

Results:

- Number of Sentences in Test File: 4
- Number of Total Tokens: 30
- Number of Unique Words (Vocabulary Size): 11
- Top 10 Unigrams with Highest Frequencies:

("came")	7	0.23333333333333334
("</s>")	4	0.13333333333333333
("i")	4	0.13333333333333333
("<s>")	4	0.13333333333333333
(",")	3	0.1
("they")	2	0.06666666666666667
("he")	2	0.06666666666666667
("come")	1	0.03333333333333333
("not")	1	0.03333333333333333
("did")	1	0.03333333333333333
- Top 10 Bigrams with Highest Frequencies:

(</s>,came)	4	0.5714285714285714
(came,i)	3	0.75
(came,they)	2	1.0
(i,<s>)	2	0.5
(i,,)	2	0.6666666666666666
(,,came)	2	0.2857142857142857
(came,he)	2	1.0
(they,,)	1	0.3333333333333333
(,,come)	1	1.0
(come,not)	1	1.0
- After UNK addition and Smoothing Operations:
 - Top 10 Bigrams with Highest Frequencies:

(</s>,came)	4	0.25
(came,i)	3	0.23333333333333334
(came,they)	2	0.19230769230769232
(i,<s>)	2	0.16666666666666666
(i,,)	2	0.17857142857142858
(,,came)	2	0.13888888888888889
(came,he)	2	0.19230769230769232
(they,,)	1	0.10714285714285714
(,,come)	1	0.125
(come,not)	1	0.125
 - Top 10 Bigrams with UNK:

(</s>,came)	0.28125	0.28125
(came,i)	0.2692307692307692	0.2692307692307692
(came,they)	0.22727272727272727	0.22727272727272727
(i,<s>)	0.19230769230769232	0.19230769230769232
(i,,)	0.20833333333333334	0.20833333333333334
(,,came)	0.15625	0.15625
(came,he)	0.22727272727272727	0.22727272727272727
(they,,)	0.125	0.125
(,,come)	0.15	0.15
(come,unk)	0.125	0.125
- iT could have, been a very NICE cat!: 4.826440695908175e-11

-I was here when they came.: 2.5628906249999996e-07

Sentences:

- he came, i came
- i came and he came
- they came, i came
- i did not come, they came

Frequency List:

("and": 1), ("did": 1), ("not": 1), ("come": 1), ("he": 2), ("they": 2), (" ": 3), ("("i": 4), ("

Bigram Frequency List:

(he,<s>) =1 (and,came) =1 (he,and) =1 (they,<s>) =1 (did,i) =1
(not,did) =1 (come,not) =1 (,,come) =1 (they,,) =1 (came,he) =2
(,,came) =2 (i,,) =2 (i,<s>) =2 (came,they) =2 (came,i) =3
(</s>,came) =4

Unigram Probability List:

("and")= 0.03333 ("did")= 0.03333 ("not")= 0.03333 ("come")= 0.03333
("they")= 0.06667 (" ")= 0.1 ("("came")= 0.23333

Bigram Probability List:

(he,<s>)= 0.25 (and,came)= 0.143 (he,and)= 1.0 (they,<s>)= 0.25
(not,did)= 1.0 (come,not)= 1.0 (,,come)= 1.0 (they,,)= 0.333
(,,came)= 0.286 (i,,)= 0.667 (i,<s>)= 0.5 (came,they)= 1.0
(</s>,came)= 0.571

Sentences After Replace With UNK:

- <s> he came , i came </s>
- <s> i came unk he came </s>
- <s> they came , i came </s>
- <s> i unk unk come , they came </s>

Frequency List After Replace with UNK:

("come": 1), ("he": 2), ("they": 2), (" ": 3), ("unk": 3), ("("came": 7),

Bigram Frequency List After Replace with UNK:

(he,<s>) =1 (unk,came) =1 (he,unk) =1 (they,<s>) =1 (unk,i) =1
(unk,unk) =1 (come,unk) =1 (,,come) =1 (they,,) =1 (came,he) =2
(,,came) =2 (i,,) =2 (i,<s>) =2 (came,they) =2 (came,i) =3
(</s>,came) =4

Bigram Probability List After Replace with UNK and Smoothing:

(he,<s>)= 0.115

(unk,came)= 0.094

(he,unk)= 0.125

(they,<s>)= 0.115

(unk,unk)= 0.125

(come,unk)= 0.125

(,,come)= 0.15

(they,,)= 0.125

(,,came)= 0.156

(i,,)= 0.208

(i,<s>)= 0.192

(came,they)= 0.227

(</s>,came)= 0.281