P1

(b): For DostoevskyKaramazov.txt, we need to know 60 words to read 50% of text.

For DrSeuss.txt, we need to know 36 words to read 50% of text.

P2

(b)

Percentage of unique count:

n, percentage

1 33.1106

2 68.6534

3 87.5173

4 96.7292

5 99.2807

6 99.8370

The value of n that gives no common ngrams: 18.

The largest common nGrams: repulsion that s what i m afraid of that s what may be too much for me

(c)

The value of n that gives no common ngrams: 8.

The largest common nGrams: in the middle of the table and there is no such thing as a

(d)

The value of n that gives no common ngrams: 7.

The largest common nGrams: of nature and of reason the

(e)

From the same write, is the value of n that gives no common ngrams is much larger. People tend to repeat their own phrases.

P3

(b)

The result with n=1…6

again the first objective question as if he said <END>

dr <END>

through the corridor which led between the rooms he saw leni to whom the warning cry of the door handle and said <END>

come into my room for a few minutes then <END>

you can see that dr <END>

the text for an assertion of this sort was passed down to me from my father and it s quite unassailable <END>

With n increasing, the sentence looks more close to the original text. With n=1, it just appears to be some random words. With n=6, it has almost the same sentence in the text.

(c)

The result is: Alienation of humanity.

Marx and Engels talks more about politics. While Kafka is a great novel writer.

P4

(b)

-183.035

-200.678

-111.105

-187.372

P5

(b)

-137.611

-198.278

-nan(ind)

P6

(b)

P6 1 0.05 50

5.9761

121 0 0 3 0 5

0 219 11 3 4 0

0 8 245 0 7 0

1 4 2 347 4 8

0 6 6 0 113 0

0 0 0 3 0 135

P6 2 0.05 50

1.03586

125 0 0 0 0 4

0 236 1 0 0 0

0 2 257 0 1 0

0 0 1 365 0 0

0 0 2 0 123 0

1 0 0 1 0 136

P6 3 0.05 50

43.8247

5 8 0 116 0 0

0 217 0 20 0 0

0 15 58 187 0 0

0 0 0 366 0 0

0 2 0 123 0 0

0 4 0 75 0 59

(c)

P6 1 0 50

7.33068

114 0 0 2 0 13

5 214 11 3 4 0

0 9 244 0 7 0

1 4 2 347 4 8

4 6 6 0 109 0

0 0 0 3 0 135

P6 2 0 50

15.5378

124 0 0 2 0 3

24 213 0 0 0 0

29 3 221 6 1 0

30 0 0 335 1 0

38 0 0 2 85 0

53 0 0 3 0 82

P6 3 0 50

50.8367

128 0 0 1 0 0

105 132 0 0 0 0

144 1 113 2 0 0

164 0 0 202 0 0

104 0 0 0 21 0

117 0 0 0 0 21

(d)

P6 3 0.05 50

43.8247

5 8 0 116 0 0

0 217 0 20 0 0

0 15 58 187 0 0

0 0 0 366 0 0

0 2 0 123 0 0

0 4 0 75 0 59

P6 3 0.005 50

10.8367

100 4 0 22 0 3

0 237 0 0 0 0

0 10 241 9 0 0

2 1 0 363 0 0

0 13 3 53 56 0

1 2 1 11 1 122

P6 3 0.0005 50

3.34661

118 1 0 5 1 4

0 237 0 0 0 0

0 2 257 1 0 0

2 0 0 361 3 0

0 4 2 6 113 0

5 0 1 4 1 127

(e)

There are a few large trends here: as the length of N-Grams increases, so does the error rate, and as delta increases, so does the error rate.

(f)

P6 2 0.05 50

1.03586

125 0 0 0 0 4

0 236 1 0 0 0

0 2 257 0 1 0

0 0 1 365 0 0

0 0 2 0 123 0

1 0 0 1 0 136

P6 2 0.05 50

1.03586

125 0 0 0 0 4

0 236 1 0 0 0

0 2 257 0 1 0

0 0 1 365 0 0

0 0 2 0 123 0

1 0 0 1 0 136

P6 2 0.05 50

1.03586

125 0 0 0 0 4

0 236 1 0 0 0

0 2 257 0 1 0

0 0 1 365 0 0

0 0 2 0 123 0

1 0 0 1 0 136

P7

(b)

2 3 1 1

size: 124

size: 126

size: 83

size: 107

size: 87

i would love to her thae story

you will ged in the garden

hello from the top of the world

i will dink mlk in the morning

i will read thew story

2 3 0.1 1

size: 124

size: 126

size: 83

size: 107

size: 87

i would love to hear the story

you will read in the garden

hello from the top of the world

i will dink mlk in the morning

i will read thew story

2 3 0.01 1

size: 124

size: 126

size: 83

size: 107

size: 87

i would love to hear the story

you will read in the garden

hello from the top of the world

i will dink mlk in the morning

i will read the story

(c)

1 3 0.01 1

size: 124

size: 126

size: 83

size: 107

size: 87

i would love to he the story

you will re in the garden

hello from the to of the world

i will drink milk in the morning

i will read the story

2 3 0.01 1

size: 124

size: 126

size: 83

size: 107

size: 87

i would love to hear the story

you will read in the garden

hello from the top of the world

i will dink mlk in the morning

i will read the story

3 3 0.01 1

size: 124

size: 126

size: 83

size: 107

size: 87

i would clove to her the story

you will red ain the garden

hello froe the tp of the world

i bill drink mlk in the morning

i bill read they story

(d)

1 3 1 0

size: 124

size: 126

size: 83

size: 107

size: 87

i would love to he the story

you will re in the garden

hello from the to of the world

i will drink milk in the morning

i will read the story

2 3 1 0

size: 124

size: 126

size: 83

size: 107

size: 87

i would love to her thae story

you will ged in the garden

hello from the top of the world

i will dink mlk in the morning

i will read thew story

3 3 1 0

size: 124

size: 126

size: 83

size: 107

size: 87

i would love to fer the story

you will red ain the garden

hello froe the tp of the world

i will drink mlk ain the morning

i kill read they story