HW3 Report

Task 1

For grammar2 and lexicon, we can see that both of them have one probability for each word and gramma. So this is a UniformModel.

Task 2

```
jlai@mars:~/ECS189G_HW3$ perl cfgparse.pl grammar1 grammar2 lexicon < examples.sen
4.713e-06 9.427e-06 0.500 (ROOT (S1 (NP (Proper Arthur)) (VP (VerbT is) (NP (Det the) (Nbar (Noun king)))) .))
7.694e-11 1.539e-10 0.500 (ROOT (S1 (NP (Proper Arthur)) (VP (VerbT rides) (NP (Det the) (Nbar (Nbar (Noun horse)
) (PP (Prep near) (NP (Det the) (Nbar (Noun castle)))))) .))
1.150e-20 2.301e-20 0.500 (ROOT (S2 (+Misc (Misc riding) (+Misc (Misc to) (+Misc (Misc Camelot) (+VerbT (VerbT is
) (+Misc (Misc hard) (+Misc (Misc .))))))))))
1.238e-15 2.476e-15 0.500 (ROOT (S2 (+Misc (Misc do) (+Misc (Misc coconuts) (+Misc (Misc speak) (+Misc (Misc ?)))
))))
4.421e-17 8.842e-17 0.500 (ROOT (S2 (+Misc (Misc what) (+Misc (Misc does) (+Proper (Proper Arthur) (+Misc (Misc r
ide) (+Misc (Misc ?)))))))))
2.694e-23 5.388e-23 0.500 (ROOT (S2 (+Misc (Misc who) (+Misc (Misc does) (+Proper (Proper Arthur) (+Misc (Misc su
ggest) (+Misc (Misc she) (+Misc (Misc carry) (+Misc (Misc ?))))))))))
1.043e-22 2.087e-22 0.500 (ROOT (S2 (+Misc (Misc why) (+Misc (Misc does) (+Misc (Misc England) (+Misc (Misc have)
 (+Det (Det a) (+Noun (Noun king) (+Misc (Misc ?)))))))))))
1.052e-26 2.103e-26 0.500 (ROOT (S2 (+Misc (Misc are) (+Misc (Misc they) (+Misc (Misc suggesting) (+Proper (Prope
r Arthur) (+Misc (Misc ride) (+Misc (Misc to) (+Misc (Misc Camelot) (+Misc (Misc ?))))))))))))
7.826e-23 1.565e-22 0.500 (ROOT (S2 (+Misc (Misc five) (+Misc (Misc strangers) (+Misc (Misc are) (+Prep (Prep at)
 (+Det (Det the) (+Misc (Misc Round Table) (+Misc (Misc .)))))))))
2.211e-17 4.421e-17 0.500 (ROOT (S2 (+Proper (Proper Guinevere) (+Misc (Misc might) (+Misc (Misc have) (+Misc (Mi
sc known) (+Misc (Misc .)))))))))
2.014e-21 4.028e-21 0.500 (ROOT (S2 (+Proper (Proper Guinevere) (+Misc (Misc should) (+Misc (Misc be) (+Misc (Mis
c riding) (+Prep (Prep with) (+Proper (Proper Patsy) (+Misc (Misc .)))))))))
4.699e-21 9.398e-21 0.500 (ROOT (S2 (+Misc (Misc it) (+VerbT (VerbT is) (+Proper (Proper Sir Lancelot) (+Misc (Mi
sc who) (+Misc (Misc knows) (+Proper (Proper Zoot) (+Misc (Misc !)))))))))
1.541e-22 3.081e-22 0.500 (ROOT (S2 (+Misc (Misc either) (+Proper (Proper Arthur) (+Misc (Misc knows) (+Misc (Mis
c or) (+Proper (Proper Patsy) (+Misc (Misc does) (+Misc (Misc .)))))))))
2.455e-27 4.909e-27 0.500 (ROOT (S2 (+Misc (Misc neither) (+Proper (Proper Sir Lancelot) (+Misc (Misc nor) (+Prop
er (Proper Guinevere) (+Misc (Misc will) (+Misc (Misc speak) (+Prep (Prep of) (+Misc (Misc it) (+Misc (Misc .))))
))))))))
1.690e-26 3.380e-26 0.500 (ROOT (S2 (+Det (Det the) (+Misc (Misc Holy Grail) (+Misc (Misc was) (+Misc (Misc cover
ed) (+Prep (Prep by) (+Det (Det a) (+Misc (Misc yellow) (+Noun (Noun fruit) (+Misc (Misc .)))))))))))
3.803e-26 7.605e-26 0.500 (ROOT (S2 (+Proper (Proper Zoot) (+Misc (Misc might) (+Misc (Misc have) (+Misc (Misc be
en) (+Misc (Misc carried) (+Prep (Prep by) (+Det (Det a) (+Noun (Noun swallow) (+Misc (Misc .)))))))))))
3.650e-31 7.300e-31 0.500 (ROOT (S2 (+Proper (Proper Arthur) (+Misc (Misc rode) (+Misc (Misc to) (+Misc (Misc Cam
elot) (+Misc (Misc and) (+Misc (Misc drank) (+Prep (Prep from) (+Misc (Misc his) (+Noun (Noun chalice) (+Misc (Mi
sc .)))))))))))))
1.121e-33 2.241e-33 0.500 (ROOT (S2 (+Misc (Misc they) (+Misc (Misc migrate) (+Misc (Misc precisely) (+Misc (Misc
 because) (+Misc (Misc they) (+Misc (Misc know) (+Misc (Misc they) (+Misc (Misc will) (+Misc (Misc grow) (+Misc (
Misc .)))))))))))))
1.238e-15 2.476e-15 0.500 (ROOT (S2 (+Misc (Misc do) (+Misc (Misc not) (+Misc (Misc speak) (+Misc (Misc !))))))))
1.073e-28 2.146e-28 0.500 (ROOT (S2 (+Proper (Proper Arthur) (+Misc (Misc will) (+Misc (Misc have) (+Misc (Misc b
een) (+Misc (Misc riding) (+Prep (Prep for) (+Misc (Misc eight) (+Misc (Misc nights) (+Misc (Misc .)))))))))))
1.732e-29 3.463e-29 0.500 (ROOT (S2 (+Proper (Proper Arthur) (+Misc (Misc ,) (+Misc (Misc sixty) (+Misc (Misc inc
hes) (+Misc (Misc ,) (+VerbT (VerbT is) (+Det (Det a) (+Misc (Misc tiny) (+Noun (Noun king) (+Misc (Misc .)))))))
)))))
2.131e-23 4.262e-23 0.500 (ROOT (S2 (+Proper (Proper Arthur) (+Misc (Misc knows) (+Proper (Proper Patsy) (+Misc (
Misc ,) (+Det (Det the) (+Misc (Misc trusty) (+Noun (Noun servant) (+Misc (Misc .)))))))))))
3.947e-19 7.895e-19 0.500 (ROOT (S2 (+Proper (Proper Arthur) (+Misc (Misc and) (+Proper (Proper Guinevere) (+Misc
 (Misc migrate) (+Misc (Misc frequently) (+Misc (Misc .)))))))))
6.489e-31 1.298e-30 0.500 (ROOT (S2 (+Misc (Misc he) (+Misc (Misc knows) (+Misc (Misc what) (+Misc (Misc they) (+
Misc (Misc are) (+Prep (Prep with) (+Det (Det that) (+Noun (Noun story) (+Misc (Misc .))))))))))
))))))))
```

When running the code with only gramma1, it can only parse the first two sentences successfully but the rest cannot parse. When running the code with also gramma2, all sentences could be parsed successfully.

There are several reasons accounting for this difference. Firstly, the grammar1 dose not cover all POS tags and has little rules, like it has no Misc. Secondly, considering examples.sen, there are many sentences ending with question mark and exclamation. The grammar1 has few rules for these while the grammer2 has Misc tag. So there are fewer lexicon rules and recursive grammar rules in the garmmar1 file. Hence merging with grammr2, all of the sentences could be parsed successfully.

Task 3

```
jlai@mars:~/ECS189G_HW3$ perl cfggen.pl --text 10 grammar1 lexicon
1:   another winter rides no husk .
2:   a swallow covers a fruit .
3:   that story carries another sovereign .
4:   no servant carries any servant .
5:   no corner rides every home .
6:   that king rides this king .
7:   a weight from that fruit for no king rides another husk .
8:   each land below every winter is any winter .
9:   no horse carries the husk .
10:  a castle rides the sun .
```

```
jlai@mars:~/ECS189G_HW3$ perl cfggen.pl --text 10 grammar2 lexicon
1:  Patsy Arthur home drinks
2:  Sir Lancelot rides successfully over carries Patsy has unable near above each story
3:  ants this any husk
4:  story every Dingo any pound rides coconut
5:  that
6:  drinks goes rides the another swallow successfully has Guinevere
7:  sovereign
8:  every below covers Uther Pendragon through pound unfortunately swallow has coconut chalice across trustier de
feater Dingo of no migrated sun carries fruit rides Arthur every no Patsy chalice every swallow covers Sir Bedeve
re no the by another Sir Lancelot above at might
9:  sovereign is
10:  king how every corner sun trustiest carries sun Arthur husk chalice coconut Dingo that each
```

```
jlai@mars:~/ECS189G_HW3$ perl cfggen.pl --text 10 grammar1 grammar2 lexicon
1:   every quest has this master .
2:   each quest across every chalice has another quest .
3:   each land is another sun .
4:   Sir Lancelot is a land .
5:   each husk carries Zoot .
6:   a chalice has Guinevere .
7:   each chalice carries that sovereign .
8:   the winter has each swallow .
9:   no home at the sovereign carries each fruit .
10:  this king is each husk .
```

Using only grammar1, we can see the sentences have almost the same length, which is five words. There are few very short sentences which has less than four words or very long sentences which has more than seven word. And the length cannot be very long because grammar1 has no recursive rules.

Using only grammar2, we can see the sentences have different lengths. There are very short sentences and also very long sentences because there are many recursive rules in grammar2.

Using both grammar1 and grammar2, we can see that there are difference lengths of sentences but mostly are five-word, which indicates that grammar1 is called more frequently because ROOT derives into S1 in 99 times out of 100 times basically. And grammar2 is called less frequently. So there exists some different length sentences.

And the probabilities are also changed as the weights from grammar1 and grammar2 are merged.

Task 4

I use NLTK package of Python to generate my production rules and lexicon rules and combine Penn Treebank and Brown's production rules. But it seems not ideal. For example, they have many rules with same probabilities. So I amplify some of the probability of the regular rules. Besides, some relation tags are too complex that I make them into groups together, like grouping JJR, JJS into JJ.

Task 5

There are 6 sentences are grammatically correct in the 20 sentences.