Homework #2

1. Consider the following three English sentences:

   - The book on FST is in the library.
   - I returned a book to the library.
   - I read a book about Linguistics.

   Think about their constituent structure: which groups of words make up constituents given the meaning of each sentence? You don't need to submit an answer – just do this for yourself, using boxes or square brackets to group words.

   Now consider the following grammar, which includes some of the rules introduced in sections 12.2 and 12.3 of the textbook *Speech & Language Processing*:

Lexicon		Phrase structure rules		
N	$\rightarrow$ *book* | *library*	S	$\rightarrow$	$NP\ \ VP$
V	$\rightarrow$ *is* | *read* | *returned*	NP	$\rightarrow$	$Det\ \ Nom$
Pro	$\rightarrow$ *I* | *you*	NP	$\rightarrow$	Pro
PN	$\rightarrow$ *FST* | *Linguistics*	NP	$\rightarrow$	PN
Det	$\rightarrow$ *the* | *this* | *a*	Nom	$\rightarrow$	N
P	$\rightarrow$ *to* | *about* | *in* | on	Nom	$\rightarrow$	$Nom\ \ PP$
		VP	$\rightarrow$	$V\ \ NP$
		VP	$\rightarrow$	$V\ \ PP$
		VP	$\rightarrow$	$V\ \ NP\ \ PP$
		PP	$\rightarrow$	$P\ \ NP$

   (a) The three sentences above can be parsed with this grammar. For each sentence, say how many possible parse trees the grammar can assign to it. [1.5 points]

   (b) Using this grammar, draw one parse tree for each of these sentence. Select the parse tree that corresponds to the constituent structure that is compatible with the meaning of each sentence. [1.5 points]

   (c) This grammar can generate an infinite number of sentences. Why? Which grammar rules make this possible? [1.5 points]

   (d) Which grammar rules from those introduced in section 12.3 of *Speech & Language Processing* would have to be added to the grammar above to be able to parse the following sentence? Assume that *Which* is a determiner (*Det*). [1.5 points]

[1] Assignment created by Dr. Tejaswini Deoskar

- Which book do you like?

Give the additional rules and the resulting parse tree for the sentence.

2. Consider formal language $L = \{x \mid x$ contains an even number of a's followed by an even number of b's (zero counts as even)$\}$ over alphabet $\Sigma = \{\epsilon, a, b\}$.

   (a) This language is regular. Give i) a regular expression, ii) an FSA, and iii) a right-linear grammar that are equivalent and that generate language L. [1.5 points]

   (b) State the condition specified by the Pumping lemma; then show by example that L meets this condition. [1.5 points]

3. Consider language $L = \{a^n b^2 a^n | n \geq 0\}$. This language includes the following types of strings: *bb*, *abba*, *aabbaa*, *aaabbaaa*, ...

   (a) Is L a regular language? If yes, give a regular expression that defines it. If not, use the Pumping Lemma to show that it is not regular. [3 points]

   (b) Is L a context-free language? If so, give a context-free grammar that generates it. [1 points]