CS6320, Fall 2020
Dr. Mithun Balakrishna
Homework 5
Due Sunday, November 8th, 2020 11:59pm

A. Submission Instructions:

- Submit your solutions via eLearning.
- Please submit a single zip file containing **ALL** the relevant homework solution files. The zip filename should follow the pattern "HW#_FirstnameLastname.zip" (Example: HW3_Claire Underwood.zip)
    - **Penalty of 5 points** if not followed
- For all non-programming questions:
    - Please include **ALL** the solutions in a **single** PDF/Doc/PS/Image file
    - The filename should follow the pattern "HW#_FirstnameLastname.FileExtension" (Example: HW3_Claire Underwood.pdf)
    - **Penalty of 5 points** if not followed
- For programming questions:
    - Write the programming solutions in C/C++, Java, or Python. For using any other programming language, please get prior approval from the TA.
    - Include a Readme file with instructions on how to build and run your programming question solution
        - Instructions should be very simple:
            - python bigram.py input_arguments

              OR

              python bigram.py (if the input arguments are hard coded)
        - Hard coding the input arguments to your program is fine unless the TA cannot run your code directly. Do **NOT** include instructions such as: "Please modify the path in my main function. Then copy the training data in the same folder."
        - Provide your training data together unless the dataset is too large.
        - **Penalty of 10 points** if not followed
    - Submit ALL your source code files
        - Do not write your solutions in the readme file
        - **Penalty of 10 points** if not followed
- Late Submission Penalty:
    - up to 2 hours late — 10% deduction
    - 2 - 4 hours late — 20% deduction
    - 4 - 12 hours late — 35% deduction
    - 12 - 24 hours late — 50% deduction
    - 24 - 48 hours late — 75% deduction
    - more than 48 hours late — 100% deduction (zero credit)

B. Problems:

1. Chart Parsing

POS Tag Lexicon:

the: ART
large: ADJ
can: N, AUX, V
hold: N, V
water: N,V

Grammar:

1. S → NP VP
2. NP → ART ADJ N
3. NP → ART N
4. NP → ADJ N
5. VP → AUX VP
6. VP → V NP

Using the above lexicon and grammar rules, manually create all charts for the following sentence applying the bottom-up chart parser:

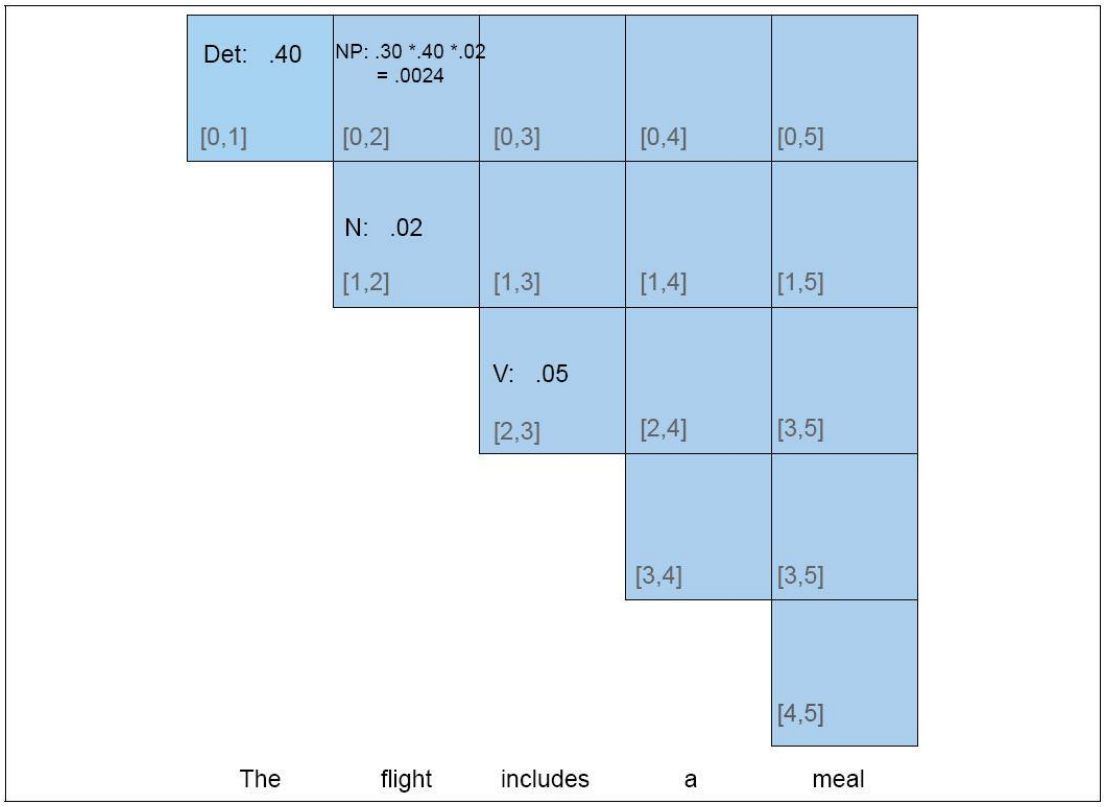
The large can can hold the water

Using the final chart, please draw the parse tree structure(s) for the above sentence.

2. Statistical Parsing (25 points)

$S \rightarrow NP\ VP$	.80	$Det \rightarrow the$	.40
$NP \rightarrow Det\ N$	.30	$Det \rightarrow a$	.40
$VP \rightarrow V\ NP$	.20	$N \rightarrow meal$	.01
$V \rightarrow includes$	.05	$N \rightarrow flight$	.02

Using the above grammar rules, manually fill out the rest of the probabilistic CKY chart in the figure below:


Det: .40
[0,1]
NP: .30 *.40 *.02 = .0024
[0,2]
[0,3]
[0,4]
[0,5]
N: .02
[1,2]
[1,3]
[1,4]
[1,5]
V: .05
[2,3]
[2,4]
[3,5]
[3,4]
[3,5]
[4,5]
The
flight
includes
a
meal