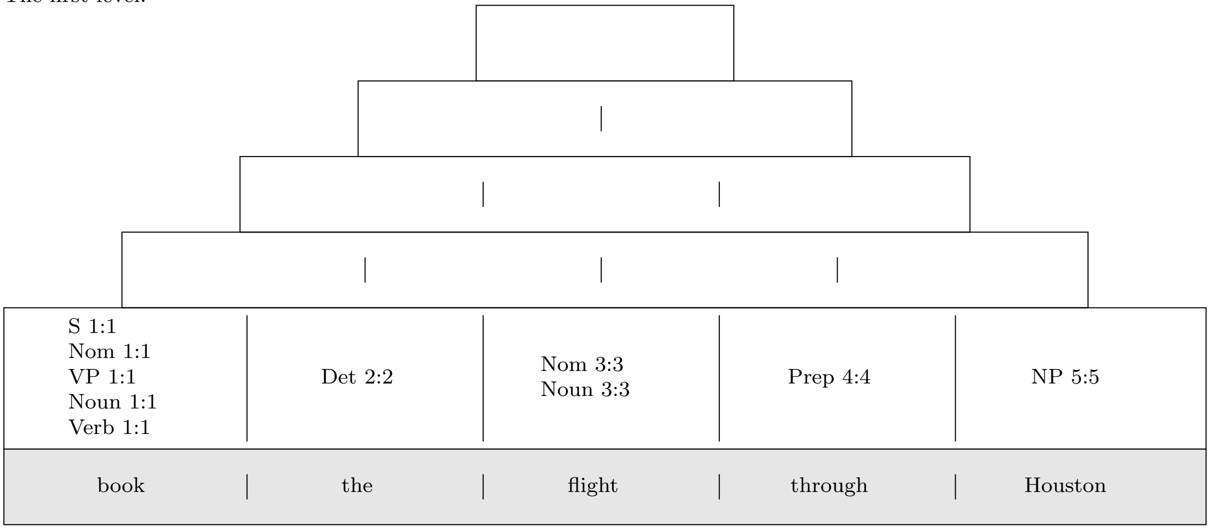
VP → book | include | prefer
VP → Verb NP
VP → X2 PP
X2 → Verb NP
VP → Verb PP
VP → VP PP
PP → Prep NP
Det → that | this | the | a
Noun → book | flight | meal | money
Verb → book | include | prefer
Aux → does
Prep → from | to | on | near | through

book the flight through Houston

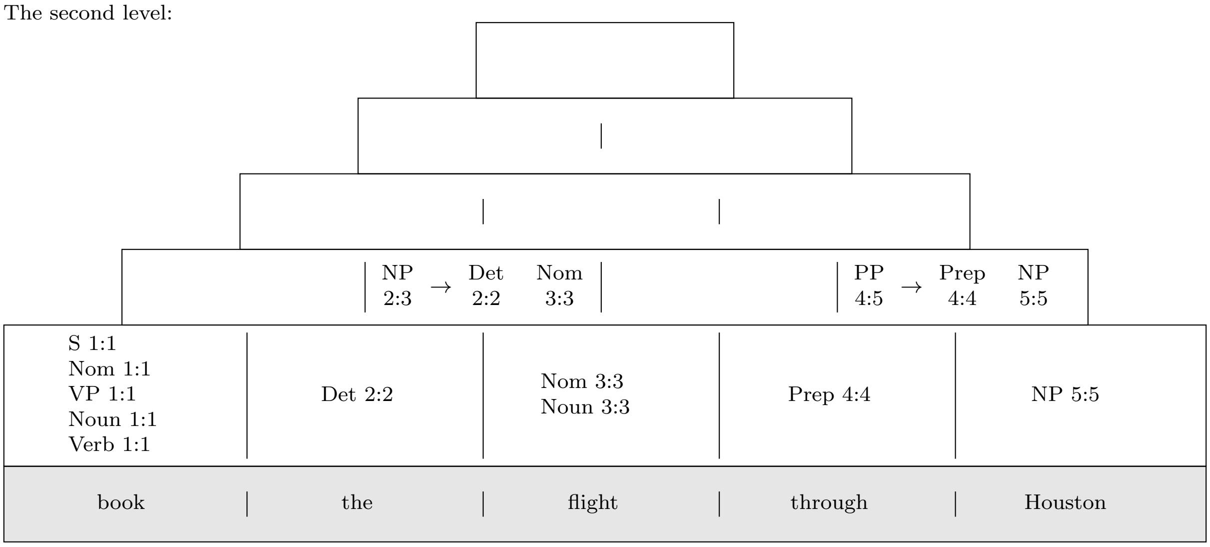
Empty parse table:

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

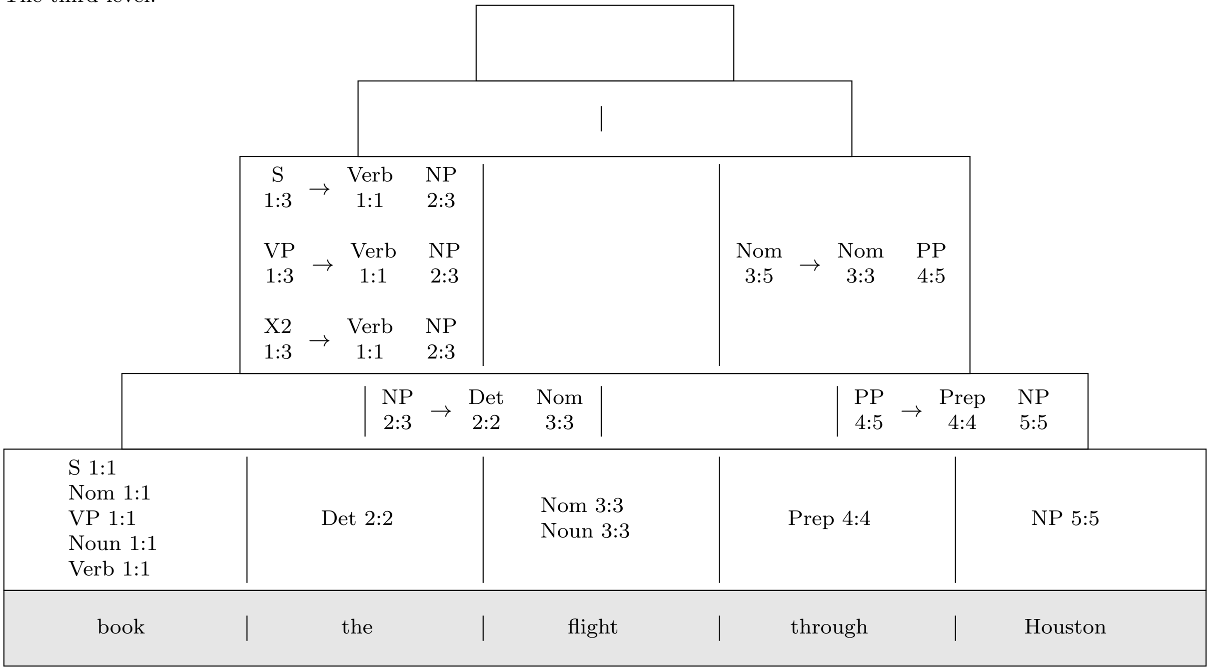
The first level:



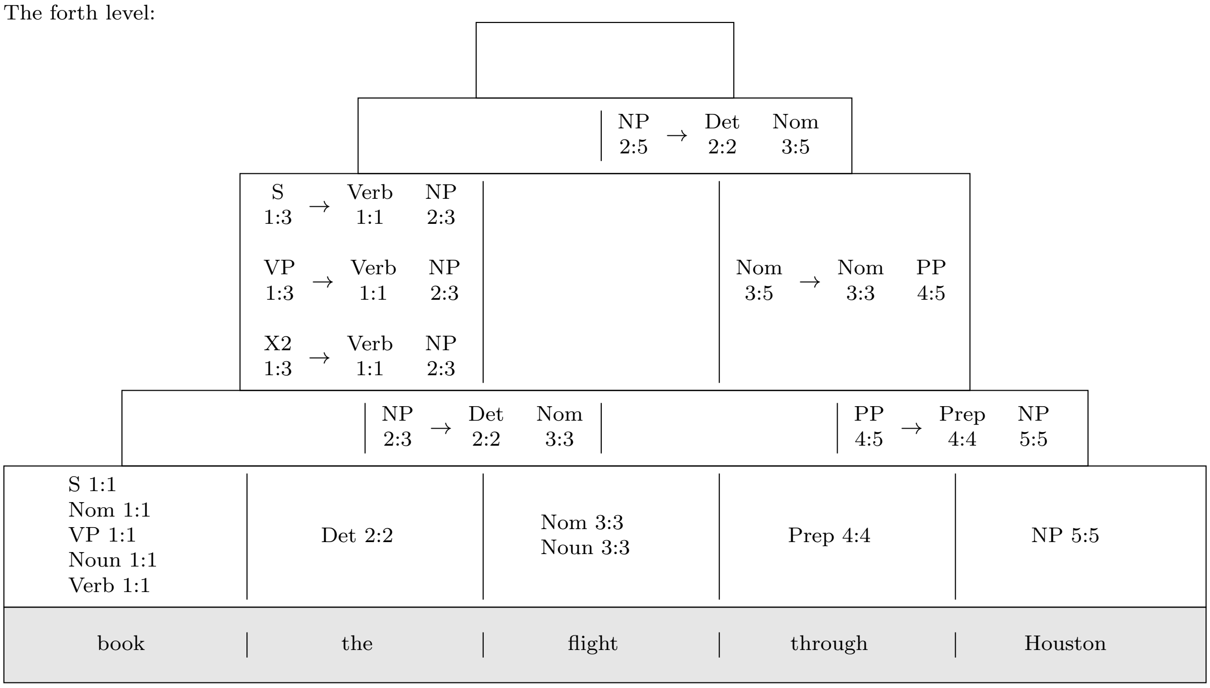
The second level:



The third level:



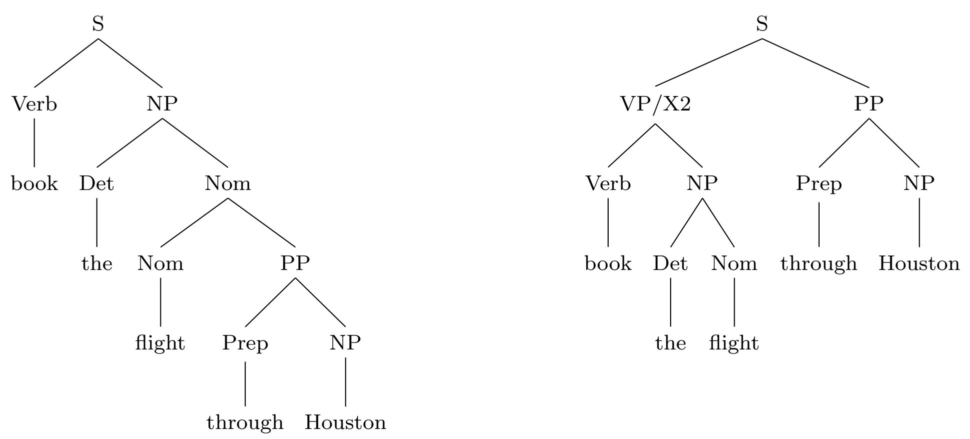
The forth level:



The final level:

S 1:5 → Verb 1:1 NP 2:5 S 1:5 → VP 1:3 PP 4:5 S 1:5 → X2 1:3 PP 4:5 VP 1:5 → Verb 1:1 NP 2:5 VP 1:5 → VP 1:3 PP 4:5 VP 1:5 → X2 1:3 PP 4:5 X2 1:5 → Verb 1:1 NP 2:5				
	NP 2:5 → Det 2:2 Nom 3:5			
S 1:3 → Verb 1:1 NP 2:3 VP 1:3 → Verb 1:1 NP 2:3 X2 1:3 → Verb 1:1 NP 2:3		Nom 3:5 → Nom 3:3 PP 4:5		
	NP 2:3 → Det 2:2 Nom 3:3		PP 4:5 → Prep 4:4 NP 5:5	
S 1:1 Nom 1:1 VP 1:1 Noun 1:1 Verb 1:1	Det 2:2	Nom 3:3 Noun 3:3	Prep 4:4	NP 5:5
book	the	flight	through	Houston

Parse Trees:

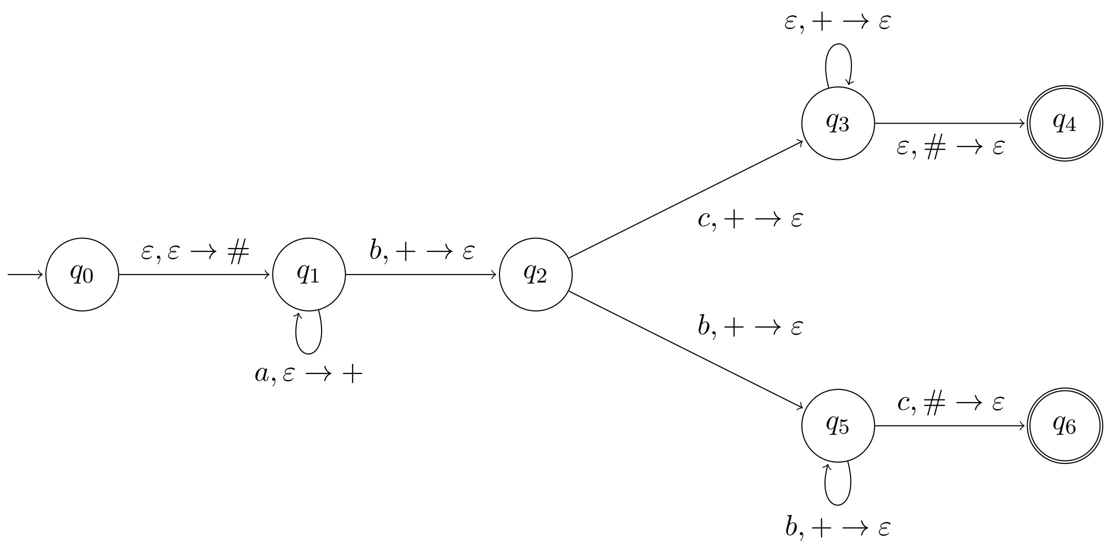


7 Deterministic Pushdown Automata (not graded)

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

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

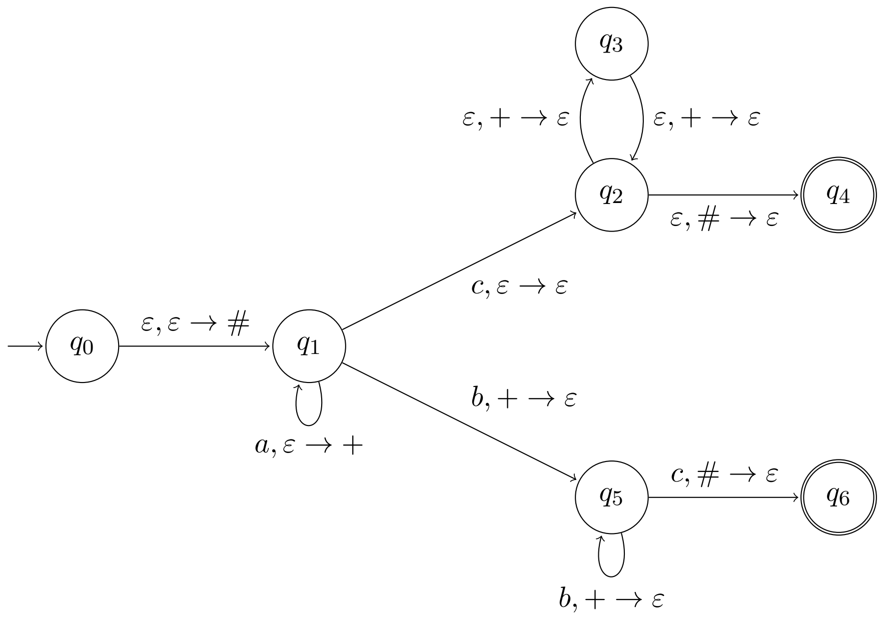
The transition function δ of a DPDA must satisfy the following condition:
For every $q \in Q$, $a \in \Sigma$, and $x \in \Gamma$,
exactly one of $\delta(q,a,x)$, $\delta(q,a,\varepsilon)$, $\delta(q,\varepsilon,x)$, and $\delta(q,\varepsilon,\varepsilon)$ is not $\emptyset$.



where $\Sigma = \{a,b,c\}$ and $\Gamma = \{+,\#\}$

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

The transition function δ of a DPDA must satisfy the following condition:
For every $q \in Q$, $a \in \Sigma$, and $x \in \Gamma$,
exactly one of $\delta(q, a, x)$, $\delta(q, a, \varepsilon)$, $\delta(q, \varepsilon, x)$, and $\delta(q, \varepsilon, \varepsilon)$ is not $\emptyset$.



where $\Sigma = \{a, b, c\}$ and $\Gamma = \{+, \#\}$

Submission

- **Late Submission:** You have 2 days in total for late submission of all take-home exams. All submissions will be graded as normal during this period. No further late submissions are accepted.
- You should submit your solutions as a single file named **the2-e1234567.tex**. Please use the template provided on COW with appropriate modifications. THE should compile and produce a PDF file with a single command:

```
pdflatex the2-e1234567.tex
```

- You do not need to submit solutions for not-graded questions. Yet solving them is advisable for studying for the midterm.

Regulations

1. **Cheating: We have zero tolerance policy for cheating**. People involved in cheating will be punished according to the university regulations.
2. **Newsgroup:** You must follow the newsgroup (news.ceng.metu.edu.tr) for discussions and possible updates on a daily basis.

References

Various LaTeX examples on drawing and mathematical symbols:

- `https://en.wikibooks.org/wiki/LaTeX/Mathematics`
- `https://en.wikibooks.org/wiki/LaTeX/Linguistics`
- `https://www.texample.net/tikz/examples/`
- `https://www1.essex.ac.uk/linguistics/external/clmt/latex4ling/`