

new sensible word count

8000

6000

4000

2000

0

0.0

0.2

0.4

0.6

0.8

1.0

1.2

1.4

1.6

number of tokens in the reference data

1e9

