

# Report of Assignment 2

## Language modelling exercise

Devbrat Anuragi

17078

Q1. Vocabulary:= ['<s>', 'peter', 'piper', 'picked', 'a', 'peck', 'of', 'pickled', 'peppers', '</s>', 'if', 'where', 's', 'the']

Q2 . unigram with highest probability is '<s>' with probability 0.09302325581395349  
unigram with second highest Probability 'peter' with probability 0.09302325581395349

Actually <s>, peter, piper, picked, peck, of, pickled, peppers, </s> has the same highest probability, followed by a, if, where, s, the

Q3.

	<s>	peter	piper	picked	a	peck	of	pickled	peppers	</s>	if	where	s	the
<s>	0.0	0.25	0.0	0.0	0.25	0.0	0.0	0.0	0.0	0.0	0.25	0.25	0.0	0.0
peter	0.0	0.00	1.0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.0	0.0
piper	0.0	0.00	0.0	1.0	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.0	0.0
picked	0.0	0.00	0.0	0.0	0.50	0.0	0.0	0.0	0.0	0.5	0.00	0.00	0.0	0.0
a	0.0	0.00	0.0	0.0	0.00	1.0	0.0	0.0	0.0	0.0	0.00	0.00	0.0	0.0
peck	0.0	0.00	0.0	0.0	0.00	0.0	1.0	0.0	0.0	0.0	0.00	0.00	0.0	0.0
of	0.0	0.00	0.0	0.0	0.00	0.0	0.0	1.0	0.0	0.0	0.00	0.00	0.0	0.0
pickled	0.0	0.00	0.0	0.0	0.00	0.0	0.0	0.0	1.0	0.0	0.00	0.00	0.0	0.0
peppers	0.0	0.50	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.5	0.00	0.00	0.0	0.0
</s>	0.0	0.00	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.0	0.0
if	0.0	1.00	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.0	0.0
where	0.0	0.00	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	1.0	0.0
s	0.0	0.00	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.0	1.0
the	0.0	0.00	0.0	0.0	0.00	1.0	0.0	0.0	0.0	0.0	0.00	0.00	0.0	0.0

Q4.

	Bigram	Bigram Count	Bigram Prob
3	s the	1	1
5	if peter	1	1
16	of pickled	4	1
9	peter piper	4	1
1	peck of	4	1
12	pickled peppers	4	1
13	the peck	1	1
14	where s	1	1
15	a peck	3	1
8	piper picked	4	1

Q5.

[illegible]

Q6. {

```
"</s> </s>": 0.05555555555555555,  
"</s> <s>": 0.05555555555555555,  
"</s> a": 0.05555555555555555,  
"</s> if": 0.05555555555555555,  
"</s> of": 0.05555555555555555,  
"</s> peck": 0.05555555555555555,  
"</s> peppers": 0.05555555555555555,  
"</s> peter": 0.05555555555555555,  
"</s> picked": 0.05555555555555555,  
"</s> pickled": 0.05555555555555555,  
"</s> piper": 0.05555555555555555,  
"</s> s": 0.05555555555555555,  
"</s> the": 0.05555555555555555,  
"</s> where": 0.05555555555555555,  
"<s> </s>": 0.05555555555555555,  
"<s> <s>": 0.05555555555555555,  
"<s> a": 0.05555555555555555,  
"<s> if": 0.05555555555555555,  
"<s> of": 0.05555555555555555,  
"<s> peck": 0.05555555555555555,  
"<s> peppers": 0.05555555555555555,  
"<s> peter": 0.05555555555555555,  
"<s> picked": 0.05555555555555555,  
"<s> pickled": 0.05555555555555555,  
"<s> piper": 0.05555555555555555,  
"<s> s": 0.05555555555555555,  
"<s> the": 0.05555555555555555,  
"<s> where": 0.05555555555555555,  
"a </s>": 0.058823529411764705,  
"a <s>": 0.058823529411764705,  
"a a": 0.058823529411764705,  
"a if": 0.058823529411764705,  
"a of": 0.058823529411764705,  
"a peppers": 0.058823529411764705,  
"a peter": 0.058823529411764705,  
"a picked": 0.058823529411764705,  
"a pickled": 0.058823529411764705,  
"a piper": 0.058823529411764705,  
"a s": 0.058823529411764705,  
"a the": 0.058823529411764705,  
"a where": 0.058823529411764705,  
"if </s>": 0.06666666666666667,  
"if <s>": 0.06666666666666667,  
"if a": 0.06666666666666667,  
"if if": 0.06666666666666667,  
"if of": 0.06666666666666667,  
"if peck": 0.06666666666666667,  
"if peppers": 0.06666666666666667,  
"if picked": 0.06666666666666667,  
"if pickled": 0.06666666666666667,  
"if piper": 0.06666666666666667,  
"if s": 0.06666666666666667,
```

"if the": 0.06666666666666667,  
"if where": 0.06666666666666667,  
"of </s>": 0.05555555555555555,  
"of <s>": 0.05555555555555555,  
"of a": 0.05555555555555555,  
"of if": 0.05555555555555555,  
"of of": 0.05555555555555555,  
"of peck": 0.05555555555555555,  
"of peppers": 0.05555555555555555,  
"of peter": 0.05555555555555555,  
"of picked": 0.05555555555555555,  
"of piper": 0.05555555555555555,  
"of s": 0.05555555555555555,  
"of the": 0.05555555555555555,  
"of where": 0.05555555555555555,  
"peck </s>": 0.05555555555555555,  
"peck <s>": 0.05555555555555555,  
"peck a": 0.05555555555555555,  
"peck if": 0.05555555555555555,  
"peck peck": 0.05555555555555555,  
"peck peppers": 0.05555555555555555,  
"peck peter": 0.05555555555555555,  
"peck picked": 0.05555555555555555,  
"peck pickled": 0.05555555555555555,  
"peck piper": 0.05555555555555555,  
"peck s": 0.05555555555555555,  
"peck the": 0.05555555555555555,  
"peck where": 0.05555555555555555,  
"peppers </s>": 0.05555555555555555,  
"peppers <s>": 0.05555555555555555,  
"peppers a": 0.05555555555555555,  
"peppers if": 0.05555555555555555,  
"peppers of": 0.05555555555555555,  
"peppers peck": 0.05555555555555555,  
"peppers peppers": 0.05555555555555555,  
"peppers picked": 0.05555555555555555,  
"peppers pickled": 0.05555555555555555,  
"peppers piper": 0.05555555555555555,  
"peppers s": 0.05555555555555555,  
"peppers the": 0.05555555555555555,  
"peppers where": 0.05555555555555555,  
"peter </s>": 0.05555555555555555,  
"peter <s>": 0.05555555555555555,  
"peter a": 0.05555555555555555,  
"peter if": 0.05555555555555555,  
"peter of": 0.05555555555555555,  
"peter peck": 0.05555555555555555,  
"peter peppers": 0.05555555555555555,  
"peter peter": 0.05555555555555555,  
"peter picked": 0.05555555555555555,  
"peter pickled": 0.05555555555555555,  
"peter s": 0.05555555555555555,  
"peter the": 0.05555555555555555,  
"peter where": 0.05555555555555555,

"picked </s>": 0.05555555555555555,  
"picked <s>": 0.05555555555555555,  
"picked if": 0.05555555555555555,  
"picked of": 0.05555555555555555,  
"picked peck": 0.05555555555555555,  
"picked peppers": 0.05555555555555555,  
"picked peter": 0.05555555555555555,  
"picked picked": 0.05555555555555555,  
"picked pickled": 0.05555555555555555,  
"picked piper": 0.05555555555555555,  
"picked s": 0.05555555555555555,  
"picked the": 0.05555555555555555,  
"picked where": 0.05555555555555555,  
"pickled </s>": 0.05555555555555555,  
"pickled <s>": 0.05555555555555555,  
"pickled a": 0.05555555555555555,  
"pickled if": 0.05555555555555555,  
"pickled of": 0.05555555555555555,  
"pickled peck": 0.05555555555555555,  
"pickled peter": 0.05555555555555555,  
"pickled picked": 0.05555555555555555,  
"pickled pickled": 0.05555555555555555,  
"pickled piper": 0.05555555555555555,  
"pickled s": 0.05555555555555555,  
"pickled the": 0.05555555555555555,  
"pickled where": 0.05555555555555555,  
"piper </s>": 0.05555555555555555,  
"piper <s>": 0.05555555555555555,  
"piper a": 0.05555555555555555,  
"piper if": 0.05555555555555555,  
"piper of": 0.05555555555555555,  
"piper peck": 0.05555555555555555,  
"piper peppers": 0.05555555555555555,  
"piper peter": 0.05555555555555555,  
"piper pickled": 0.05555555555555555,  
"piper piper": 0.05555555555555555,  
"piper s": 0.05555555555555555,  
"piper the": 0.05555555555555555,  
"piper where": 0.05555555555555555,  
"s </s>": 0.06666666666666667,  
"s <s>": 0.06666666666666667,  
"s a": 0.06666666666666667,  
"s if": 0.06666666666666667,  
"s of": 0.06666666666666667,  
"s peck": 0.06666666666666667,  
"s peppers": 0.06666666666666667,  
"s peter": 0.06666666666666667,  
"s picked": 0.06666666666666667,  
"s pickled": 0.06666666666666667,  
"s piper": 0.06666666666666667,  
"s s": 0.06666666666666667,  
"s where": 0.06666666666666667,  
"the </s>": 0.06666666666666667,  
"the <s>": 0.06666666666666667,

```
"the a": 0.06666666666666667,  
"the if": 0.06666666666666667,  
"the of": 0.06666666666666667,  
"the peppers": 0.06666666666666667,  
"the peter": 0.06666666666666667,  
"the picked": 0.06666666666666667,  
"the pickled": 0.06666666666666667,  
"the piper": 0.06666666666666667,  
"the s": 0.06666666666666667,  
"the the": 0.06666666666666667,  
"the where": 0.06666666666666667,  
"where </s>": 0.06666666666666667,  
"where <s>": 0.06666666666666667,  
"where a": 0.06666666666666667,  
"where if": 0.06666666666666667,  
"where of": 0.06666666666666667,  
"where peck": 0.06666666666666667,  
"where peppers": 0.06666666666666667,  
"where peter": 0.06666666666666667,  
"where picked": 0.06666666666666667,  
"where pickled": 0.06666666666666667,  
"where piper": 0.06666666666666667,  
"where the": 0.06666666666666667,  
"where where": 0.06666666666666667  
}
```

Q7.

	Bigram	Bigram Count
0	<s> where	1
1	peck of	4
2	<s> peter	1
3	s the	1
4	peppers peter	2
5	if peter	1
6	picked a	2
7	peppers </s>	2
8	piper picked	4
9	peter piper	4
10	picked </s>	2
11	<s> if	1
12	pickled peppers	4
13	the peck	1
14	where s	1
15	a peck	3
16	of pickled	4
17	<s> a	1

Q8. We perform smoothing on the count of the counts table in order to account for the unseen words

Q9.

a. MLE Probability = 0.0

b. Laplace Probability =  $4.202641275989133e-05$

Bonus: I have tried to do it for the brown corpus but some part is taking very long to complete it took ~ 2 hrs to process 8% of the total bigram. And my system got hanged. Then i tried the same thing on the google colab But there it was giving me the error that your program end up using all the RAM, therefore suspended , So i didn't proceed further.