Context Free Languages: Problems for Week 4

Exercise 1 *Consider the language generated by the grammar*

$$\Rightarrow S \quad ::= \quad \mathtt{b}SS \quad | \quad \mathtt{a}S \quad | \quad \mathtt{a}$$

and the string `aabbaaa`.

1. *Find a leftmost derivation for this string.*
2. *Draw the derivation tree.*

Solution 1

1. *Leftmost derivation is given below:*

$$\begin{aligned}\dot{S} &\Rightarrow a\dot{S}\\ &\Rightarrow aa\dot{S}\\ &\Rightarrow aab\dot{S}S\\ &\Rightarrow aabb\dot{S}SS\\ &\Rightarrow aabba\dot{S}S\\ &\Rightarrow aabbaa\dot{S}\\ &\Rightarrow aabbaaa\end{aligned}$$

2. *The derivation tree for the above string would be:*

```
  S
 / \
-----
a    S
    / \
   -----
   a    S
       /   \
      /      \
     /         \
-------------
b      S       S
      /   \   / \
     /     \ -----
    /        \  a
-------------
b      S     S
      / \   / \
     ----- -----
       a     a
```

Exercise 2 *Let's look at a "Natural Language" example. The alphabet is*

{ *the, a, cat, dog, happy, tired, slept, died, ate, dinner, and, .* }

The grammar is

Sentence	$\Rightarrow S$	::=	*C .*
Clause	*C*	::=	*NP VP* | *C and C*
Noun phrase	*NP*	::=	*Art N* | *dinner*
Noun	*N*	::=	*Adj N* | *cat* | *dog*
Adjective	*Adj*	::=	*happy* | *tired*
Verb phrase	*VP*	::=	*VI* | *VT NP*
Intransitive verb	*VI*	::=	*slept* | *died*
Transitive verb	*VT*	::=	*ate*
Article	*Art*	::=	*a* | *the*

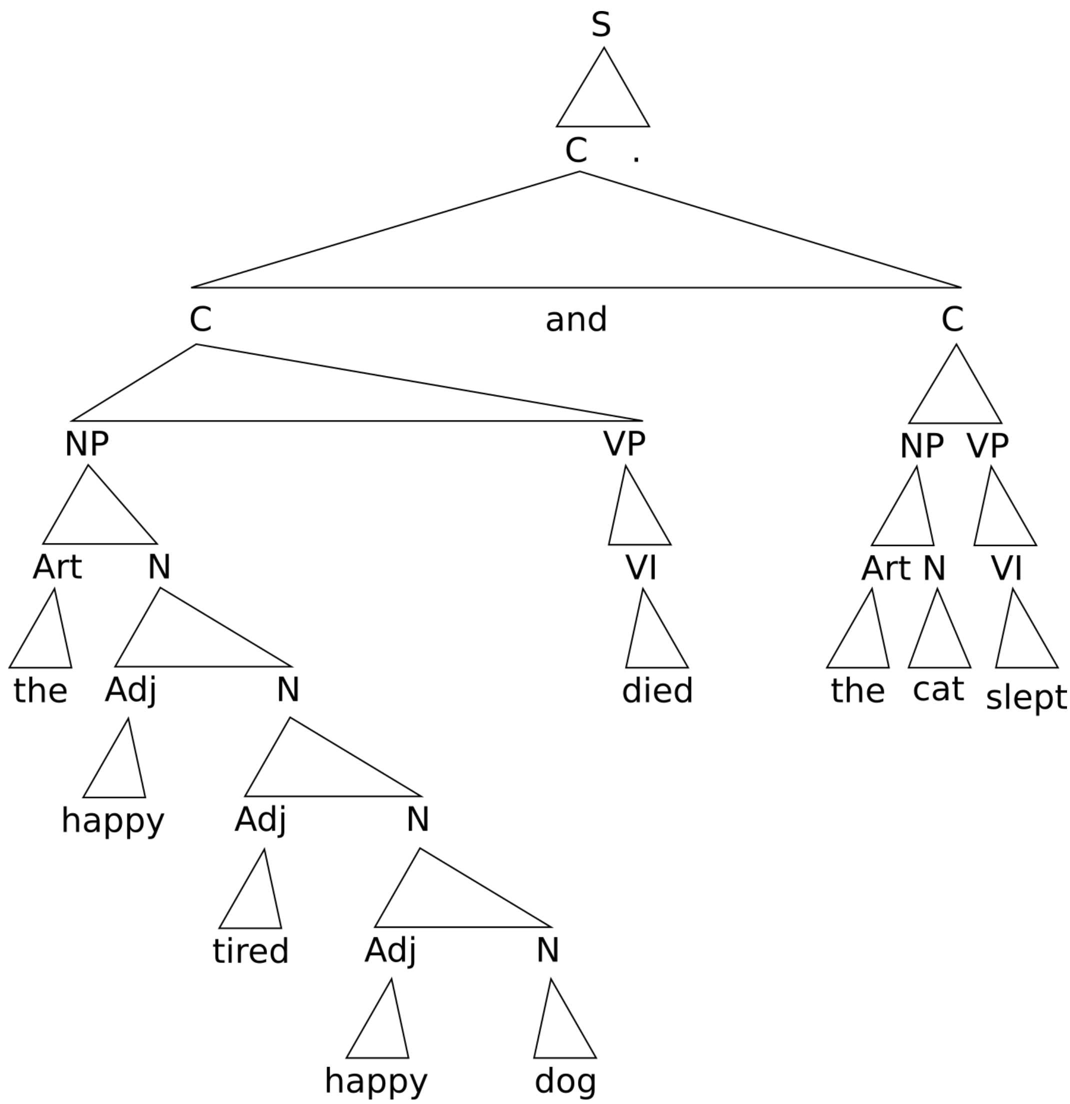
This grammar accepts "words" such as

the happy tired happy dog died and the cat slept.
the tired tired cat ate dinner.
dinner ate a happy dog.

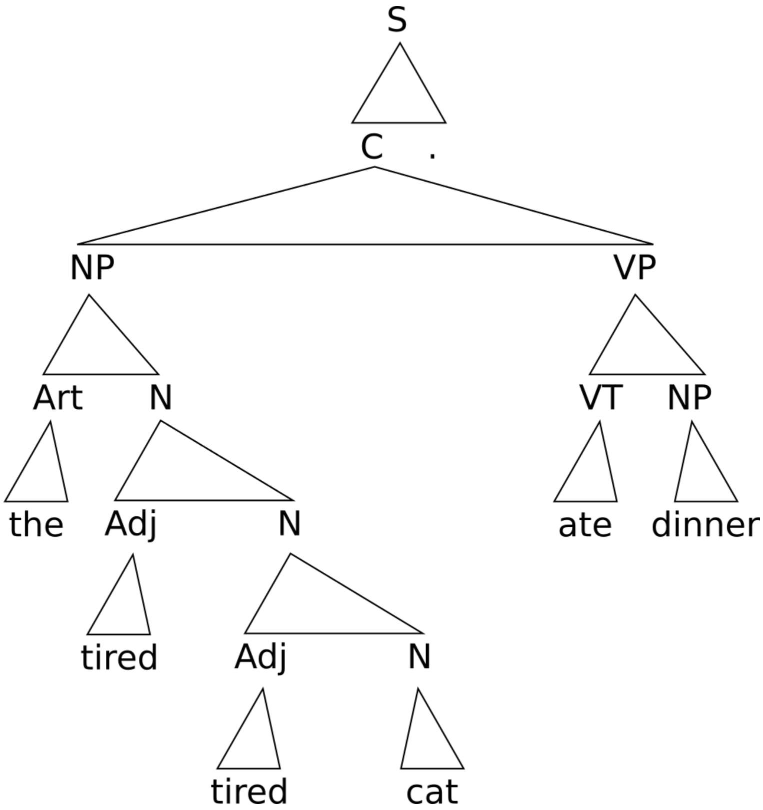
Try writing derivations and derivation trees for these sentences.

Solution 2

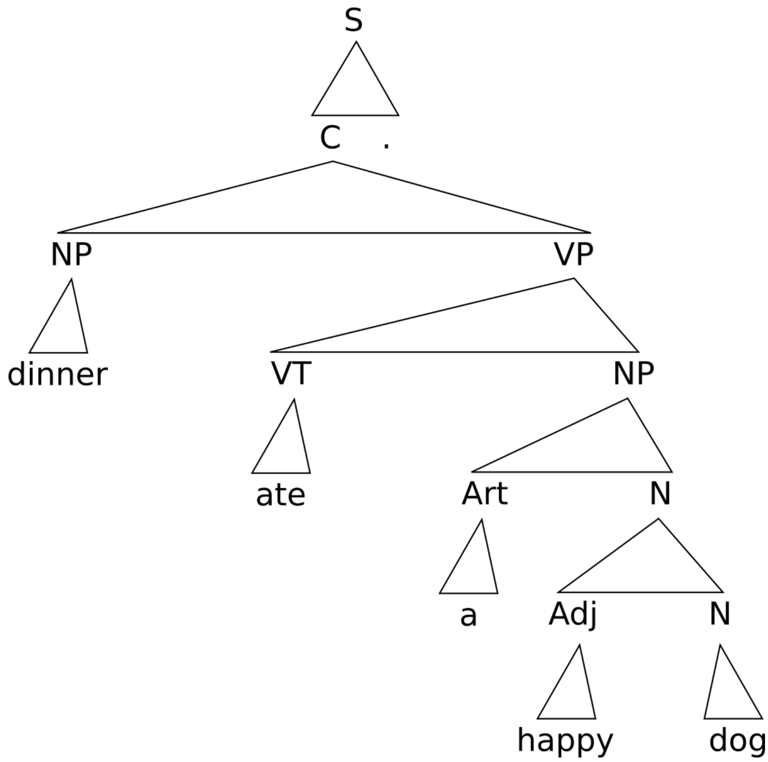
1. *The derivation tree for "the happy tired happy dog died and the cat slept."*



2. *The derivation tree for "the tired tired cat ate dinner."*

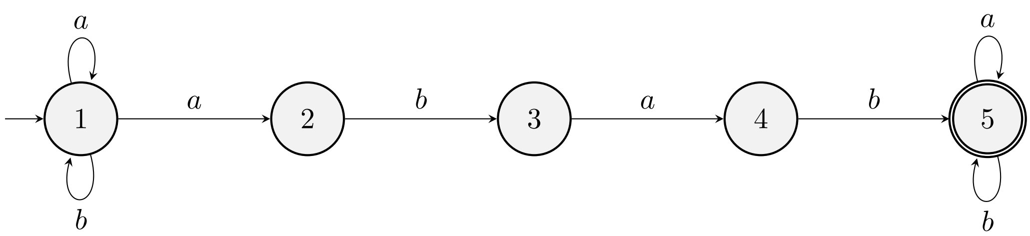


3. *The derivation tree for "dinner ate a happy dog."*



We will leave the derivations to you!

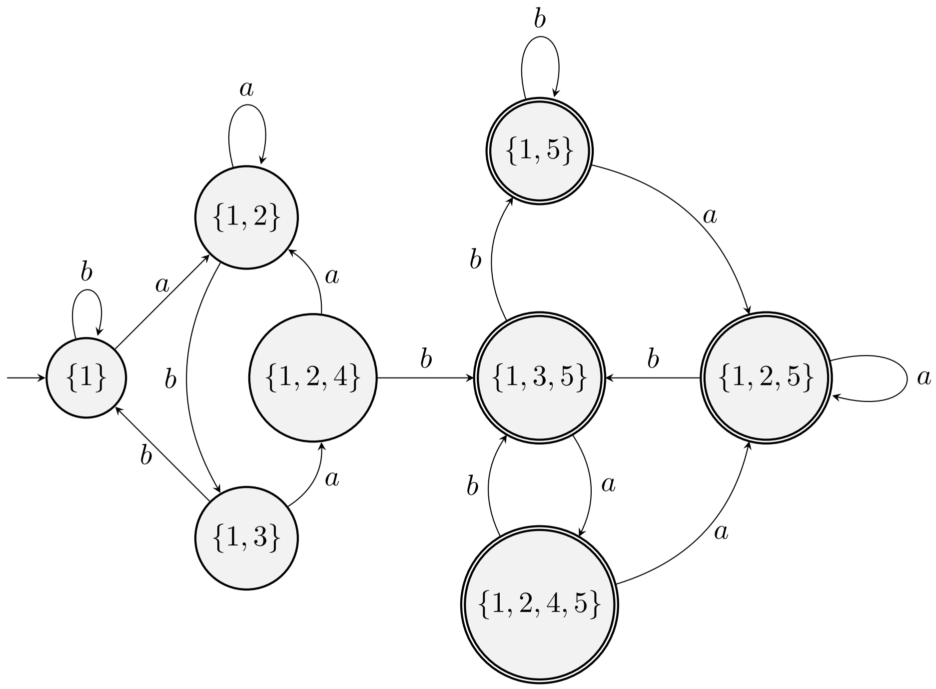
Exercise 3 *Let's consider an NFA that accepts any string that contains the substring "abab".*



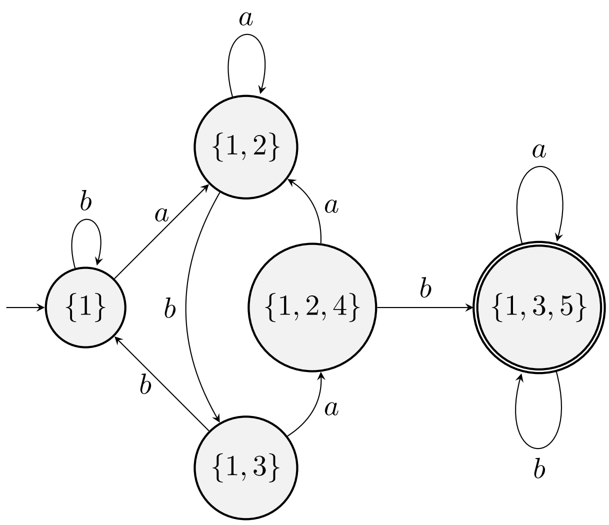
1. *Convert the above NFA into its equivalent total DFA.*
2. *Convert the resultant DFA in an equivalent CFG. It is suggested to minimize the DFA before writing CFG.*

Solution 3

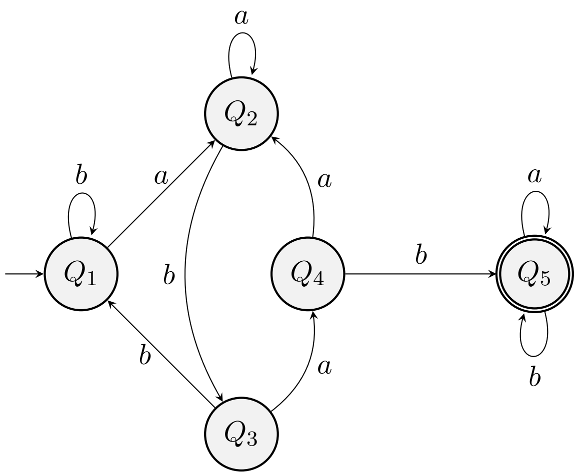
1. *On converting the NFA to its equivalent total DFA, we get the following:*



2. *We can see that the states* $\{1,3,5\}$, $\{1,5\}$, $\{1,2,4,5\}$, *and* $\{1,2,5\}$ *are equivalent and could be removed. On minimizing, we get the following DFA:*



We can rename the states, to make it easier for us to write the equivalent CFG.



Now, we can very easily write the CFG for the above DFA:

$$
\begin{array}{rcl}
R_1 & ::= & aR_2 \mid bR_1 \\
R_2 & ::= & aR_2 \mid bR_3 \\
R_3 & ::= & aR_4 \mid bR_1 \\
R_4 & ::= & aR_2 \mid bR_5 \\
R_5 & ::= & aR_5 \mid bR_5 \mid \varepsilon \\
\textit{Start:} & & R_1
\end{array}
$$

Exercise 4 *Give a context free grammar for the set of palindromes over the alphabet* $\{a, b\}$.

Solution 4
A CFG for the set of palindromes over the alphabet $\{a, b\}$ *is given below:*

$$\Rightarrow S \quad ::= \quad \varepsilon \mid \mathtt{a} \mid \mathtt{b} \mid \mathtt{aSa} \mid \mathtt{bSb}$$

Exercise 5 *Try deriving the string* $3 + 5 \times 3$ *in two different ways using leftmost derivation only, using the grammar given below:*

$$
\begin{array}{rcl}
\Rightarrow A & ::= & A + B \mid B \\
B & ::= & B \times C \mid C \\
C & ::= & (A) \mid 3 \mid 5
\end{array}
$$

Solution 5

There is only one possible leftmost derivation, as shown below (because the grammar is unambiguous):

$$
\begin{array}{rcl}
\dot{A} & \Rightarrow & \dot{A} + B \\
& \Rightarrow & \dot{B} + B \\
& \Rightarrow & \dot{C} + B \\
& \Rightarrow & 3 + \dot{B} \\
& \Rightarrow & 3 + \dot{B} \times C \\
& \Rightarrow & 3 + \dot{C} \times C \\
& \Rightarrow & 3 + 5 \times \dot{C} \\
& \Rightarrow & 3 + 5 \times 3
\end{array}
$$

Exercise 6 *Show that the following grammar is ambiguous. The alphabet is* $\{\mathtt{a}, \mathtt{b}\}$.

$$
\begin{array}{rcl}
\Rightarrow P & ::= & \varepsilon \mid Q\mathtt{a} \mid \mathtt{a}Q \\
Q & ::= & \mathtt{aa}P \mid \mathtt{b}R \\
R & ::= & Q\mathtt{a}
\end{array}
$$

Solution 6

Here is the first leftmost derivation for the string "aaa":

$$\begin{array}{rcl} \dot{P} & \Rightarrow & \dot{Q}a \\ & \Rightarrow & aa\dot{P}a \\ & \Rightarrow & aaa \end{array}$$

We can get the same "aaa" string using the following leftmost derivation:

$$\begin{array}{rcl} \dot{P} & \Rightarrow & a\dot{Q} \\ & \Rightarrow & aaa\dot{P} \\ & \Rightarrow & aaa \end{array}$$

Therefore, the given grammar is ambiguous!

Exercise 7 *Convert the following CFG into an equivalent CFG in Chomsky normal form*

$$\begin{array}{rcl} \Rightarrow A & ::= & BAB \mid B \mid \varepsilon \\ B & ::= & 00 \mid \varepsilon \end{array}$$

Solution 7

1. *Add a new start variable S_0*

$$\begin{array}{rcl} \Rightarrow \boldsymbol{S_0} & \boldsymbol{::=} & \boldsymbol{A} \\ A & ::= & BAB \mid B \mid \varepsilon \\ B & ::= & 00 \mid \varepsilon \end{array}$$

2a. *Remove ε-rule $B ::= \varepsilon$*

$$\begin{array}{rcl} \Rightarrow S_0 & ::= & A \\ A & ::= & BAB \mid B \mid \boldsymbol{AB} \mid \boldsymbol{BA} \mid \boldsymbol{A} \mid \varepsilon \\ B & ::= & 00 \end{array}$$

2b. *Remove ε-rule $A ::= \varepsilon$*

$$\begin{array}{rcl} \Rightarrow S_0 & ::= & A \mid \boldsymbol{\varepsilon} \\ A & ::= & BAB \mid B \mid AB \mid BA \mid A \mid \boldsymbol{BB} \\ B & ::= & 00 \end{array}$$

***Note**: $S_0 ::= \varepsilon$ is allowed in CNF.*

3a. *Remove the unit rule $A ::= A$*

$$\begin{array}{rcl} \Rightarrow S_0 & ::= & A \mid \varepsilon \\ A & ::= & BAB \mid B \mid AB \mid BA \mid BB \\ B & ::= & 00 \end{array}$$

3b. *Remove the unit rule $A ::= B$*

$$\begin{array}{rcl} \Rightarrow S_0 & ::= & A \mid \varepsilon \\ A & ::= & BAB \mid \mathbf{00} \mid AB \mid BA \mid BB \\ B & ::= & 00 \end{array}$$

3c. Remove the unit rule $S_0 ::= \quad A$

$$\begin{aligned} \Rightarrow S_0 \quad &::= \quad \boldsymbol{BAB} \mid \mathbf{00} \mid \boldsymbol{AB} \mid \boldsymbol{BA} \mid \boldsymbol{BB} \mid \varepsilon \\ A \quad &::= \quad BAB \mid 00 \mid AB \mid BA \mid BB \\ B \quad &::= \quad 00 \end{aligned}$$

4a. Convert the rules into proper form; introduce rule $D ::= \quad 0$

$$\begin{aligned} \Rightarrow S_0 \quad &::= \quad BAB \mid \boldsymbol{DD} \mid AB \mid BA \mid BB \mid \varepsilon \\ A \quad &::= \quad BAB \mid \boldsymbol{DD} \mid AB \mid BA \mid BB \\ B \quad &::= \quad \boldsymbol{DD} \\ \boldsymbol{D} \quad &\boldsymbol{::=} \quad \mathbf{0} \end{aligned}$$

4b. Convert the rules into proper form; introduce rule $C ::= \quad AB$

$$\begin{aligned} \Rightarrow S_0 \quad &::= \quad B\boldsymbol{C} \mid DD \mid AB \mid BA \mid BB \mid \varepsilon \\ A \quad &::= \quad B\boldsymbol{C} \mid DD \mid AB \mid BA \mid BB \\ B \quad &::= \quad DD \\ \boldsymbol{C} \quad &\boldsymbol{::=} \quad \boldsymbol{AB} \\ D \quad &::= \quad 0 \end{aligned}$$

The above CFG is now in Chomsky Normal Form.

Exercise 8 *Give grammars for the following two languages:*

1. *All binary strings with both an even number of zeroes and an even number of ones.*
2. *All strings of the form $0^a1^b0^c$ where $a + c = b$.*

Solution 8

1. *All binary strings with both an even number of zeroes and an even number of ones.*

$$\begin{aligned} \Rightarrow S \quad &::= \quad 0X \mid 1Y \mid \varepsilon \\ X \quad &::= \quad 0S \mid 1Z \\ Y \quad &::= \quad 1S \mid 0Z \\ Z \quad &::= \quad 0Y \mid 1X \end{aligned}$$

The idea here is that:

(a) *Production S corresponds to having an even number of zeroes and an even number of ones.*
(b) *Production X corresponds to having an odd number of zeroes and an even number of ones.*
(c) *Production Y corresponds to having an even number of zeroes and an odd number of ones.*
(d) *Production Z corresponds to having an odd number of zeroes and an odd number of ones.*

You can check this yourself by producing some binary strings using the grammar and keeping track of the parity of the number of zeroes and ones as you move through the productions.

2. *All strings of the form $0^a1^b0^c$ where $a + c = b$.*

$$\begin{aligned} \Rightarrow S \quad &::= \quad TU \\ T \quad &::= \quad 0T1 \mid \varepsilon \\ U \quad &::= \quad 1U0 \mid \varepsilon \end{aligned}$$

This works because for every 0 generated on the right, we generate a 1 in the middle (see U) and for every 0 generated on the left, we generate another 1 in the middle (see T).