Exercise 8: Probabilistic Context-Free Grammars

You can earn up to 10 points on this exercise.
You may work as a group of up to 3 people, but please submit your own version.
You may use any programming language you wish, but any submission that we cannot run on our computers without installing things must be presented to the class.
Submissions without names on the document will receive no credit.

Please email your solution in PDF to `claytong@coli.uni-saarland.de` *or submit before the tutorial by* **10:14:59 AM GMT+1, January 13, 2016**.

TASK 1

(a) Draw a tree that exhibits left embedding and a tree that exhibits right embedding. No terminal nodes needed. (1 point)

(b) Write a natural language sentence that exhibits left embedding OR write a natural language sentence that exhibits right embedding. Draw the tree and explain. (1 point)

(c) Draw a tree that exhibits center embedding. (1 point)

(d) Write a natural language sentence that exhibits center embedding. Draw the tree and explain. (1 point)

TASK 2

The following is a probabilistic context-free grammar in Chomsky Normal Form. The probability for each rule is given in parentheses.

S	→	NP VP	(1.0)
NP	→	Det N	(0.8)
NP	→	NP PP	(0.2)
PP	→	P NP	(1.0)
VP	→	V NP	(0.7)
VP	→	VP PP	(0.3)

Det	→	"the"	(1.0)
N	→	"man"	(0.3)
N	→	"woman"	(0.3)
N	→	"telescope"	(0.1)
N	→	"hill"	(0.3)
P	→	"with"	(0.4)
P	→	"on"	(0.6)
V	→	"saw"	(1.0)

(a) Using the grammar to the left, create the CKY chart for "The man saw the woman with the telescope on the hill". Write down all appropriate duplicate symbols. The cell for the entire string should contain five S symbols. (3 points)

(b) What is the probability of the string? Which trees would the language model consider? Which trees would a parser using the probabilistic CKY algorithm consider? (3 points)