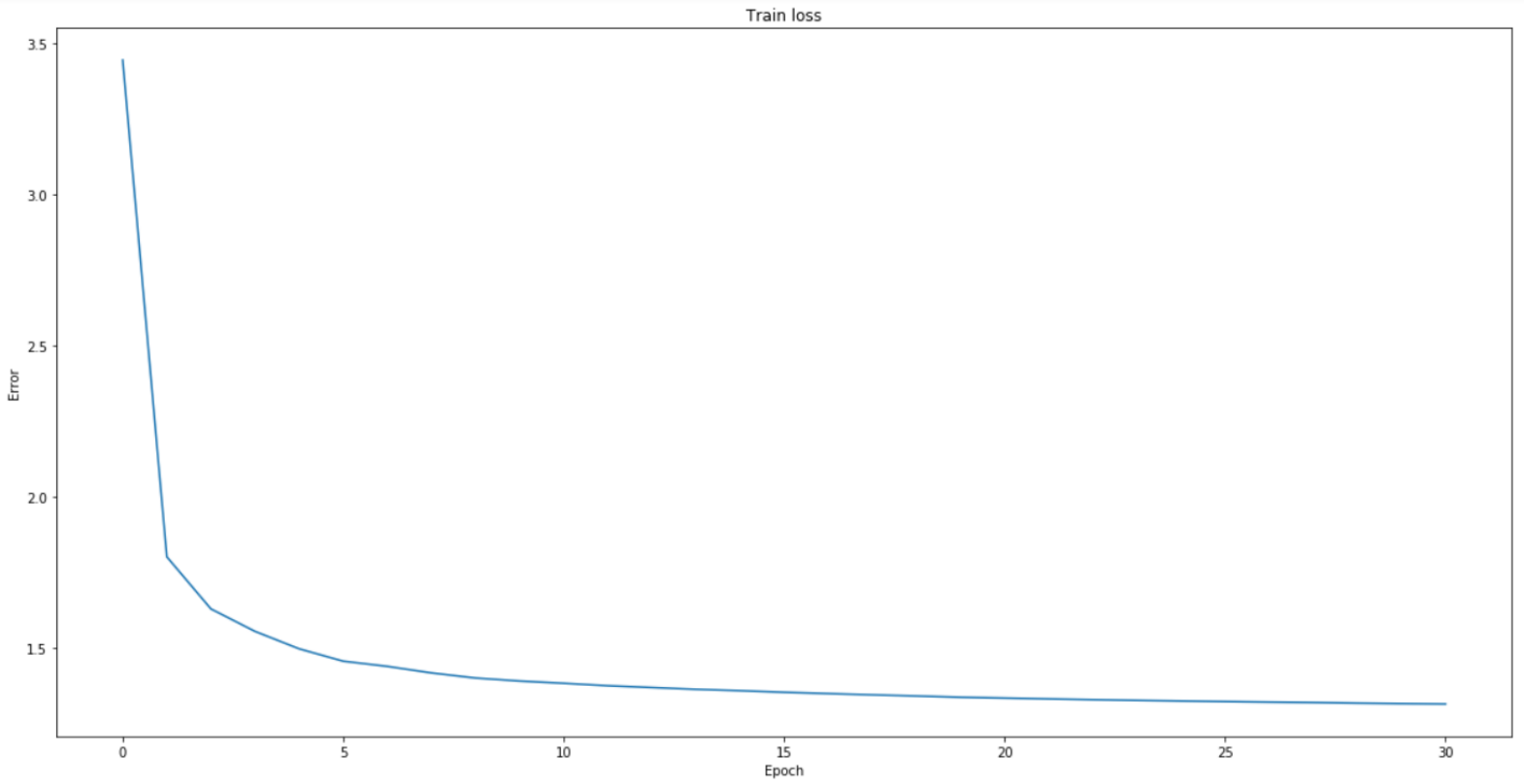
HW3 SHORT ANSWERS

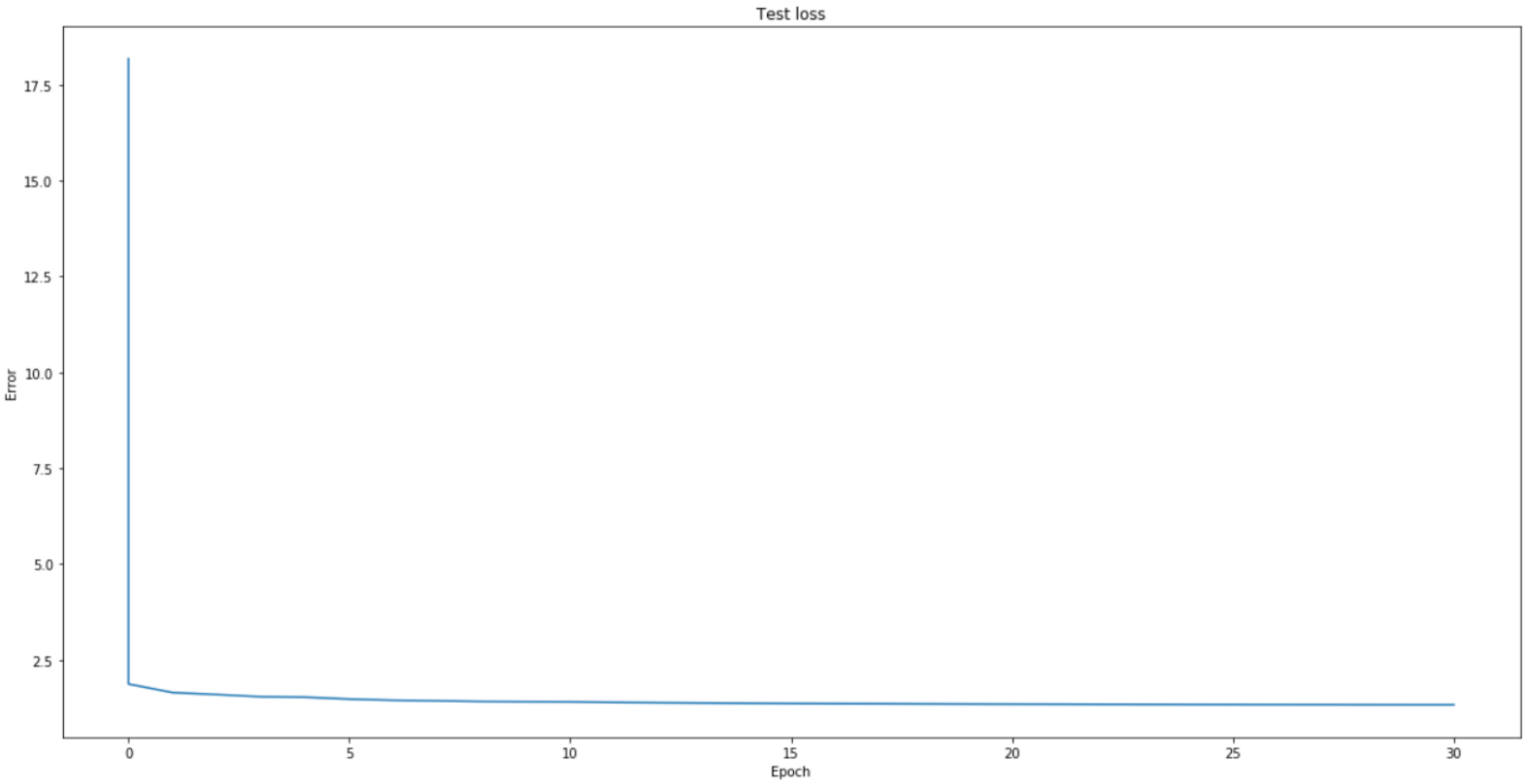
Short Answers for the ordinary RNN trained on harry_potter.txt:

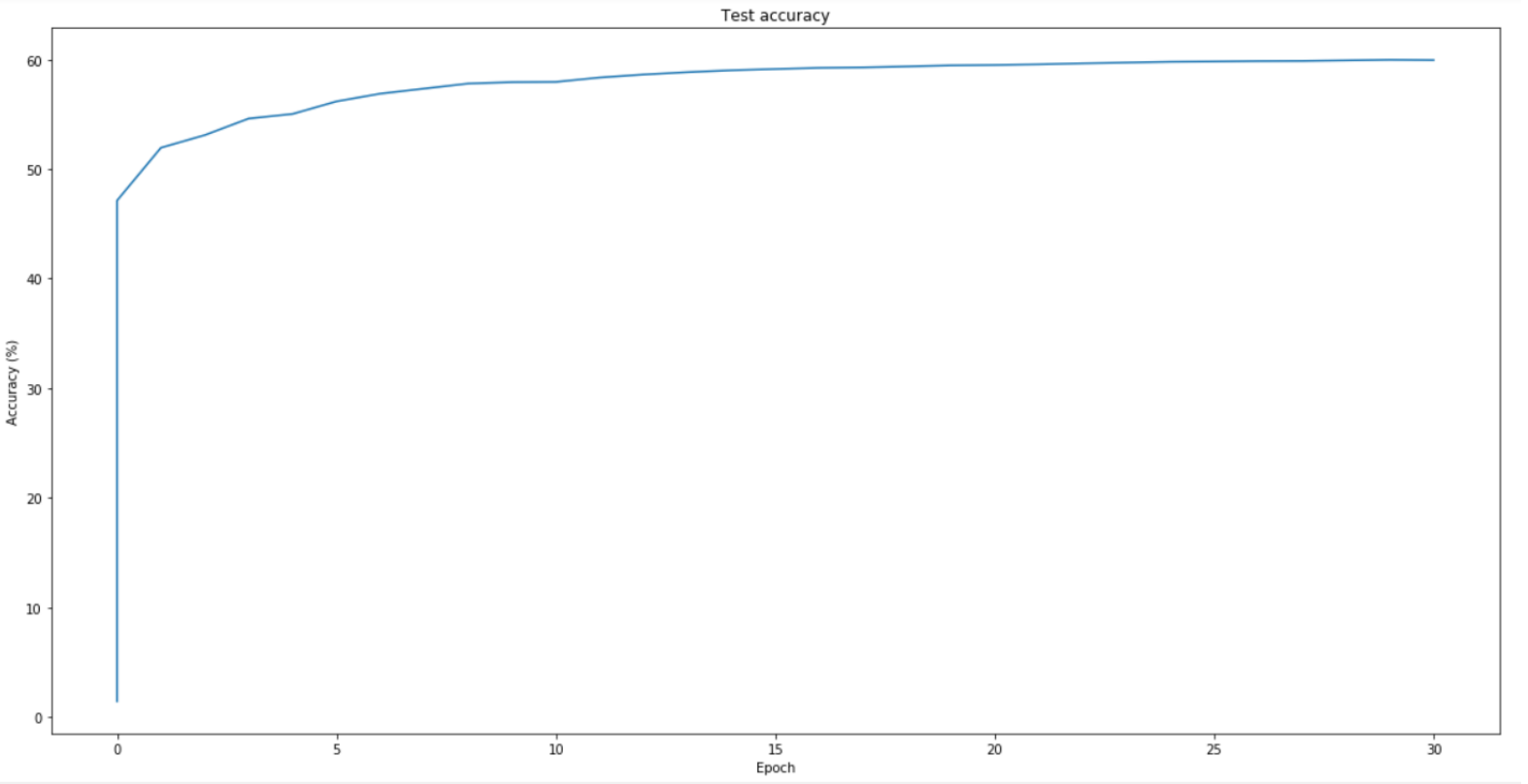
1) The modifications I made to the network are:
    a. I changed the number of epochs to 30.
    b. I increased the number of features to 1024.

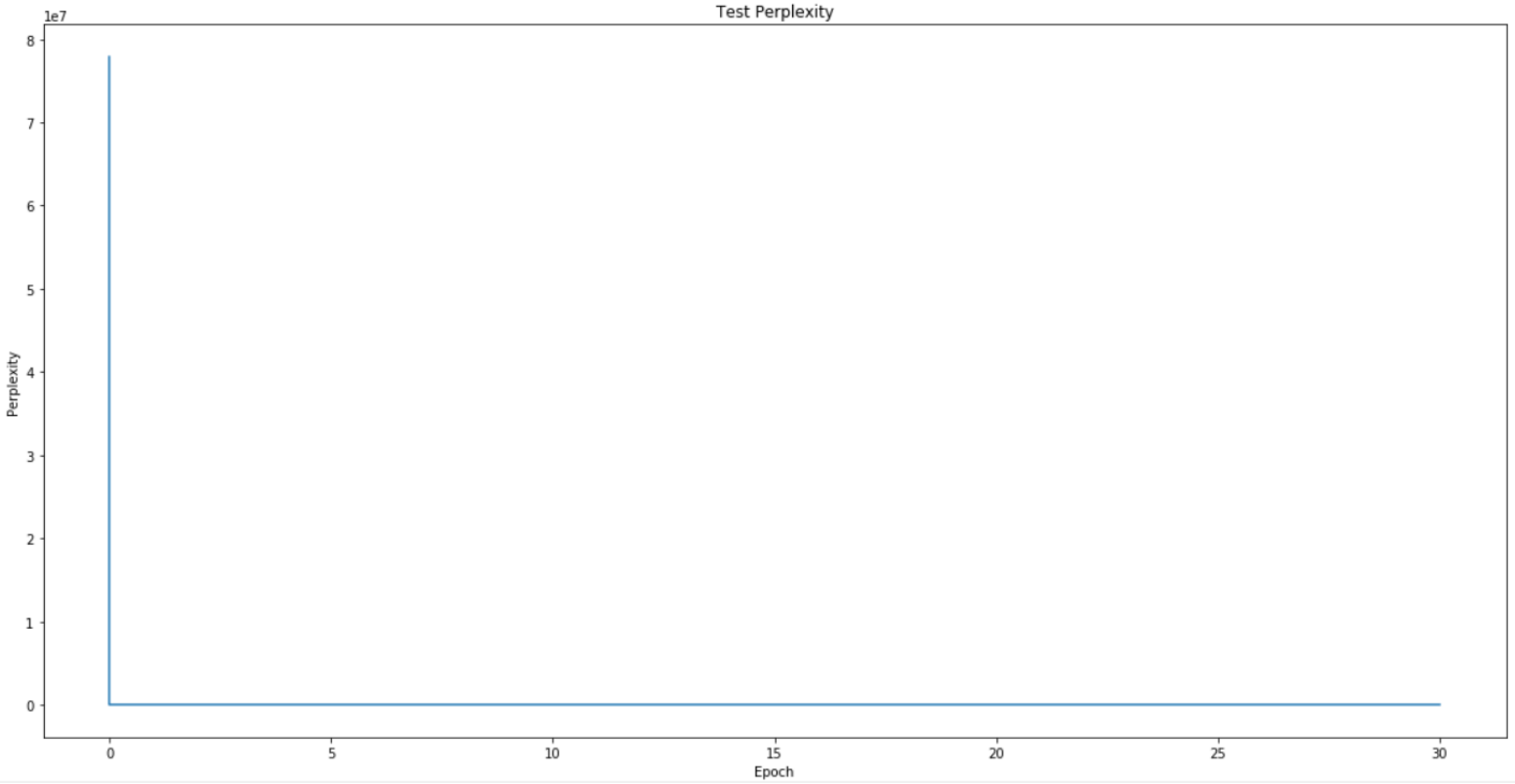
    After I increased the number of features my accuracy dropped by 2%, hence I got the best performance when I had the default hyper – parameter values with the exception of changing number of epochs to 30.

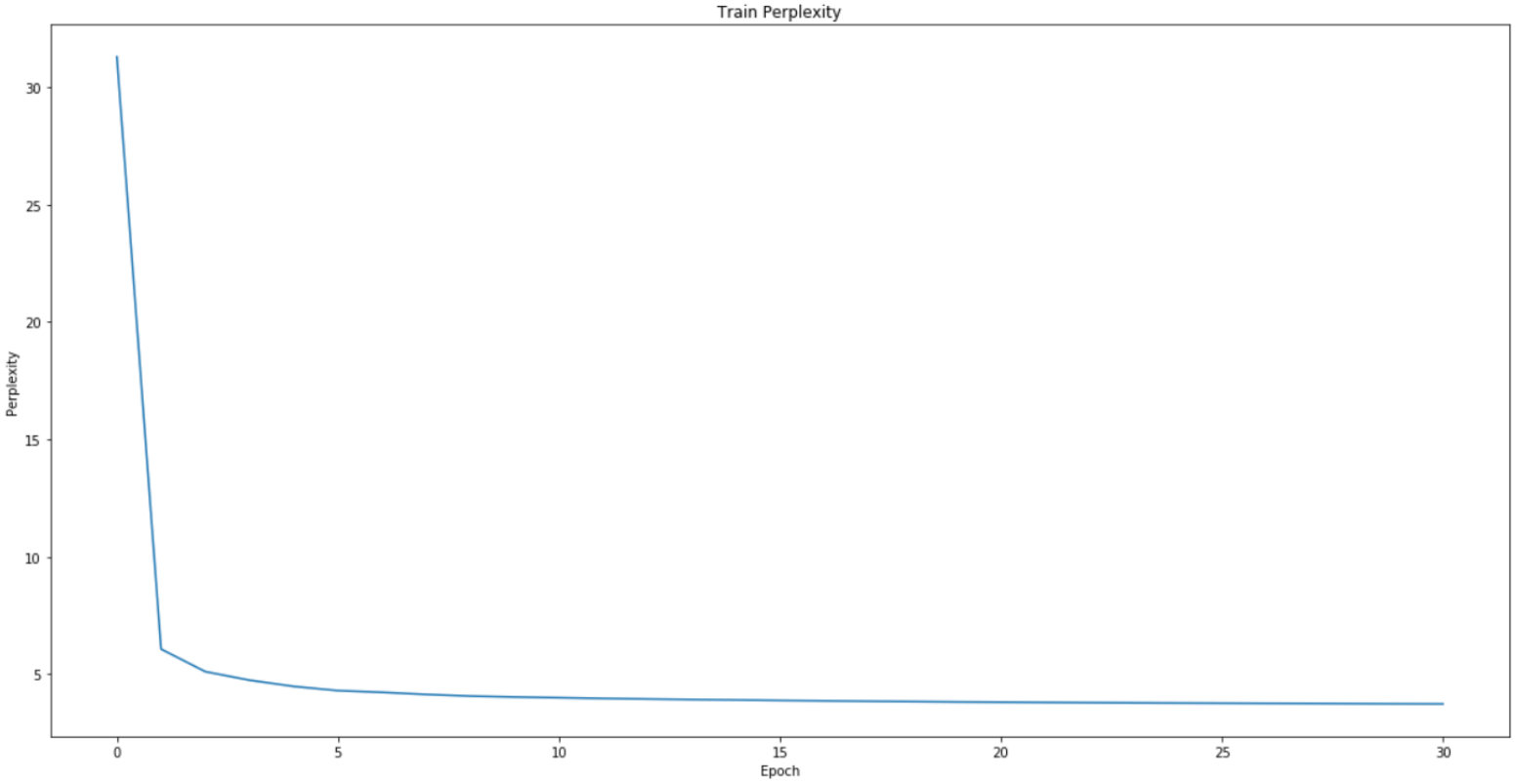
    The plots are below:




Test loss
Error
Epoch
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0
5
10
15
20
25
30


Test accuracy
Accuracy (%)
Epoch
60
50
40
30
20
10
0
0
5
10
15
20
25
30


Test Perplexity
1e7
8
7
6
5
4
3
2
1
0
Perplexity
0
5
10
15
20
25
30
Epoch


Train Perplexity
30
25
20
15
10
5
Perplexity
0
5
10
15
20
25
30
Epoch

2) Final accuracy was 60% and final test perplexity was 3.725
3) Favorite sentence generated with each of the sampling methods (the prompt I gave was harry grabbed the wand and):
    a. `generated with max: Harry grabbed the wand and said to hear the stairs and said in the castle of the castle of the castle`
    b. `generated with sample: Harry grabbed the wand and pulled out of the corridor to his feet and shaked to his back and the staring and because so what he had terribled the Dark and George was all so concentured in the garden the holding green move`
    c. `generated with beam: Harry grabbed the wand and Harry and Hermione was standing in the back of the darkness. Harry looked around at the entrance of the corner of the back of the door`

4) Beam sampling seems to generate the best results because beam sampling tries to maximize the probability of a sequence whereas max sampling and sample sampling only look at the probability of the next character. By trying to maximize the joint probability, beam sampling gives best result.
5) The best result I got was with temperature = 0.5, by best result I mean the sentence that was the funniest: `Harry looked at the pain in his eyes. "He's not under the tries and of the summer of the Minister of the Dark Transpinal was back to the office in the moment that he was under the corridors of the moment.` Temperature controls the randomness of prediction by scaling the logits before softmax is applied. The higher the temperature, the smoother the probability distribution, meaning the sampling from less likely candidates will increase. When the temperature is lowered, the model becomes more confident and sampling from candidates with higher probabilities increases. Having negative temperature will cause a serious problem because log(logits/temperature) will result in log of a negative number which will give you a NaN or a runtime error.

OTHER SECTION:

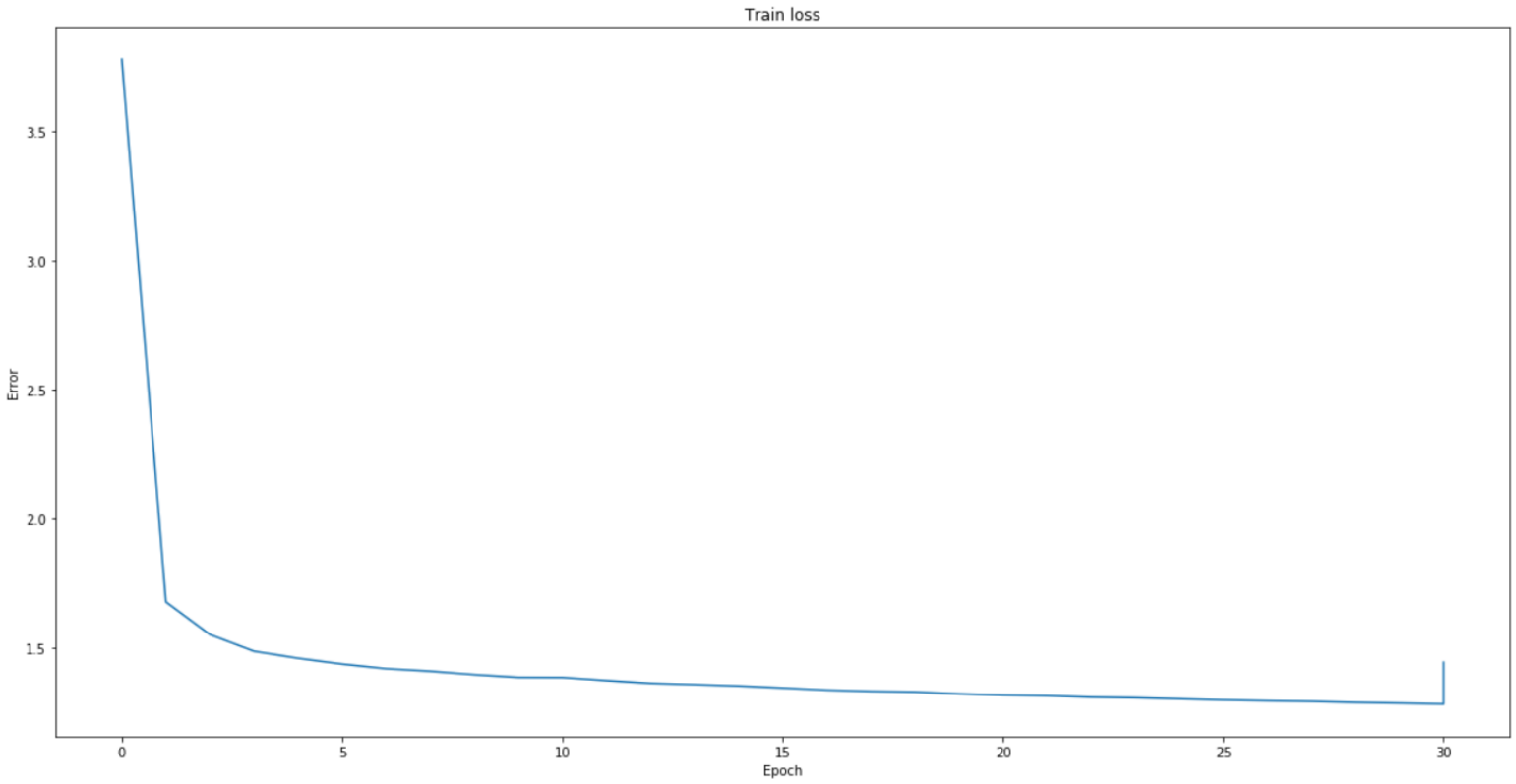
Trying on a different text corpus:

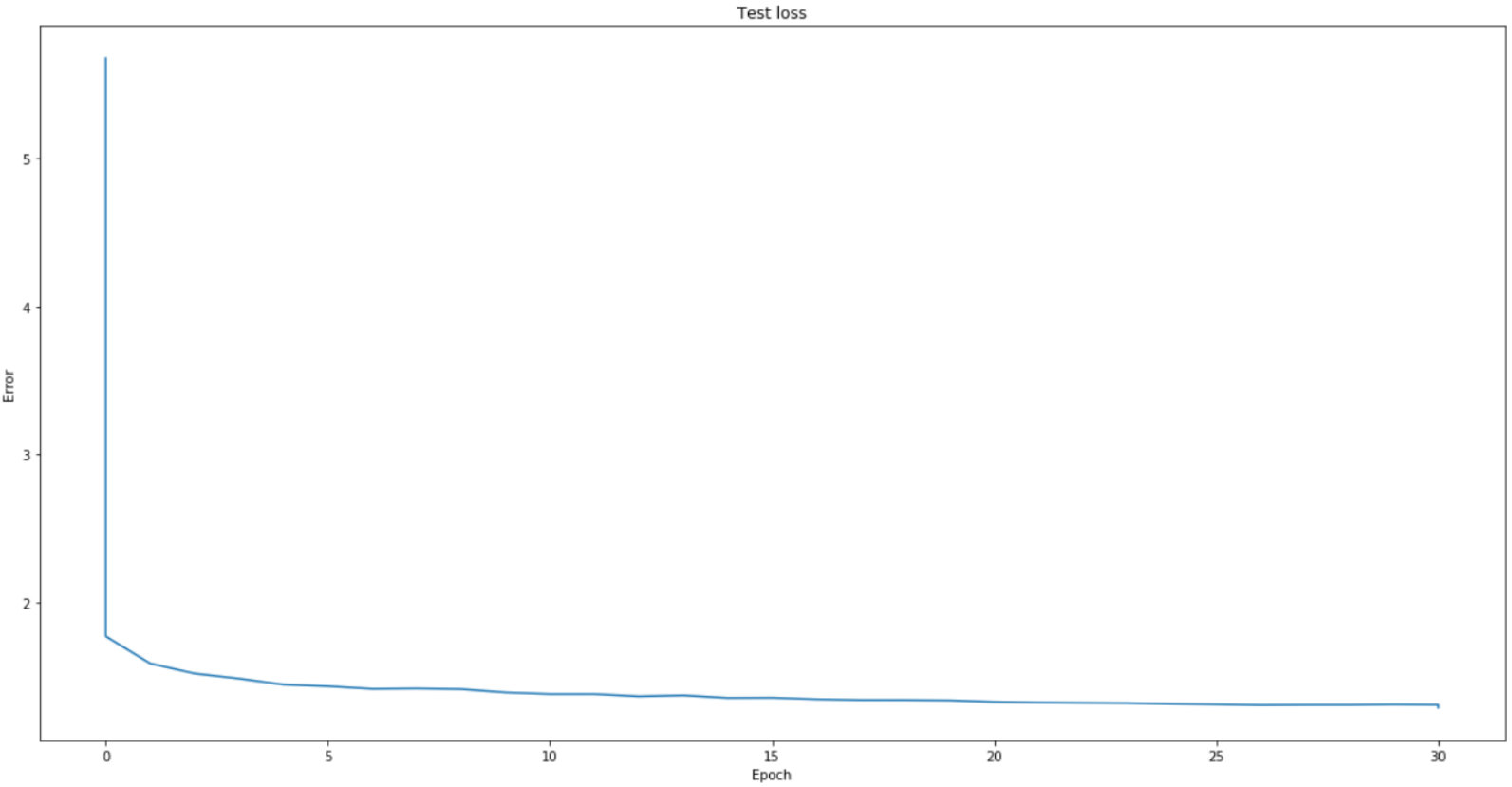
1) The corpus I used was a Lord Of the Rings book, it has 1,694,918 characters in it.
2) The accuracy with the new corpus was 55% hence the sentences were slightly less accurate than the harry potter case.
3) With beam width 10 and temperature = 0.5 I got the following sequences:
    a. `generated with max: Frodo put the ring on and strong the last for a shadow of the last for a shadow of the last for a shadow of the last for a shadow of the last for a shadow of the last for a shadow of the last for a shadow of the last`
    b. `generated with sample: Frodo put the ring on and let the skiling soft for it.`
    c. `generated with beam: Frodo put the ring on and singing in the Shire, and the light of the light of the Shire, and the light`

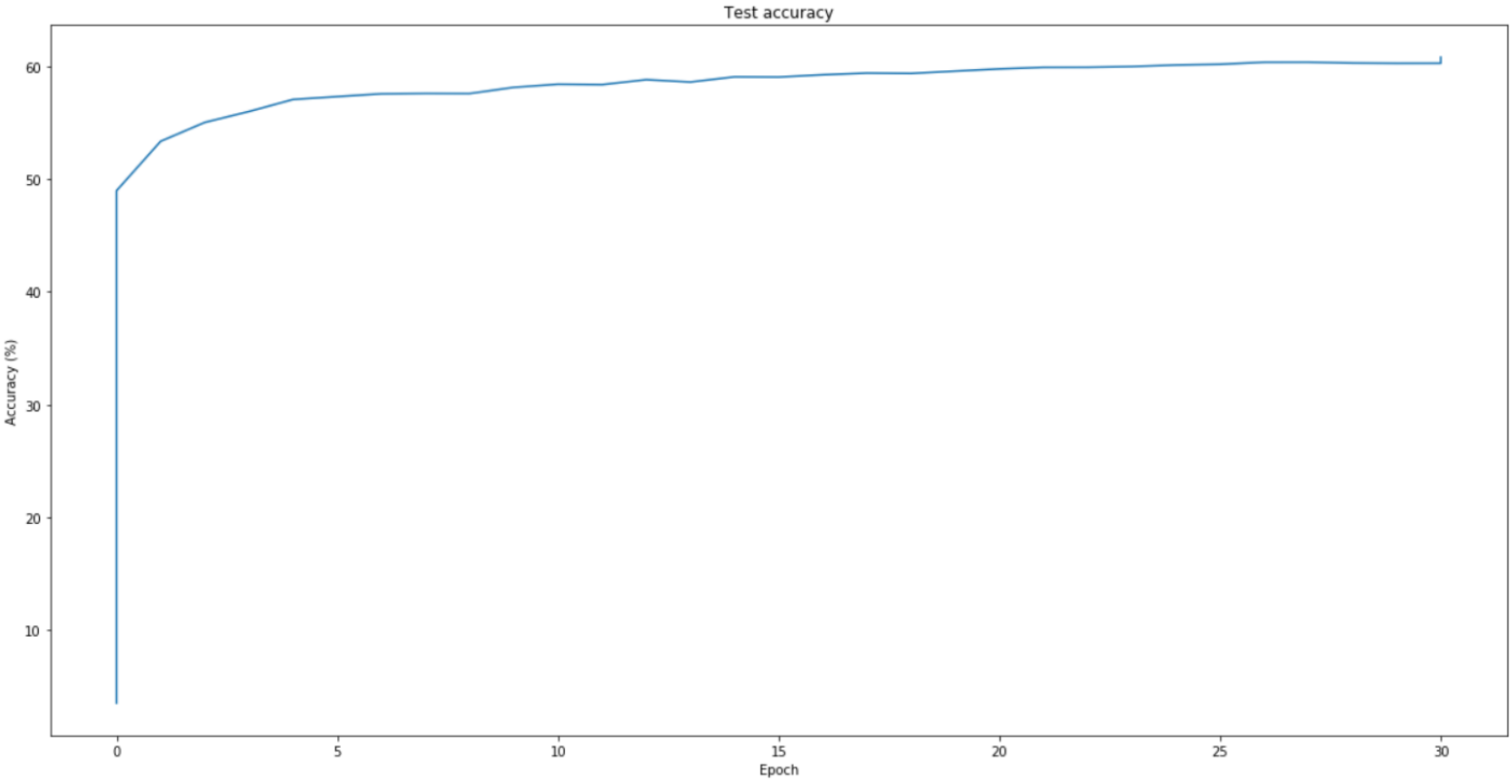
```
of the Shire, and the light of the light of the Shire, and the
light of the Shire, and the light of the Shire, and he said
```

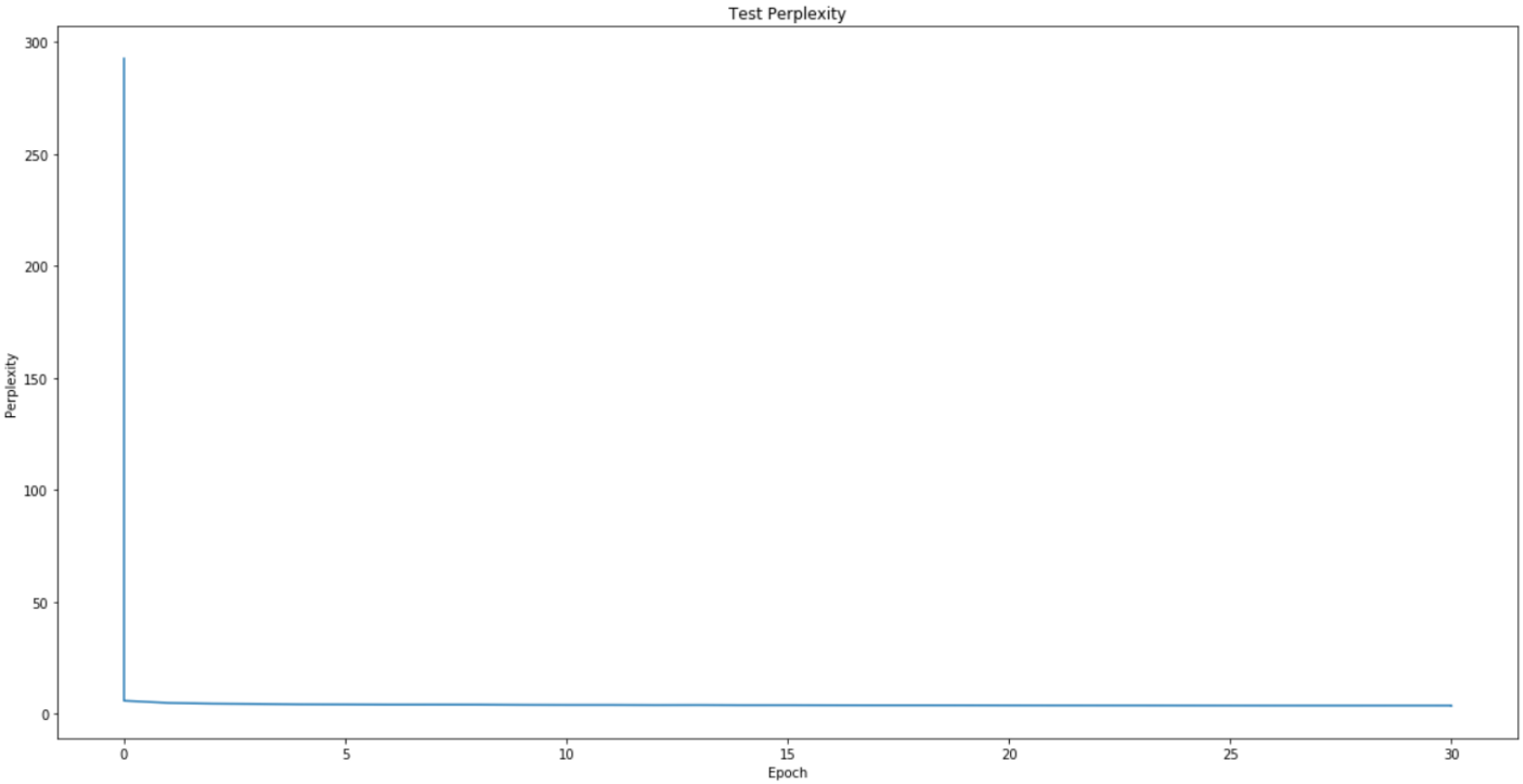
Changing the network architecture:

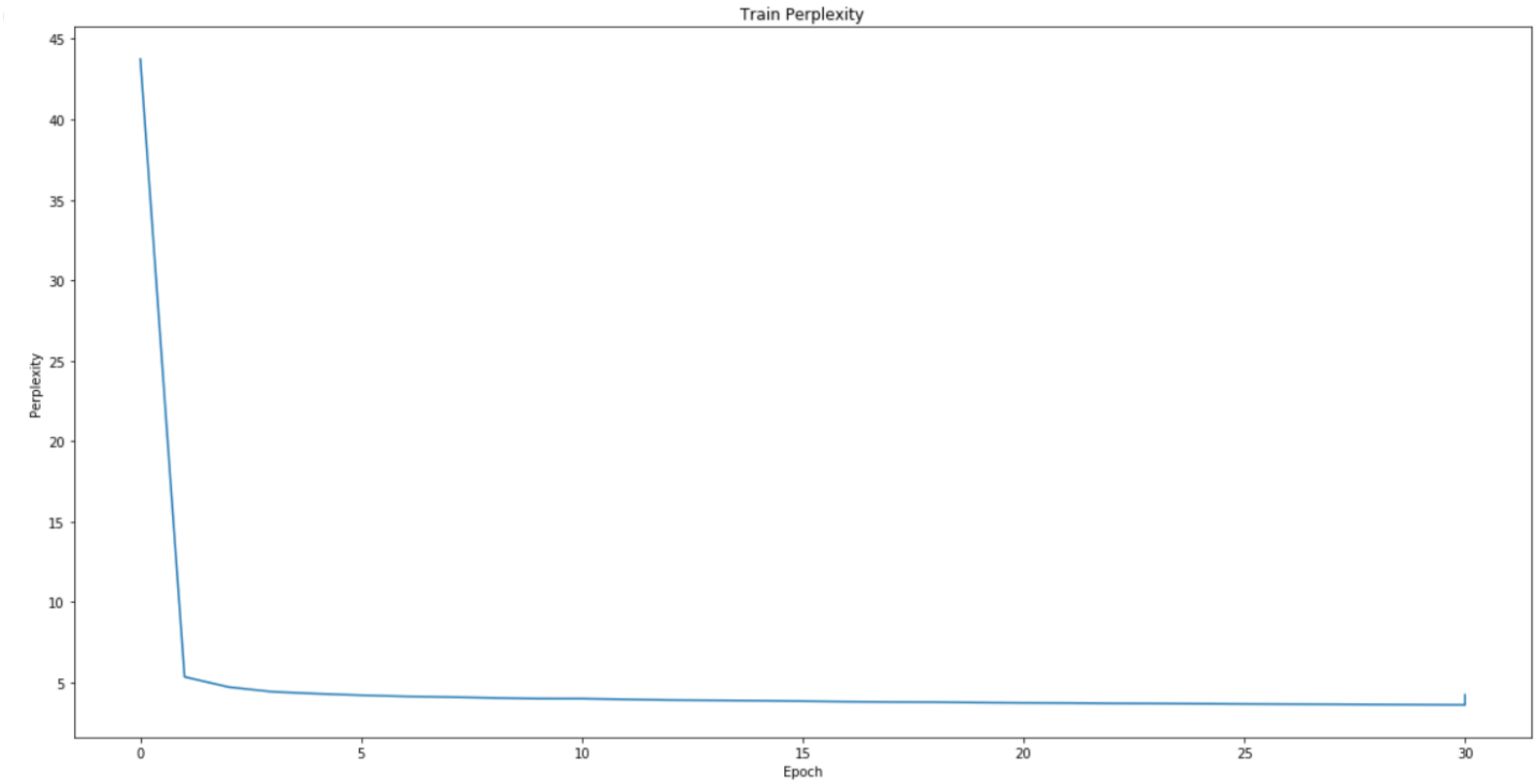
1) I increase the num_layers parameter for the GRU from 1 to 3 and the ran the training for 30 epochs. After making the changes, the highest accuracy I got was 61% whereas before I got 60%. The improvement is slightly better, but the difference is not significant. The plots are presented below:




Test loss
Error
Epoch
5
4
3
2
0
5
10
15
20
25
30


Test accuracy
Accuracy (%)
Epoch
60
50
40
30
20
10
0
5
10
15
20
25
30



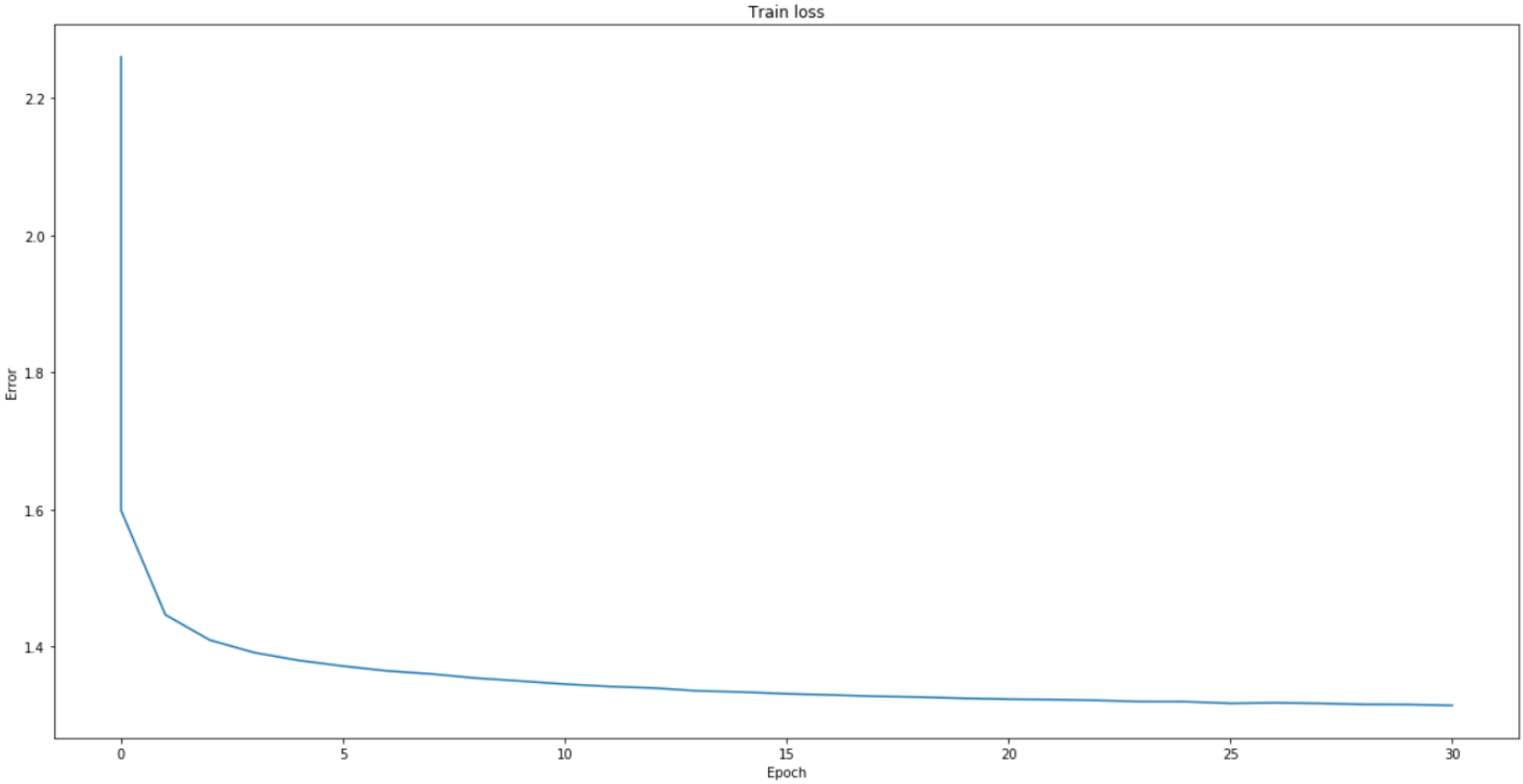


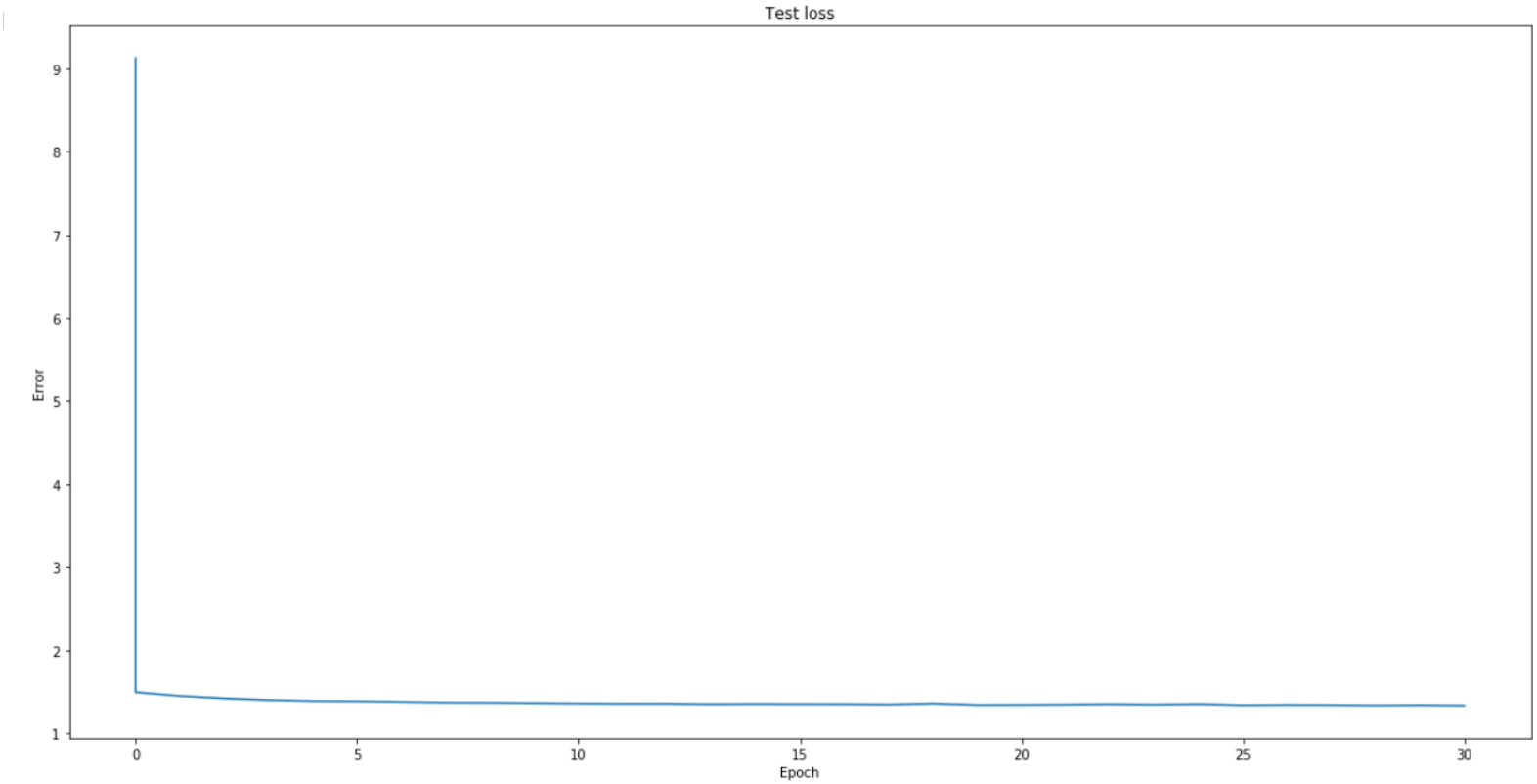
2) The final test accuracy was 61% and the lowest test perplexity was 4.22.

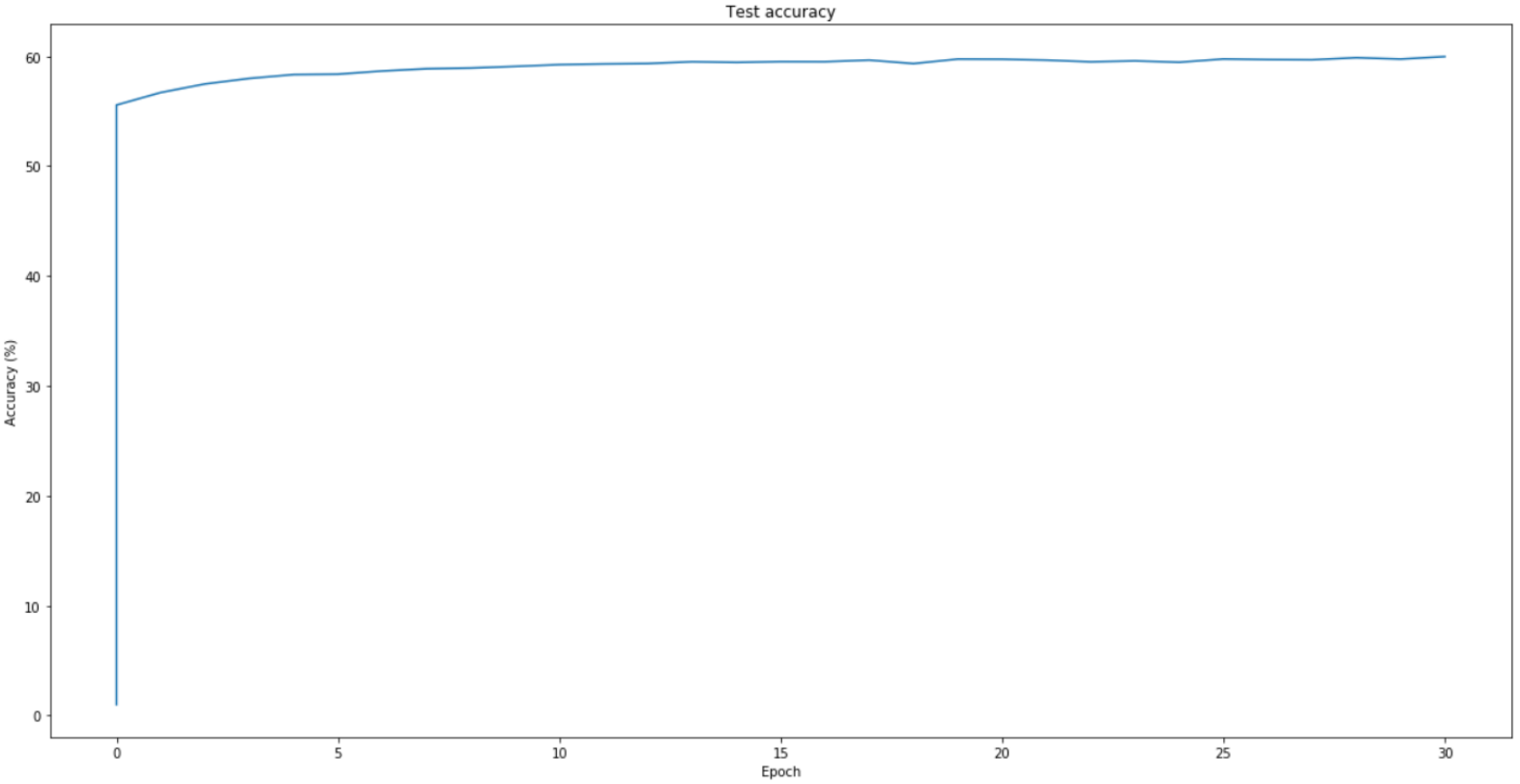
3) The seed words I gave to the sampling methods were (Temperature = 0.5): ‘Harry looked at the’
    a. Max sampling often resulted in repetition for example: `Harry looked at the stairs and started to the stairs and started to the stairs and started to the stairs and started to the stairs and started to the stairs and started to the stairs and started to the stairs and started`
    b. Sample sampling had interesting results, my favorite one is: `Harry looked at the real eye to the left of his head. "I don't like that?" said Harry and Harry saw it was a look of the windows.`
    c. Beam sampling also had interesting results: `Harry looked at the corridors and looked around at him. Harry was looking at Harry and Hermione and Hermione and Hermione looked around at the corridor of the window.`

LSTM:

1) The biggest difficulty I had was to modify the code so that I could accommodate the cell state outputted by the LSTM.
2) The results were very similar for the LSTM and the GRU. Plots:




Test loss
Error
Epoch
1
2
3
4
5
6
7
8
9
0
5
10
15
20
25
30


Test accuracy
Accuracy (%)
Epoch
0
10
20
30
40
50
60
0
5
10
15
20
25
30

3) Outputs of the text generators:
    a. generated with max: Harry grabbed the wand and said, "I said the first time that was a stand of the first time that he was standing and seemed to be seen the first time that he was starting to see the castle and started to see the castle and start
    b. generated with sample: Harry grabbed the wand and staring at the back of a smile seat of school and the only of them was all the only remotion for the windows.
    c. generated with beam: Harry grabbed the wand and Harry seemed to be able to see what he would have been going to get out of the front of the first time that they were not supposed to have been going to get out of the first time,"