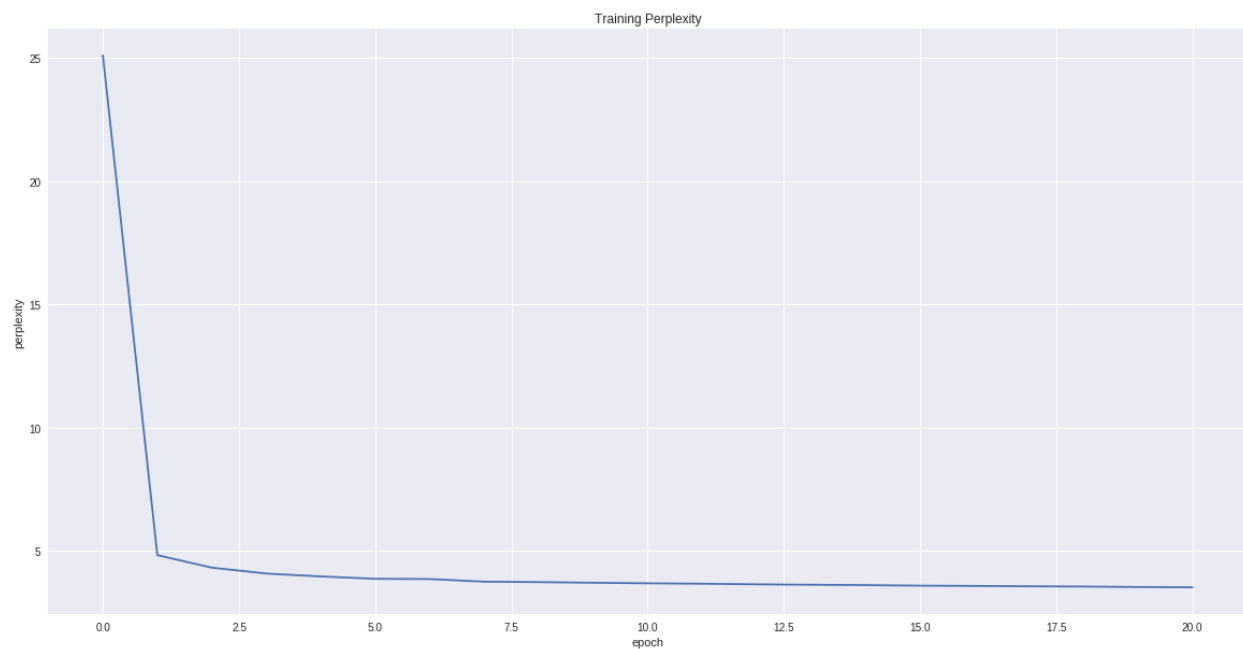
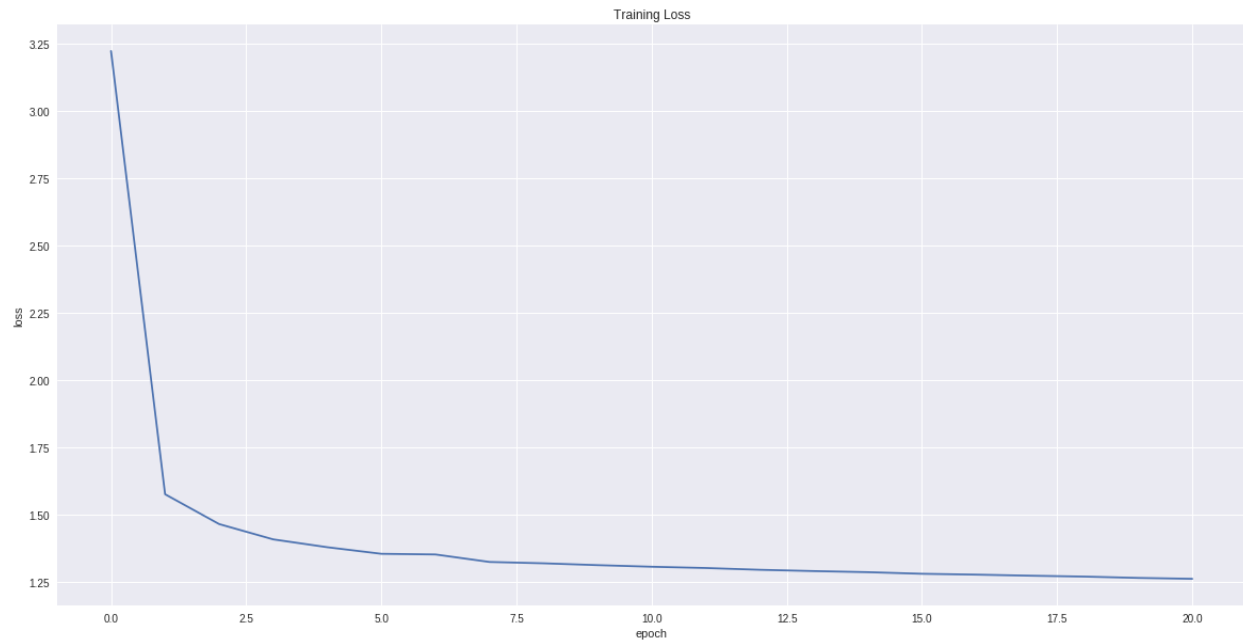
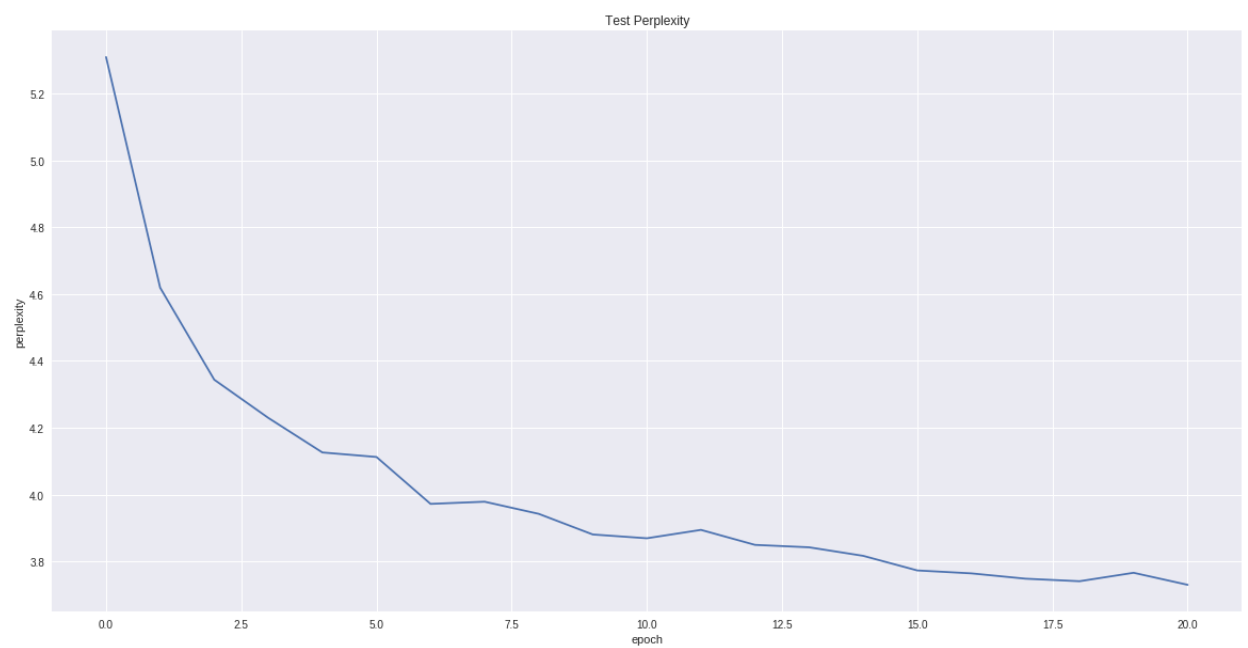
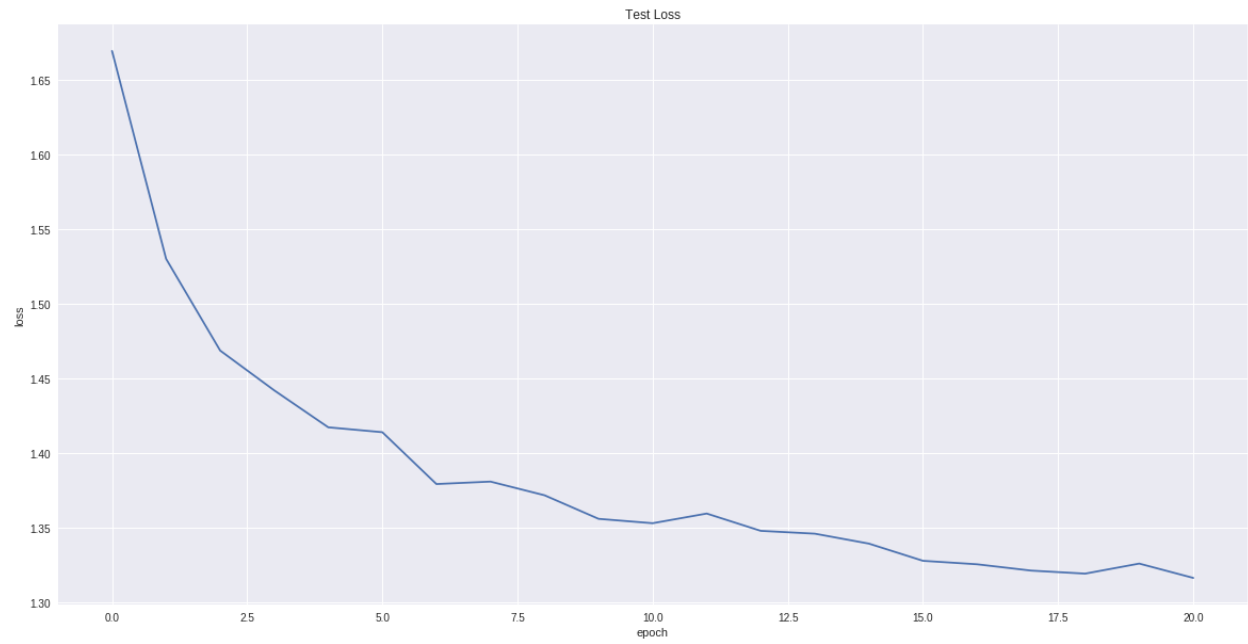


# Short Answer Questions – Word Model

## Train / Test Performance Figures







- generated with sample (temperature = 0.5) → Harry Potter and the room was a self sign of watch and the fire of the hall was wandering the expression of him and staring at the back of the boy had been purpled out of the portrait of the corridors and said, "Harry wobeam
- generated with beam (temperature = 0.5, beam width = 5) → Harry Potter and the wad been *slleady* to be *shet te* wad been *sotting tp ao* the *snpression* of the sir "he e was a lost *aeck oo* the *sarerti* of the sir "t was *aooking ap aith hhe soom wnd* she sir *af* the r sand "he e was

It seems that sample produces the most reasonable results. There are many non-sense words (italicized) in beam output.

For temperature in range (0, 1) we exponentially scale the values before normalizing over them which means the values we normalize over are either much smaller than their original (if the exponent was negative) or are much bigger than their original value (if the exponent was positive). That is, the probability distribution has more pronounced peaks and valleys. This makes sampling of the max more likely (under the asymptotic case of temperature → 0, sampling is equivalent to max). The opposite is true if temperature > 1, as we are smoothing out the probability distribution. This would encourage exploration for beam sampling that would allow us possibly produce better results. It will likely confuse sampling as the distribution we are working with now is smoother. If temperature = 1, we are working with the original probabilities.

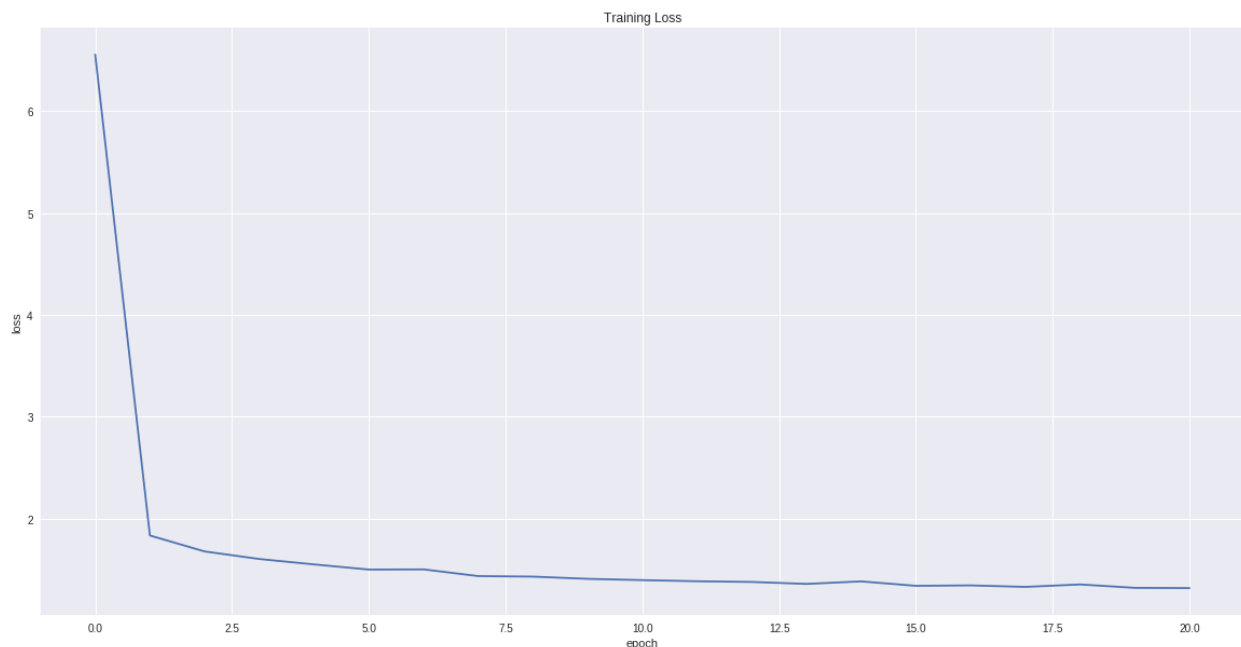
## Short Answer Questions – New Corpus

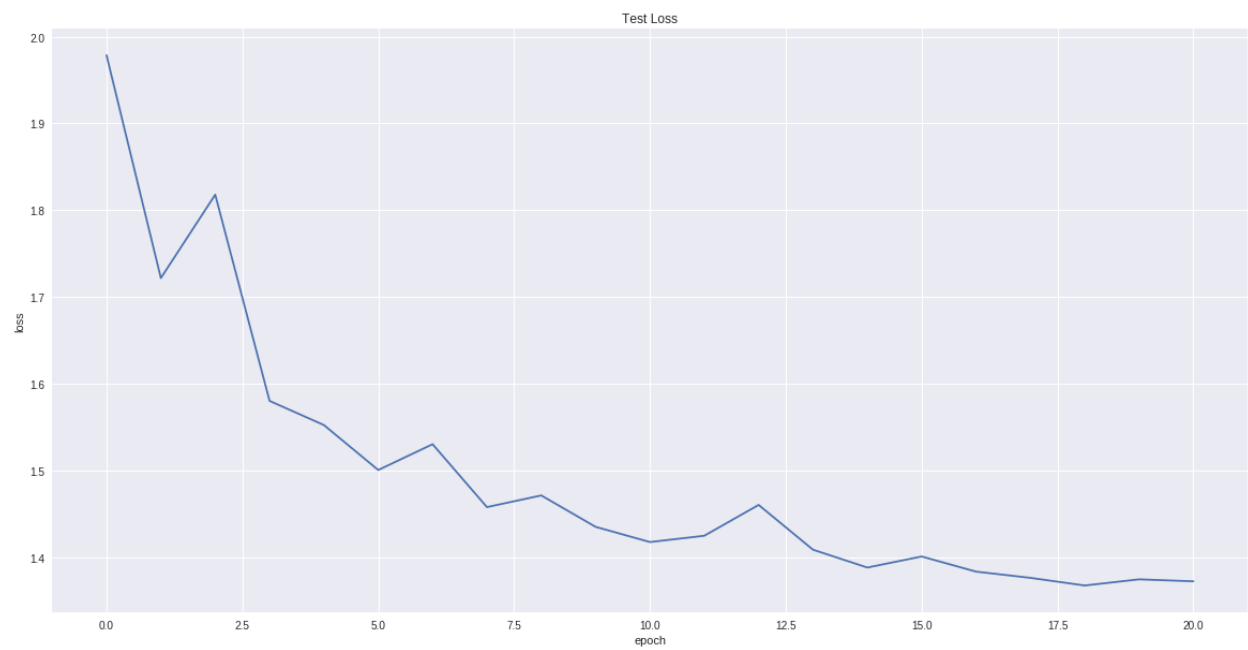
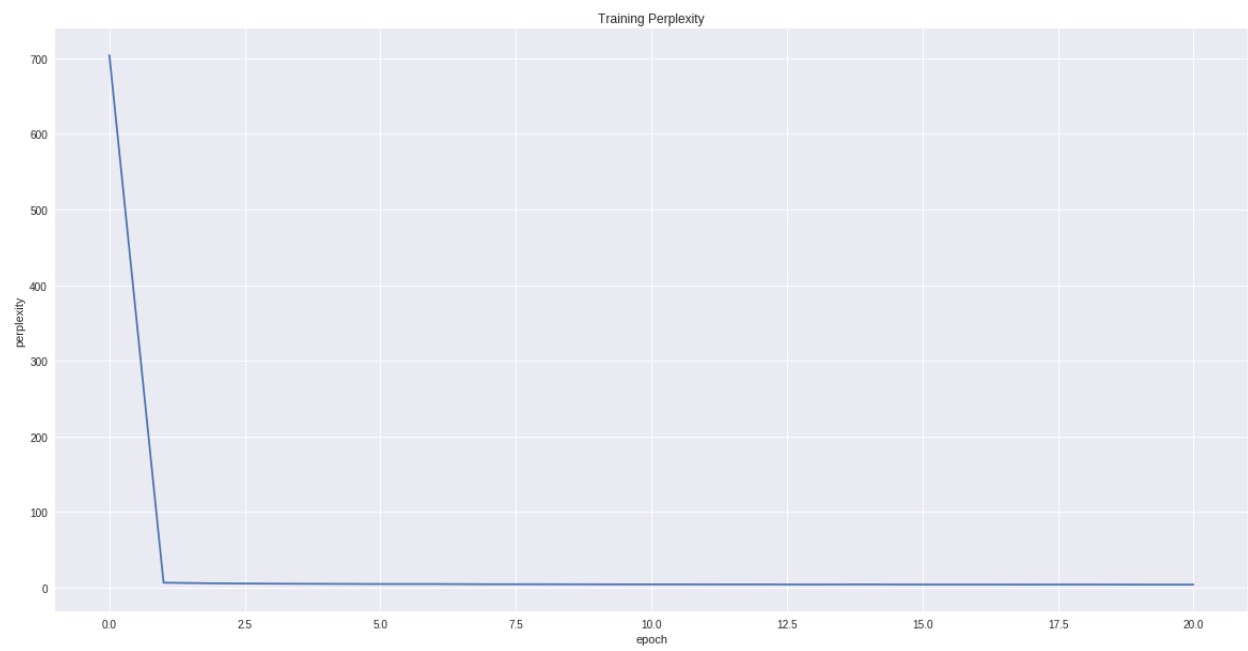
### Corpus Information

Name: Lord of the Rings, Number of Characters = 2579193

### Comparison of Sentence Generation on the Old and New Corpus

The sentences get into repeating a particular sentence a lot sooner with this new data. The test performance is lower too (see images below). This is likely because the new data set I have used is smaller (one third of the harry potter dataset in terms of file size).







- [illegible]

It's cool that 'Ring' but not 'ring' appears in the generated sentences. Also note that with a small temperature beam search has been caught up in repetition so is max. These are both the result of considering a probability distribution with more pronounced valleys and peaks.

## Short Answer Questions – Student Forcing

## Difficulties with Student Forcing

Training becomes slow and unstable.

## Comparison with Teacher Forcing

The results are better with teacher forcing as by feeding in the correct value we prevent the error from being accumulated. By helping the network in this way we prevent it from falling in a local minima.