CS-5340/6340, Written Assignment #2
DUE: Thursday September 20, 2018 by 11:59pm

1. (24 pts)

   (a) Suppose a shallow parser (chunker) is applied to the sentences below. Label all prepositional phrase (PP) chunks and base noun phrase (NP) chunks that should be produced for each sentence. Be sure to label an NP even if it is nested in a PP. For examples *'He stays in Salt Lake City'* should be '[NP: He] stays [PP: in [NP: Salt Lake City]]'.

      i. [**NP**: John] gave [**NP**: Mary] [**NP**: a book] as [**NP**: a gift].

      ii. [**NP**: A person] [**PP**: in [**NP**: the park] threw [**NP**: a bone] [**PP**: to [**NP**: his dog]].

      iii. [**PP**: In [**NP**: July]] [**NP**: Bob] will sell [**NP**: his food truck].

      iv. [**NP**: The mouse] [**NP**: the cat] caught died.

      v. [**NP**: The storm] destroyed [**NP**: a large number] [**PP**: of [**NP**:properties]] [**PP**: in [**NP**: the city]].

   (b) Label each sentence below with BIO tags corresponding to the NP chunks that you identified in part (a).

      i. John/**B** gave/**O** Mary/**B** a/**B** book/**I** as/**O** a/**B** gift/**I** .

      ii. A/**B** person/**I** in/**O** the/**B** park/**I** threw/**O** a/**B** bone/**I** to/**O** his/**B** dog/**I** .

      iii. In/**O** July/**B** Bob/**B** will/**O** sell/**O** his/**B** food/**I** truck/**I** .

      iv. The/**B** mouse/**I** the/**B** cat/**I** caught/**O** died/**O** .

      v. The/**B** storm/**I** destroyed/**O** a/**B** large/**I** number/**I** of/**O** properties/**B** in/**O** the/**B** city/**I** .

2. (24 pts) In this question, you are going to use the CKY algorithm to parse a sentence. (Use the ordinary CKY algorithm, not probabilistic CKY.)

   (a) List all table entries produced by the CKY algorithm for the sentence *'John eats the steak with chopsticks'* using the grammar below. Each entry table[i,j] refers to the cell for row i and column j in the table. For example, table[1,4] should contain constituents that span words 1 through 4.

S → NP VP
NP → NP PP
NP → art noun
PP → prep NP
VP → VP PP
VP → verb NP
NP → John
NP → chopsticks
NP → steak
noun → John
noun → chopsticks
noun → steak
verb → eats
prep → with
art → the

table[1,1]: {noun, NP}
table[1,2]: {}
table[1,3]: {}
table[1,4]: {S}
table[1,5]: {}
table[1,6]: {S, S}
table[2,2]: {verb}
table[2,3]: {}
table[2,4]: {VP}
table[2,5]: {}
table[2,6]: {VP, VP}
table[3,3]: {art}
table[3,4]: {NP}
table[3,5]: {}
table[3,6]: {NP}
table[4,4]: {noun, NP}
table[4,5]: {}
table[4,6]: {NP}
table[5,5]: {prep}
table[5,6]: {PP}
table[6,6]: {noun, NP}

(b) Draw the parse tree for every S constituent produced in the table for part (a). For each one, please indicate which cell the S constituent appears in.

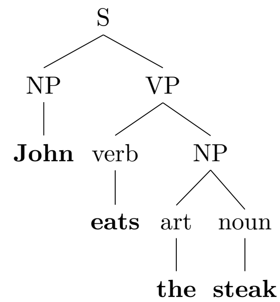


Figure 1: table[1,4]

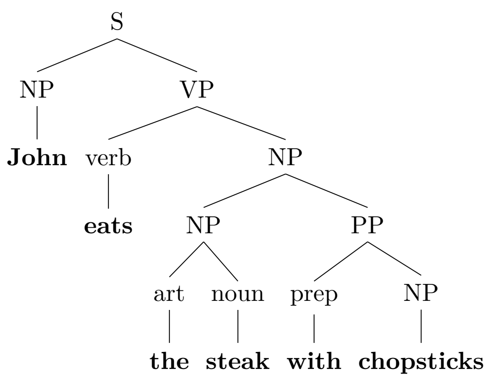


Figure 2: table[1,6]

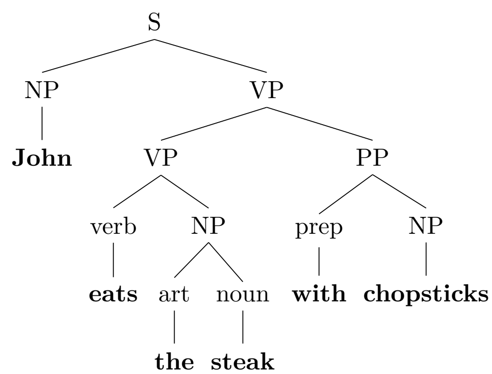


Figure 3: table[1,6]

3. (18 Pts) Given the grammar G2 and the sentence *'the dog park is closed'*, fill in a chart using the Earley chart parsing algorithm. Each chart entry should be a constituent or a rule, with a start and end position indicating the range of words that have been matched by the constituent or rule. For example, 'NP[1-4]' should be used for an NP that matches words in positions 1-3 (thus ends at position 4). Similarly, 'S $\rightarrow$ NP * VP [1,3]' means that the NP has matched the words in positions 1-2 and the rule is anticipating a VP starting in position 3. Assume that the first word in the sentence is in position 1.

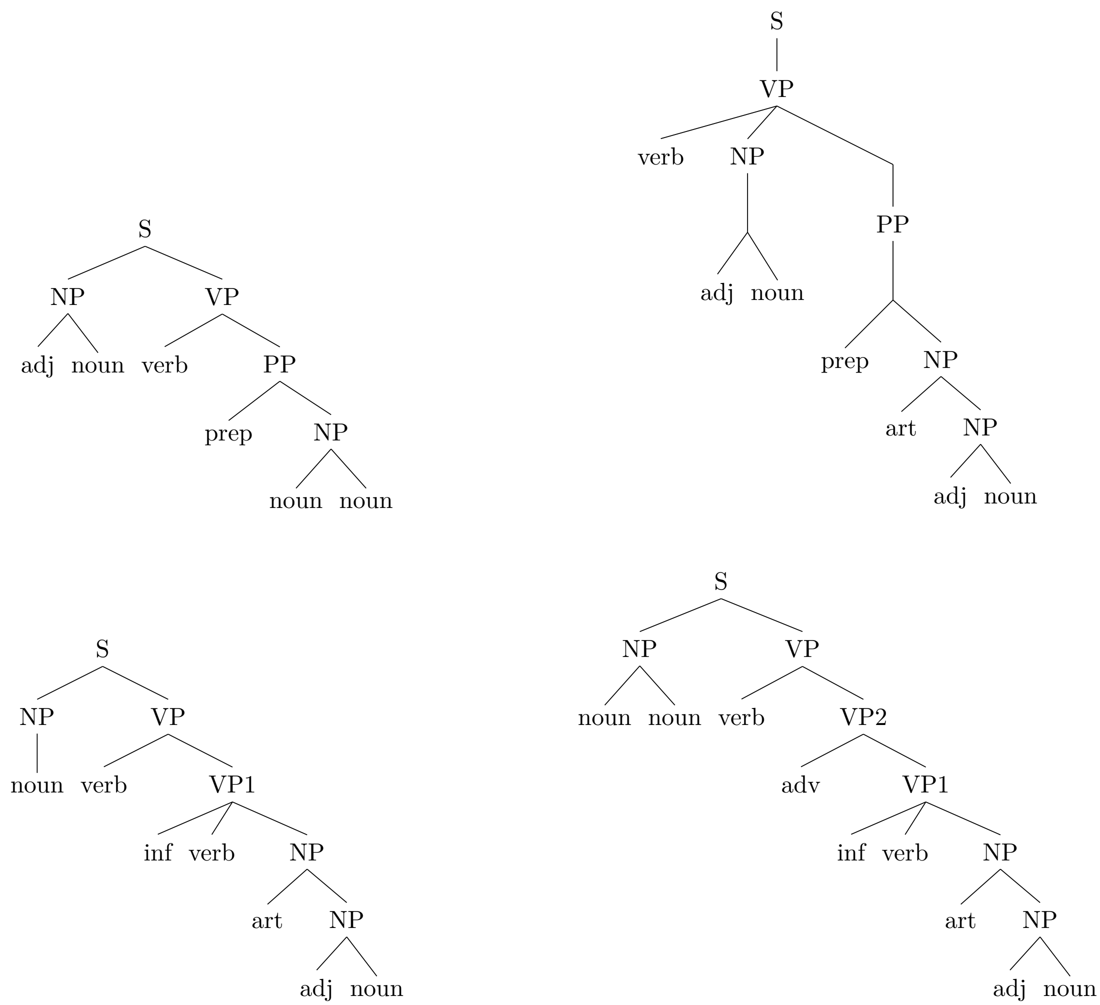
G2
S $\rightarrow$ NP VP
S $\rightarrow$ VP NP
VP $\rightarrow$ verb VP
VP $\rightarrow$ verb
NP $\rightarrow$ art NP
NP $\rightarrow$ noun noun
noun $\rightarrow$ dog | park
verb $\rightarrow$ closed | park | is
art $\rightarrow$ the

Below, we have created the part-of-speech constituents that belong in the chart to get you started. Please fill in the remaining entries!

Constituent or Rule	**Start-End**
art(“the”)	[1-2]
noun(“dog”)	[2-3]
noun(“park”)	[3-4]
verb(“park”)	[3-4]
verb(“is”)	[4-5]
verb(“closed”)	[5-6]

S $\rightarrow$ * NP VP	[1-1]
NP $\rightarrow$ * art NP	[1-1]
NP $\rightarrow$ * noun noun	[1-1]
S $\rightarrow$ * VP NP	[1-1]
VP $\rightarrow$ * verb VP	[1-1]
VP $\rightarrow$ * verb	[1-1]
NP $\rightarrow$ art * NP	[1-2]
NP $\rightarrow$ * art NP	[2-2]
NP $\rightarrow$ * noun noun	[2-2]
NP $\rightarrow$ noun * noun	[2-3]
NP $\rightarrow$ noun noun *	[2-4]
NP	[2-4]
NP $\rightarrow$ art NP *	[1-4]
NP	[1-4]
S $\rightarrow$ NP * VP	[1-4]
VP $\rightarrow$ * verb VP	[4-4]
VP $\rightarrow$ * verb	[4-4]
VP $\rightarrow$ verb * VP	[4-5]
VP $\rightarrow$ * verb VP	[5-5]
VP $\rightarrow$ * verb	[5-5]
VP $\rightarrow$ verb *	[4-5]
VP	[4-5]
S $\rightarrow$ NP VP *	[1-5]
S	[1-5]
VP $\rightarrow$ verb * VP	[5-6]
VP $\rightarrow$ * verb VP	[6-6]
VP $\rightarrow$ * verb	[6-6]
VP $\rightarrow$ verb *	[5-6]
VP	[5-6]
VP $\rightarrow$ verb VP *	[4-6]
VP	[4-6]
S $\rightarrow$ NP VP *	[1-6]
S	[1-6]

4. (14 pts) Consider the four parse trees below:



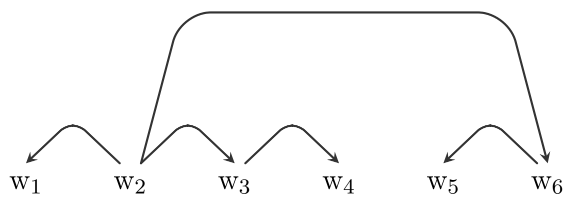
Consider these four parse trees to be a (tiny!) Treebank, and construct a probabilistic context-free grammar (PCFG) from this Treebank. List all distinct context-free grammar rules that are depicted in the trees above and compute their probabilities.

Rules	**Probability**
S → NP VP	.75
S → VP	.25
NP → adj noun	.46
NP → noun noun	.18
NP → art NP	.27
NP → noun	.09
VP → verb PP	.25
VP → verb NP PP	.25
VP → verb VP1	.25
VP → verb VP2	.25
VP1 → inf verb NP	1
VP2 → adv VP1	1
PP → prep NP	1

5. (20 pts) For each sentence below, draw the dependency relations that would be produced by applying the given sequence of shift-reduce operations. w_i means the i^{th} word in the sentence. (The dependency relations will be *unlabeled* directed edges.)

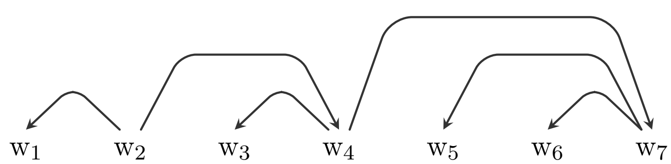
(a) SENTENCE: w_1 w_2 w_3 w_4 w_5 w_6

OPERATIONS: Shift, Shift, LeftArc, Shift, Shift, RightArc, RightArc, Shift, Shift, LeftArc, RightArc



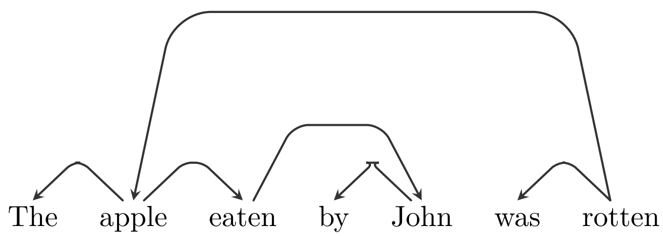
(b) SENTENCE: w_1 w_2 w_3 w_4 w_5 w_6 w_7

OPERATIONS: Shift, Shift, LeftArc, Shift, Shift, LeftArc, Shift, Shift, Shift, LeftArc, LeftArc, RightArc, RightArc

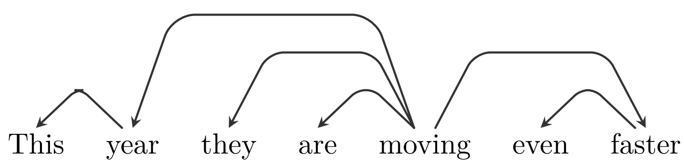


<u>Question #6 is for CS-6340 students ONLY!</u>

6. (13 pts) For each dependency graph below, determine the sequence of shift-reduce operations that should be predicted by the oracle in order to produce the graph.



OPERATIONS: Shift, Shift, LeftArc, Shift, Shift, Shift, LeftAre, RightArc, RightArc, Shift, Shift, LeftArc, LeftArc



OPERATIONS: Shift, Shift, LeftArc, Shift, Shift, Shift, LeftArc, LeftArc, LeftArc, Shift, Shift, LeftArc, RightArc

ELECTRONIC SUBMISSION INSTRUCTIONS

You should submit the answers to this assignment **in pdf format** on our course's CANVAS site by 11:59pm on Thursday, September 20.