Assignment 3

Task 1

❑ Run:

./cfgparse.pl grammar2 lexicon < examples.sen

```
PS D:\ucdavis\ECS 189G\Homework\hw3> Get-Content examples.sen |perl ./cfgparse.pl grammar2 lexicon
5.973e-012 1.195e-011 0.500 (ROOT (S2 (+Proper (Proper Arthur) (+VerbT (VerbT is) (+Det (Det the) (+Noun (Noun king) (+Misc
 (Misc .))))))))
6.581e-018 1.316e-017 0.500 (ROOT (S2 (+Proper (Proper Arthur) (+VerbT (VerbT rides) (+Det (Det the) (+Noun (Noun horse) (+
Prep (Prep near) (+Det (Det the) (+Noun (Noun castle) (+Misc (Misc .)))))))))))
1.150e-018 2.301e-018 0.500 (ROOT (S2 (+Misc (Misc riding) (+Misc (Misc to) (+Misc (Misc Camelot) (+VerbT (VerbT is) (+Misc
 (Misc hard) (+Misc (Misc .)))))))))
1.238e-013 2.476e-013 0.500 (ROOT (S2 (+Misc (Misc do) (+Misc (Misc coconuts) (+Misc (Misc speak) (+Misc (Misc ?)))))))
4.421e-015 8.842e-015 0.500 (ROOT (S2 (+Misc (Misc what) (+Misc (Misc does) (+Proper (Proper Arthur) (+Misc (Misc ride) (+M
isc (Misc ?))))))))
2.694e-021 5.388e-021 0.500 (ROOT (S2 (+Misc (Misc who) (+Misc (Misc does) (+Proper (Proper Arthur) (+Misc (Misc suggest) (
+Misc (Misc she) (+Misc (Misc carry) (+Misc (Misc ?)))))))))
1.043e-020 2.087e-020 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-024 2.103e-024 0.500 (ROOT (S2 (+Misc (Misc are) (+Misc (Misc they) (+Misc (Misc suggesting) (+Proper (Proper Arthur
) (+Misc (Misc ride) (+Misc (Misc to) (+Misc (Misc Camelot) (+Misc (Misc ?)))))))))))
7.826e-021 1.565e-020 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-015 4.421e-015 0.500 (ROOT (S2 (+Proper (Proper Guinevere) (+Misc (Misc might) (+Misc (Misc have) (+Misc (Misc known
) (+Misc (Misc .))))))))
2.014e-019 4.028e-019 0.500 (ROOT (S2 (+Proper (Proper Guinevere) (+Misc (Misc should) (+Misc (Misc be) (+Misc (Misc riding
) (+Prep (Prep with) (+Proper (Proper Patsy) (+Misc (Misc .)))))))))))
4.699e-019 9.398e-019 0.500 (ROOT (S2 (+Misc (Misc it) (+VerbT (VerbT is) (+Proper (Proper Sir Lancelot) (+Misc (Misc who)
(+Misc (Misc knows) (+Proper (Proper Zoot) (+Misc (Misc !)))))))))))
1.541e-020 3.081e-020 0.500 (ROOT (S2 (+Misc (Misc either) (+Proper (Proper Arthur) (+Misc (Misc knows) (+Misc (Misc or) (+
Proper (Proper Patsy) (+Misc (Misc does) (+Misc (Misc .)))))))))))
2.455e-025 4.909e-025 0.500 (ROOT (S2 (+Misc (Misc neither) (+Proper (Proper Sir Lancelot) (+Misc (Misc nor) (+Proper (Prop
er Guinevere) (+Misc (Misc will) (+Misc (Misc speak) (+Prep (Prep of) (+Misc (Misc it) (+Misc (Misc .)))))))))))
1.690e-024 3.380e-024 0.500 (ROOT (S2 (+Det (Det the) (+Misc (Misc Holy Grail) (+Misc (Misc was) (+Misc (Misc covered) (+Pr
ep (Prep by) (+Det (Det a) (+Misc (Misc yellow) (+Noun (Noun fruit) (+Misc (Misc .)))))))))))
3.803e-024 7.605e-024 0.500 (ROOT (S2 (+Proper (Proper Zoot) (+Misc (Misc might) (+Misc (Misc have) (+Misc (Misc been) (+Mi
sc (Misc carried) (+Prep (Prep by) (+Det (Det a) (+Noun (Noun swallow) (+Misc (Misc .)))))))))))
3.650e-029 7.300e-029 0.500 (ROOT (S2 (+Proper (Proper Arthur) (+Misc (Misc rode) (+Misc (Misc to) (+Misc (Misc Camelot) (+
Misc (Misc and) (+Misc (Misc drank) (+Prep (Prep from) (+Misc (Misc his) (+Noun (Noun chalice) (+Misc (Misc .)))))))))))))
1.121e-031 2.241e-031 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-013 2.476e-013 0.500 (ROOT (S2 (+Misc (Misc do) (+Misc (Misc not) (+Misc (Misc speak) (+Misc (Misc !)))))))
1.073e-026 2.146e-026 0.500 (ROOT (S2 (+Proper (Proper Arthur) (+Misc (Misc will) (+Misc (Misc have) (+Misc (Misc been) (+M
isc (Misc riding) (+Prep (Prep for) (+Misc (Misc eight) (+Misc (Misc nights) (+Misc (Misc .)))))))))))))
1.732e-027 3.463e-027 0.500 (ROOT (S2 (+Proper (Proper Arthur) (+Misc (Misc ,) (+Misc (Misc sixty) (+Misc (Misc inches) (+M
isc (Misc ,) (+VerbT (VerbT is) (+Det (Det a) (+Misc (Misc tiny) (+Noun (Noun king) (+Misc (Misc .)))))))))))))
2.131e-021 4.262e-021 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-017 7.895e-017 0.500 (ROOT (S2 (+Proper (Proper Arthur) (+Misc (Misc and) (+Proper (Proper Guinevere) (+Misc (Misc m
igrate) (+Misc (Misc frequently) (+Misc (Misc .)))))))))
6.489e-029 1.298e-028 0.500 (ROOT (S2 (+Misc (Misc he) (+Misc (Misc knows) (+Misc (Misc what) (+Misc (Misc they) (+Misc (Mi
```

❑ What kind of language model does this PCFG implement? Give your thoughts.

It implements unigram (uniform) model.

Looking into the grammar2, we could see that it assign equal probability to every possible parse rule. Also looking into the cfgparse.pl, we know the CKY inside algorithm is run alongside the "best tree" algorithm. When finally computing The probability of a parse and the probability of a sentence, the formula it uses can be simplified to:

$$c*p(tag->w\_1)*p(tag->w\_2)*...*p(tag->w\_n)$$

where c is a small constant that can be ignored, and n is the length of the sentence. The remaining formula is the same as the format of unigram and since

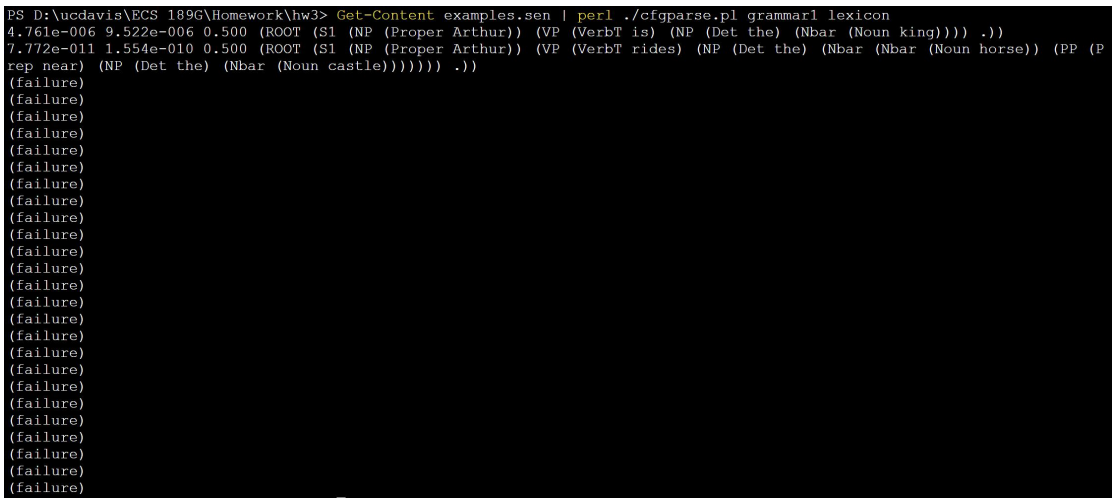
$$p(tag->w\_1) = p(tag->w\_2) = ... = p(tag->w\_n)$$

It it clear that it implements the uniform model.

Task 2

❑ Compare the outcome when you run:

./cfgparse.pl grammar1 lexicon < examples.sen



./cfgparse.pl grammar1 grammar2 lexicon < examples.sen

```
PS D:\ucdavis\ECS 189G\Homework\hw3> Get-Content examples.sen | perl ./cfgparse.pl grammar1 grammar2 lexicon
4.713e-006 9.427e-006 0.500 (ROOT (S1 (NP (Proper Arthur)) (VP (VerbT is) (NP (Det the) (Nbar (Noun king)))) .))
7.694e-011 1.539e-010 0.500 (ROOT (S1 (NP (Proper Arthur)) (VP (VerbT rides) (NP (Det the) (Nbar (Nbar (Noun horse)) (PP (P
rep near) (NP (Det the) (Nbar (Noun castle)))))) .))
1.150e-020 2.301e-020 0.500 (ROOT (S2 (+Misc (Misc riding) (+Misc (Misc to) (+Misc (Misc Camelot) (+VerbT (VerbT is) (+Misc
 (Misc hard) (+Misc (Misc .)))))))))
1.238e-015 2.476e-015 0.500 (ROOT (S2 (+Misc (Misc do) (+Misc (Misc coconuts) (+Misc (Misc speak) (+Misc (Misc ?)))))))
4.421e-017 8.842e-017 0.500 (ROOT (S2 (+Misc (Misc what) (+Misc (Misc does) (+Proper (Proper Arthur) (+Misc (Misc ride) (+M
isc (Misc ?)))))))
2.694e-023 5.388e-023 0.500 (ROOT (S2 (+Misc (Misc who) (+Misc (Misc does) (+Proper (Proper Arthur) (+Misc (Misc suggest) (
+Misc (Misc she) (+Misc (Misc carry) (+Misc (Misc ?)))))))))
1.043e-022 2.087e-022 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-026 2.103e-026 0.500 (ROOT (S2 (+Misc (Misc are) (+Misc (Misc they) (+Misc (Misc suggesting) (+Proper (Proper Arthur
) (+Misc (Misc ride) (+Misc (Misc to) (+Misc (Misc Camelot) (+Misc (Misc ?)))))))))
7.826e-023 1.565e-022 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-017 4.421e-017 0.500 (ROOT (S2 (+Proper (Proper Guinevere) (+Misc (Misc might) (+Misc (Misc have) (+Misc (Misc known
) (+Misc (Misc .)))))))
2.014e-021 4.028e-021 0.500 (ROOT (S2 (+Proper (Proper Guinevere) (+Misc (Misc should) (+Misc (Misc be) (+Misc (Misc riding
) (+Prep (Prep with) (+Proper (Proper Patsy) (+Misc (Misc .)))))))))
4.699e-021 9.398e-021 0.500 (ROOT (S2 (+Misc (Misc it) (+VerbT (VerbT is) (+Proper (Proper Sir Lancelot) (+Misc (Misc who)
(+Misc (Misc knows) (+Proper (Proper Zoot) (+Misc (Misc !)))))))))
1.541e-022 3.081e-022 0.500 (ROOT (S2 (+Misc (Misc either) (+Proper (Proper Arthur) (+Misc (Misc knows) (+Misc (Misc or) (+
Proper (Proper Patsy) (+Misc (Misc does) (+Misc (Misc .)))))))))
2.455e-027 4.909e-027 0.500 (ROOT (S2 (+Misc (Misc neither) (+Proper (Proper Sir Lancelot) (+Misc (Misc nor) (+Proper (Prop
er Guinevere) (+Misc (Misc will) (+Misc (Misc speak) (+Prep (Prep of) (+Misc (Misc it) (+Misc (Misc .))))))))))))
1.690e-026 3.380e-026 0.500 (ROOT (S2 (+Det (Det the) (+Misc (Misc Holy Grail) (+Misc (Misc was) (+Misc (Misc covered) (+Pr
ep (Prep by) (+Det (Det a) (+Misc (Misc yellow) (+Noun (Noun fruit) (+Misc (Misc .))))))))))))
3.803e-026 7.605e-026 0.500 (ROOT (S2 (+Proper (Proper Zoot) (+Misc (Misc might) (+Misc (Misc have) (+Misc (Misc been) (+Mi
sc (Misc carried) (+Prep (Prep by) (+Det (Det a) (+Noun (Noun swallow) (+Misc (Misc .))))))))))))
3.650e-031 7.300e-031 0.500 (ROOT (S2 (+Proper (Proper Arthur) (+Misc (Misc rode) (+Misc (Misc to) (+Misc (Misc Camelot) (+
Misc (Misc and) (+Misc (Misc drank) (+Prep (Prep from) (+Misc (Misc his) (+Noun (Noun chalice) (+Misc (Misc .))))))))))))
1.121e-033 2.241e-033 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-015 2.476e-015 0.500 (ROOT (S2 (+Misc (Misc do) (+Misc (Misc not) (+Misc (Misc speak) (+Misc (Misc !)))))))
1.073e-028 2.146e-028 0.500 (ROOT (S2 (+Proper (Proper Arthur) (+Misc (Misc will) (+Misc (Misc have) (+Misc (Misc been) (+M
isc (Misc riding) (+Prep (Prep for) (+Misc (Misc eight) (+Misc (Misc nights) (+Misc (Misc .))))))))))))
1.732e-029 3.463e-029 0.500 (ROOT (S2 (+Proper (Proper Arthur) (+Misc (Misc ,) (+Misc (Misc sixty) (+Misc (Misc inches) (+M
isc (Misc ,) (+VerbT (VerbT is) (+Det (Det a) (+Misc (Misc tiny) (+Noun (Noun king) (+Misc (Misc .))))))))))))
2.131e-023 4.262e-023 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-019 7.895e-019 0.500 (ROOT (S2 (+Proper (Proper Arthur) (+Misc (Misc and) (+Proper (Proper Guinevere) (+Misc (Misc m
igrate) (+Misc (Misc frequently) (+Misc (Misc .)))))))))
6.489e-031 1.298e-030 0.500 (ROOT (S2 (+Misc (Misc he) (+Misc (Misc knows) (+Misc (Misc what) (+Misc (Misc they) (+Misc (Mi
sc are) (+Misc (Misc covering) (+Prep (Prep with) (+Det (Det that) (+Noun (Noun story) (+Misc (Misc .))))))))))))
```

❑ Explain what's going on

If only use the grammar rules in grammar1, the algorithm will only be able to successfully parse the first two sentences and fail for the rest. However if use both grammar rules in grammar1 and grammar2, the algorithm will be able to parse all the sentences given in the example.sen. It is because grammar1 didn't include the rule that needed to parse the those sentence (i.e. it didn't have any rule for the "Misc" tag).

Task 3

❑ Compare the outcome when you run:

./cfggen.pl --text <N> grammar1 lexicon

```
PS D:\ucdavis\ECS 189G\Homework\hw3> perl ./cfggen.pl --text 10 grammar1 lexicon
1:  the sovereign rides each corner .
2:  that fruit drinks any winter .
3:  each horse has this coconut .
4:  that swallow covers another winter .
5:  every weight covers this corner .
6:  that home is no pound .
7:  that master rides the corner .
8:  this coconut covers any master .
9:  each corner carries another story .
10:  the master covers that story .
```

./cfggen.pl --text <N> grammar2 lexicon

```
PS D:\ucdavis\ECS 189G\Homework\hw3> perl ./cfggen.pl --text 10 grammar2 lexicon
1:  cover Dingo has either Sir Lancelot every the below tiny Dingo
2:  into king quest with
3:  any at swallow migrated drinking tropical Guinevere Sir Bedevere above home a sovereign over with horse covers any by k
ing suggest Zoot is that from chalice horse areas Patsy Patsy this
4:  home corner quest Zoot for another Patsy across with through is Guinevere has Patsy indigenous is England five the that

5:  Guinevere sovereign
6:  the covers Zoot Arthur Arthur cover migrate this every have
7:  below for coconut hard another his with through through
8:  another across from each is trustier near a sun know carries Guinevere home 5,000 Uther Pendragon any Dingo which trust
iest carries this by
9:  five carries rides their would areas whose
10:  Arthur has has the Sir Bedevere Dingo home
```

./cfggen.pl --text <N> grammar1 grammar2 lexicon

```
PS D:\ucdavis\ECS 189G\Homework\hw3> perl ./cfggen.pl --text 10 grammar1 grammar2 lexicon
1:  each sun drinks each quest .
2:  that corner covers this chalice .
3:  every sun has a land .
4:  the swallow has a chalice .
5:  each sovereign rides no husk .
6:  Dingo rides the quest .
7:  this home rides a master .
8:  each story covers another coconut .
9:  no chalice covers a horse .
10:  Sir Bedevere rides any winter .
```

❑ Explain what's going on

When using grammar1 alone to generate sentences, due to the limit rules it can use, almost all the sentences generated are grammatically correct. The structure of the sentence are also simple.

When using grammar2 alone, we can see that the proportion of grammatically correct sentences is very low, and there's no punctuation at the end of each sentence. It is because the number of rules in grammar2 are much larger (less constrains) and most of them are not correct. Also, those incorrect rule are simply given a equal probability compare to the correct ones which can lead to a worse performance.

When using grammar1 and grammar2 together, the result is very similar to that provided by grammar1. The reason is the probability of *Root -> S1(99)* is much

larger *Root -> S2(1)*, so the algorithm will in end have a 99/100 possibility to chose the rules in grammar1 to generate a sentence but only have a 1/100 possibility to chose grammar2, which explain the result.

Task 4

❑ **Explain the strategies you used to design your grammar.**

First, I made a more detailed division of all words that currently labeled "Misc" according to their different parts of speech.

I added a total of 15 new tags which is shown below:

Tag	**Example**
VBE	is, are
VBN	grown, spoken
VBG	knowing
CC	and, because
CD	5,000
RB	currently
JJ	bloody
JJR	hotter
JJS	hardest
Pron	he, she
PRP	him, them
PRPS	his, her, their

MD	will, would
Punc	--, ?, !

I added all possible mappings between those new tags and existing tags to meet the requirement, so it will never fail to parse a sentence.

Then, to improve the parse performance, I manually assign higher probability to those rule which align with the real world grammatical rules (i.e. PRPS(i.e. his) -> Noun(i.e. castle)) and lower probability to those that hardly occur in real world (i.e. PRPS(i.e. his) -> PRPS(i.e. her)).

After doing that, I cross entropy finally decrease to 103.637 compare to 144.721 when using grammar1 and grammar2 together.

Task 5

- **Tell us what fraction of your 20 sentences you decided were grammatical.**

The ratio of grammatically correct sentences in all generated sentences is 6/20.