SENG 480 - NLP
Assignment 2
Joe Czepil
V00774878

Question 1:

PP - Prepositional Phrase
IN - Preposition
NNP - Proper Noun
NN - Noun
VP - Verb Phrase
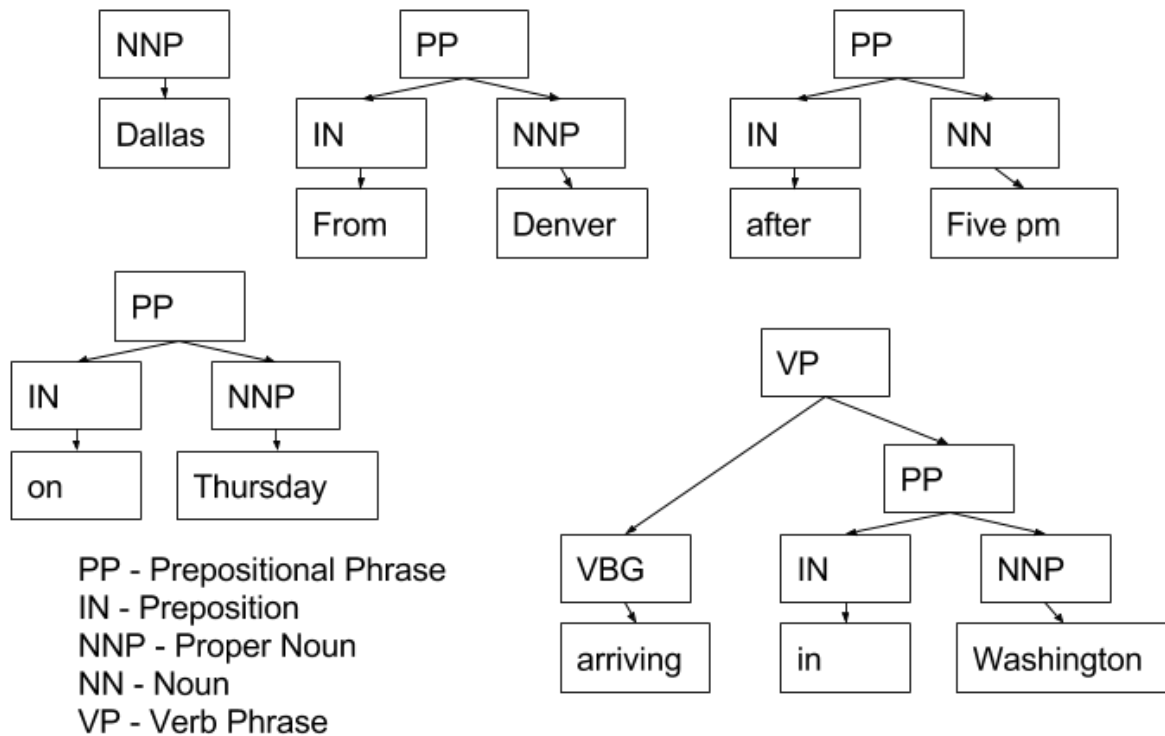
Question 2:

I believe the sentences generated by the PCFG will generally have a better structure that sentences generated by bigram and trigram models. However, bigram and trigram generated sentences will likely make more sense in context. Sentences generated by PCFGs only rely on the current non terminal of the parse tree to determine the following terminals and non terminals. Therefore, the resulting sentences will likely have good structure but the final terminals will only be relying on their parent and nothing else. Trigrams will definitely be better than bigrams if they are trained on good data and both trigrams and bigrams will occasionally return sentences that might make a bit of sense but it is unlikely that they will have good structure.

Question 3:

a)

S -> NP VP NP -> John NP -> Sally NP -> Bill NP -> Fred NP -> Jeff VP -> V1 SBAR VP -> VP ADVP VP -> V2 V1 -> said	V1 -> declared V1 -> pronounced SBAR -> COMP S COMP -> that ADVP -> loudly ADVP -> quickly ADVP -> elegantly V2 -> snored V2 -> ran V2 -> swam

b)

S_0 -> NP VP S -> NP VP NP -> John NP -> Sally NP -> Bill NP -> Fred NP -> Jeff VP -> V1 SBAR **VP -> V2 ADVP** V1 -> said	V1 -> declared V1 -> pronounced SBAR -> COMP S COMP -> that ADVP -> loudly ADVP -> quickly ADVP -> elegantly V2 -> snored V2 -> ran V2 -> swam

c)

S -> NP VP	[1]	V1 -> declared	[1/3]
NP -> John	[1/6]	V1 -> pronounced	[1/3]
NP -> Sally	[2/6]	SBAR -> COMP S	[1]
NP -> Bill	[1/6]	COMP -> that	[1]
NP -> Fred	[1/6]	ADVP -> loudly	[1/3]
NP -> Jeff	[1/6]	ADVP -> quickly	[1/3]
VP -> V1 SBAR	[1/3]	ADVP -> elegantly	[1/3]
VP -> VP ADVP	[1/3]	V2 -> snored	[1/3]
VP -> V2	[1/3]	V2 -> ran	[1/3]
		V2 -> swam	[1/3]

V1 -> said	[1/3]		[1/3]

d)

[S -> NP VP] * [NP -> Jeff] * [VP -> V1 SBAR] * [V1 -> pronounced] * [SBAR -> COMP S] * [COMP -> that] * [S -> NP VP] * [NP -> Bill] * [VP -> VP ADVP] * [VP -> V2] * [V2 -> ran] * [ADVP -> elegantly] =
[1] * [⅙] * [⅓] * [⅓] * [1] * [1] * [1] * [⅙] * [⅓] * [⅓] * [⅓] * [⅓] =3.81x10^-5

[S -> NP VP] * [VP -> VP ADVP] * [ADVP -> elegantly] * [NP -> Jeff] * [VP -> V1 SBAR] * [V1 -> pronounced] * [SBAR -> COMP S] * [COMP -> that] * [S -> NP VP] * [NP -> Bill] * [VP -> V2] * [V2 -> ran] = 3.81x10^-5

Question 4:

q4a.py
Counter-fitted
Top 5: 0.797446808511
Top 1: 0.105319148936
9400

Glove.6B.50d
Top 5: 0.807868920893
Top 1: 0.199958255062
9582

q4b.py
Counter-fitted
Correlation: 0.73533409657174387
Pvalue: 1.7874476167260104x10^-170

Glove.6B.50d
Correlation: 0.26457921929908129
Pvalue: 1.8305378556356217x10^-17