1. (19 points) Assume a bigram language model is trained on the following corpus of sentences using MLE with linear interpolation for smoothing (with the bigram $\lambda$ weight set to 0.9 and the unigram $\lambda$ weight set to 0.1). Since the unigram model does not need to estimate $P(<s>)$, just completely ignore the start token when estimating the unigram model.

- $<s>$ man marries woman $</s>$
- $<s>$ woman marries man $</s>$
- $<s>$ woman marries woman $</s>$
- $<s>$ man divorces woman $</s>$
- $<s>$ woman divorces man $</s>$

What is the estimated probability of the following test string? Show your work. You only need to calculate the parameters of the model sufficient to solve this particular problem.

- $<s>$ man marries man $</s>$
2. (19 points) Consider the HMM below where the transition probabilities are shown in the graph and the observation probabilities (where \( V = \{A, B\} \)) are in the tables below each state.

Use the Forward algorithm to compute the probability of generating the short output string: “A B”.
Show the values computed for each of \( \alpha_t(j) \) parameters as they are computed, showing your work.
3. (19 points) Consider the following probabilistic context-free grammar (PCFG):

\[
\begin{align*}
S &\rightarrow NP \ VP \quad (1.0) \\
NP &\rightarrow NP \ PP \quad (0.4) \\
NP &\rightarrow Det \ N \quad (0.6) \\
VP &\rightarrow VP \ PP \quad (0.5) \\
VP &\rightarrow V \ NP \quad (0.5) \\
PP &\rightarrow Prep \ NP \quad (1.0) \\
Det &\rightarrow the \quad (0.6) \\
Det &\rightarrow a \quad (0.4) \\
Prep &\rightarrow with \quad (1.0) \\
V &\rightarrow kissed \quad (1.0) \\
N &\rightarrow man \quad (0.4) \\
N &\rightarrow woman \quad (0.4) \\
N &\rightarrow dog \quad (0.2)
\end{align*}
\]

Show all parse trees for the sentence “The man kissed the woman with a dog” and compute the probability for each.

Is the semantically most plausible parse tree given the highest probability? Why or why not?
4. (19 points) Consider the following simple PCFG:

\[
\begin{align*}
S &\rightarrow NP \ VP \quad 0.6 \\
S &\rightarrow VP \\np \rightarrow Pronoun \quad 0.4 \\
NP &\rightarrow Pronoun \quad 0.4 \\
NP &\rightarrow PlurNoun \\
NP &\rightarrow PlurNoun \quad 0.6 \\
VP &\rightarrow Verb \\
VP &\rightarrow VP \ PP \\
PP &\rightarrow Prep \ NP \\
\end{align*}
\]

Pronoun $\rightarrow I \quad 0.5$

Pronoun $\rightarrow YOU \quad 0.5$

PlurNoun $\rightarrow FLIES \quad 0.5$

PlurNoun $\rightarrow ARROWS \quad 0.5$

Verb $\rightarrow LIKE \quad 0.6$

Verb $\rightarrow FLIES \quad 0.4$

Prep $\rightarrow LIKE \quad 0.4$

Prep $\rightarrow WITH \quad 0.6$

Use the probabilistic CKY algorithm (the Viterbi analog for PCFGs) to find the most probable parse tree for the sentence:

FLIES LIKE ARROWS

First, below show the changes to productions that require conversion to CNF:

Next, show the triangular CKY table with each cell filled with all its constituents together with their probabilities, showing your work.

Finally, show all final parse trees for this sentence (in the original grammar) with their probabilities.
5. (24 points) Provide short answers (1-3 sentences) for each of the following questions:

Consider the following joke:

Loy: Let’s eat up the street.
Roy: No, thanks; I don’t like concrete.

Explain what specific type of ambiguity in language understanding makes this humorous.

Do the same for the following joke:

Etta: Did you hear about the kidnapping in Dallas?
Gretta: No; what happened?
Etta: They woke him up.

What sequence of numbers (which has no closed form) characterizes the number of parses of simple English sentence ending in $n$ prepositional phrases and what closed form expression in $n$ is an informative lower bound on this number.

What is the primary difference between natural languages and computer languages?
What is the basic idea behind a “back-off” approach to smoothing?

Why is EM called an “any time” algorithm?

Why does CKY parsing take $O(n^3)$ time, where $n$ is the length of the sentence?

How does a PCFG parser help resolve syntactic ambiguity compared to using a normal CFG?

(Extra credit, 2pts) What university named their school of engineering after the founder of Qualcomm and why?
(Extra credit, 2pts) What is probably the best name for the so-called “Gaussian” probability distribution and why?

(Extra credit, 2pts) Who is the world’s most highly-cited living author?