CIT 596 Homework 4

Steven Tomcavage
stomcava@seas.upenn.edu

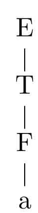
March 3, 2011

1 Exercise 2.1

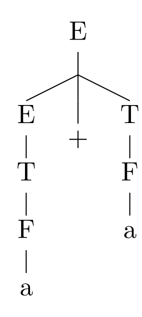
Given the following CFG, provide parse trees for the string in each part:

$$
\begin{aligned}
E &\rightarrow E + T \mid T \\
T &\rightarrow T \times F \mid F \\
F &\rightarrow (E) \mid a
\end{aligned}
$$

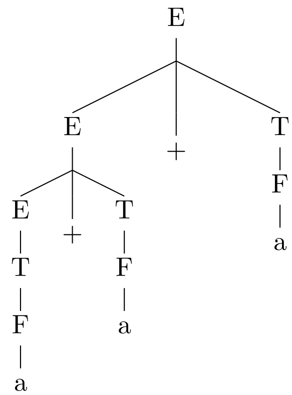
1.1 Part a: a



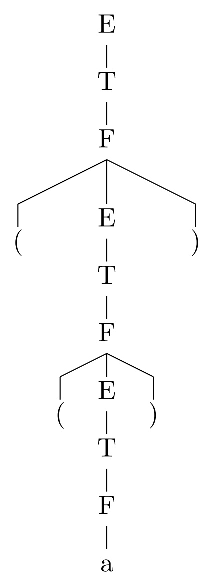
1.2 Part b: $a + a$



1.3 Part c: $a + a + a$



1.4 Part d : $((a))$



2 Exercise 2.2

2.1 Part a

Use the languages $A = \{a^m b^n c^n \mid m, n \geq 0\}$ and $B = \{a^n b^n c^m \mid m, n \geq 0\}$ together with Example 2.36 from Sipser to show the class of context-free languages is not closed under intersection.

Proof.

1. Assume the class of context-free languages is closed under intersection.
2. Given that A and B are context-free languages.
3. The intersection of A and B is the language $C = \{a^i b^i c^i \mid i \geq 0\}$.
4. By Example 2.36, C is not a context-free language.
5. Thus, the intersection of A and B is not a context-free language.
6. Therefore, the class of context-free languages is not closed under intersection. □

2.2 Part b

Use Part (a) and DeMorgan's law to show that the class of context-free languages is not closed under complementation.

Proof.

1. TODO

□

3 Exercise 2.4b

Given $\Sigma = \{0, 1\}$, give a CFG that generates the language $\{w \mid w$ starts and ends with the same symbol$\}$.

$$
\begin{aligned}
S &\rightarrow 0A0 \mid 1A1 \\
A &\rightarrow 0 \mid 1 \mid A \mid \epsilon
\end{aligned}
$$

4 Exercise 2.4c

Given $\Sigma = \{0, 1\}$, give a CFG that generates the language $\{w \mid w$ the length of w is odd$\}$.

$$
\begin{aligned}
S &\rightarrow 0A \mid 1A \\
A &\rightarrow 00 \mid 01 \mid 10 \mid 11 \mid A \mid \epsilon
\end{aligned}
$$

5 Exercise 2.4e

Given $\Sigma = \{0, 1\}$, give a CFG that generates the language $\{w \mid w = w^R$, that is$w$is a palindrome$\}$.

$$
\begin{aligned}
S &\rightarrow 0A0 \mid 1A1 \mid \epsilon \\
A &\rightarrow S
\end{aligned}
$$

6 Exercise 2.5b

Give an informal description and state diagram for the language describe by Exercise 2.4b.

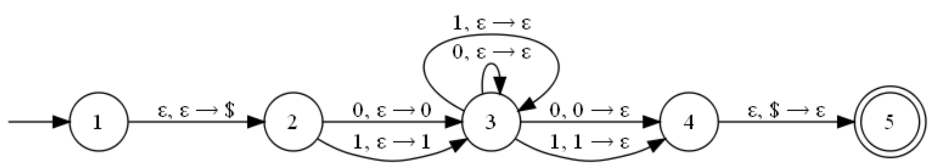


Figure 1: PDA for Exercise 2.5b

7 Exercise 2.5c

Give an informal description and state diagram for the language describe by Exercise 2.4c.

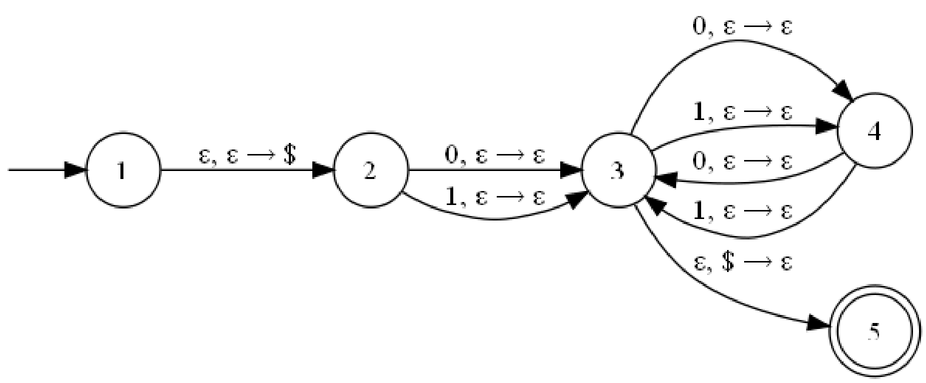


Figure 2: PDA for Exercise 2.5c

8 Exercise 2.5e

Give an informal description and state diagram for the language describe by Exercise 2.4e.

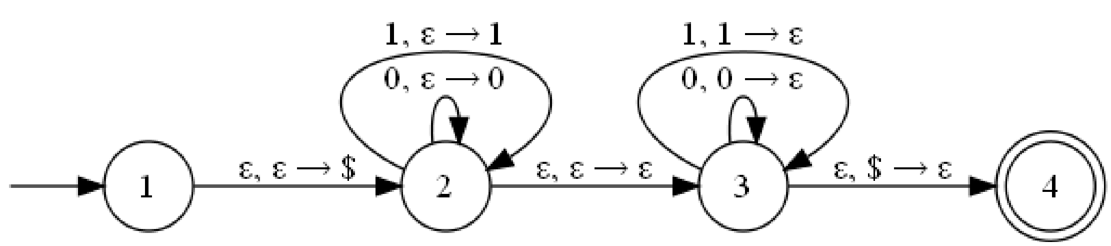


Figure 3: PDA for Exercise 2.5e

9 Exercise 2.9

Give a CFG that generates the langage $A = \{a^i b^j c^k \mid i = j \text{ or } j = k \text{ where } i, j, k \geq 0\}$. Is this CFG ambiguous?

9.1 Part a

$$\begin{aligned}
&S \to Wc \mid aX \\
&W \to aWbY \\
&X \to bXcZ \\
&Y \to W \mid \epsilon \\
&Z \to X \mid \epsilon
\end{aligned}$$

9.2 Part b

No, this CFG is not ambiguous because the leftmost derivation of any string only generates one parse tree. The productions progress linearally and the only loops in the CFG always loop back to the same location. There is never an option for a loop to have a choice of where to return to.

10 Exercise 2.13

10.1 Part a

$L(G)$ generates a string of zeros with one or two hash marks in the string. If there are two hash marks in the string, the hash marks can be at the beginning, end, or anywhere in the middle. If there is only one hash mark, the number of zeros after the hash mark is twice the number of zeros before the hash mark.

10.2 Part b

$L(G)$ is not regular because it contains two recursive rules. T is formed from either a hash mark or from some combination of 0 and T. Likewise, U is formed from either a hash mark or from some combination of 0 and U. These structures cannot be described by regular expressions. Since regular languages are described by regular expressions, the language $L(G)$ is not regular.

11 Exercise 2.14

Convert the following CFG to Chomsky Normal Form:

$$\begin{aligned}
&A \to BAB \mid B \mid \epsilon \\
&B \to 00 \mid \epsilon
\end{aligned}$$

Step 1

Add a new start variable

$$\begin{aligned}
&S \to A \\
&A \to BAB \mid B \mid \epsilon \\
&B \to 00 \mid \epsilon
\end{aligned}$$

Step 2

Eliminate $B \to \epsilon$

$$\begin{aligned}&S \to A\\&A \to BAB \mid B \mid \epsilon \mid BA \mid AB\\&B \to 00\end{aligned}$$

Step 3

Eliminate $A \to \epsilon$

$$\begin{aligned}&S \to A \mid \epsilon\\&A \to BAB \mid B \mid BA \mid AB \mid BB\\&B \to 00\end{aligned}$$

Step 4

Remove unit rules

$$\begin{aligned}&S \to BAB \mid 00 \mid BA \mid AB \mid BB \mid \epsilon\\&A \to BAB \mid 00 \mid BA \mid AB \mid BB\\&B \to 00\end{aligned}$$

12 Exercise 2.20

Let $A/B = \{w|wx \in A$ for some x in B. Show that, if A is context-free and B is regular, than A/B is context-free.

Proof.

1. Given that A/B is context-free.
2. Given that $wx \in A$ for some $x \in B$.
3. Then B is contained in the language A.
4. Therefore, B is context-free. □

13 Exercise 2.26

Show that, if G is a CFG in Chomsky normal form, then for any string $w \in L(G)$ of length $n \geq 1$, exactly $2n - 1$ steps are required for any derivation of w.

Proof.

1. Given that CNF only allows productions of the form $A \to BC$ and $A \to a$.
2. If $n = 1$, then the only production required to generate w is of the form $A \to a$, thus the length is $1 = 2n - 1$.
3. If $n = 2$, then three productions are required, one of the form $A \to BC$ and two of the form $A \to a$, thus the number of steps in the derivation is $3 = 2n - 1$.

4. If $n > 2$, then we can break w down into units of size 2 or 1 and combine those units to form the larger derivation, which will have $(2n_i - 1) + (2n_j - 1) + \ldots + (2n - 1)$ steps.

5. Therefore, the number of steps required to derive w is $2n - 1$. $\square$