I have made a very simple tokenizer which checks the following :

For paragraphs:
I am checking for double ‘enter’.
Assumption: there is a blank line between two paragraphs.

For tokens:
I am checking that if the word is not a ‘-‘ and is space separated.
Assumption: There is no special character with spaces on both sides.

For lines:
I am checking for all the tokens if they have their last character as full stops and the first character of the next word is capital. Abbreviations have been also taken care of.
Assumption: For abbreviations, the first letter of the word is capital.

Output for Development Set

Number of paragraphs:	8
Number of tokens:	647
Number of lines:	25

Output for Test set

Number of paragraphs:	62
Number of tokens:	7677
Number of lines:	274