COSC420 Assignment 2.

In this assignment, we are to first create our own word2vec models trained on the book/books pride and prejudice and war and peace. We are then to use the word 2 vec we have in part 1 to train a transformer model along with a transformer model with one hot encoding and compare the difference.

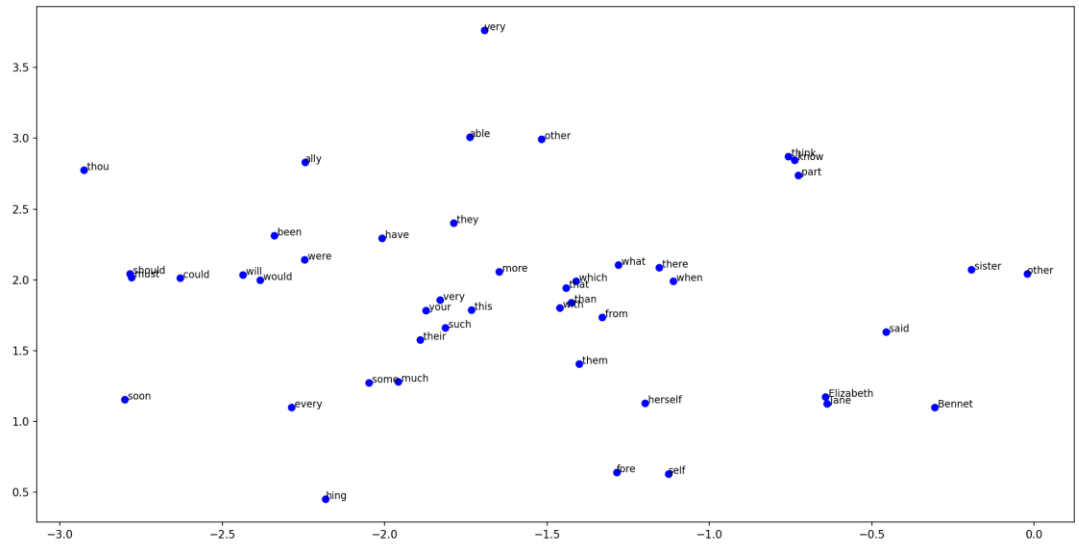
Task1

In task 1, I have created my own word2vec on just the Pride and prejudice book. I have just chosen the pride and prejudice due to memory constraints. Having both books or just War and Piece caused memory issues on my device. I begin by loading or training a tokeniser, which converts the text into numerical tokens for processing. I then create pairs of context and target tokens by pairing each word with the words surrounding it in the window size. I then separate the pairs into an array called context and an array called target. Also converting them to 1 hot encoding. For the skipgram model I created a 2 layer Neural network, with a embedding layer that outputs 200 dimensions (arbitrary number, seemed like enough without taking too long) and then a output layer that outputs a probability distribution on all the words.

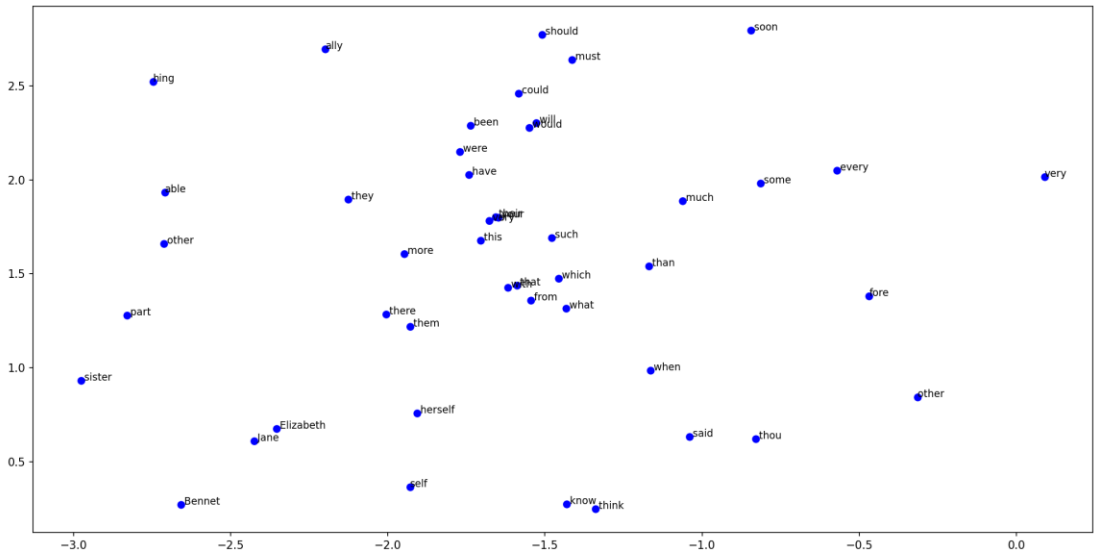
Below are some tests that I did to try to get a decent word2vec.

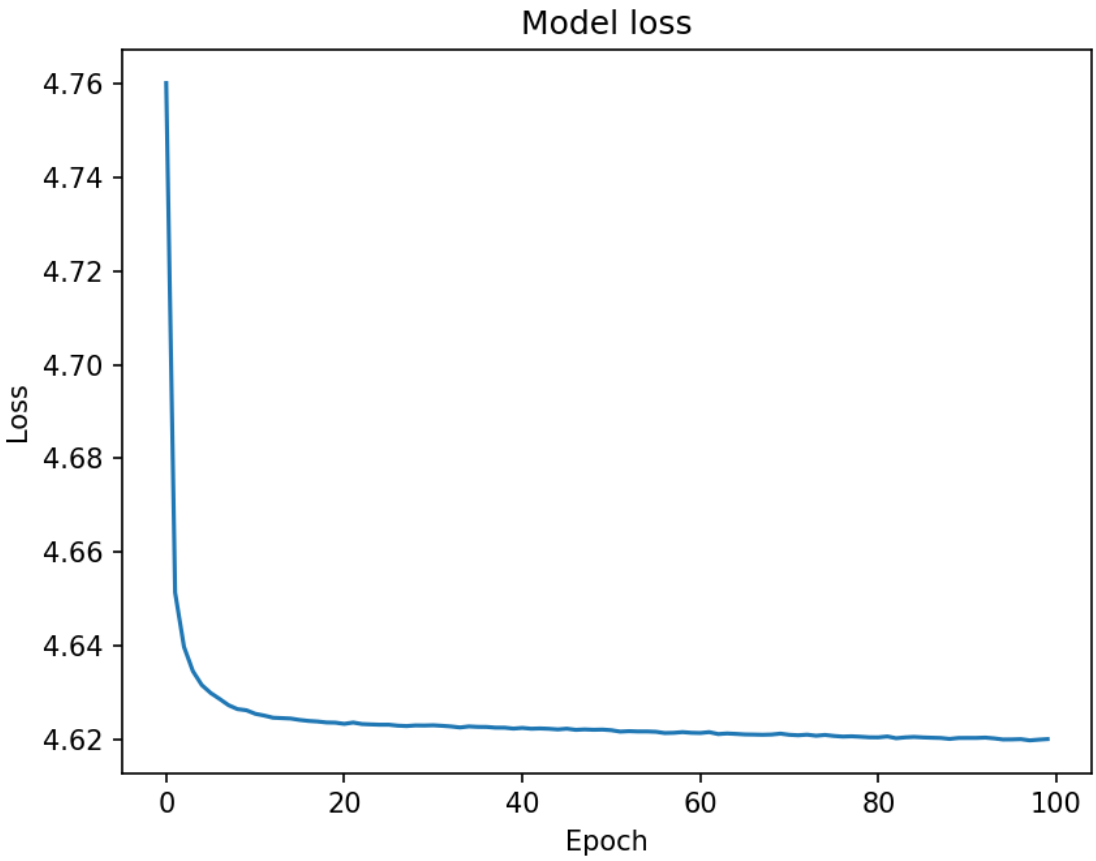
The first test I did was to see if the number of epochs mattered. So, I started with a 400-vocab sized, 200 dimensions model and trained it on 10 epochs vs 100 epochs. Below is the 2 TSNE graphs created

10 epoch



100 epoch

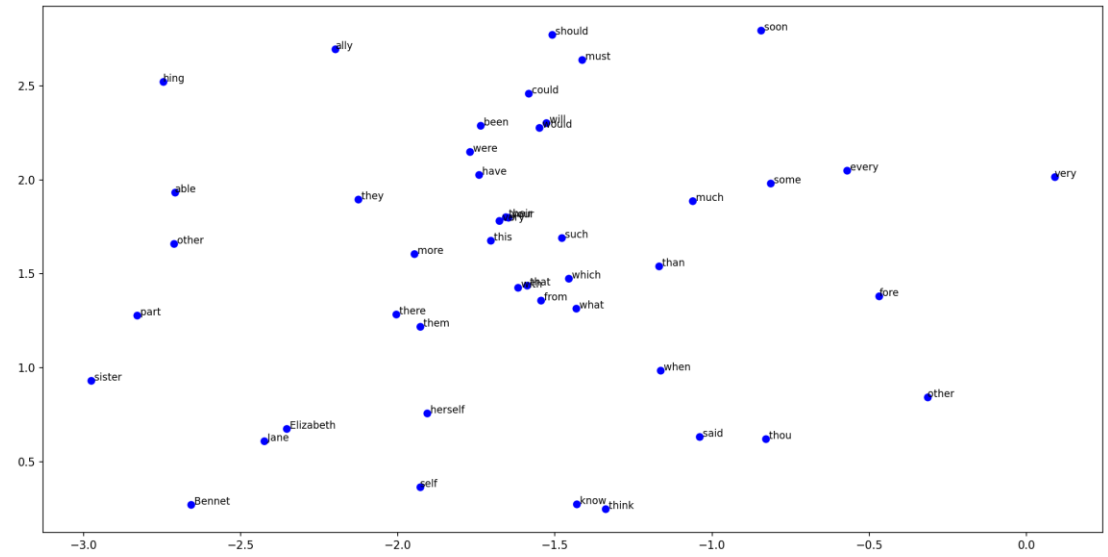




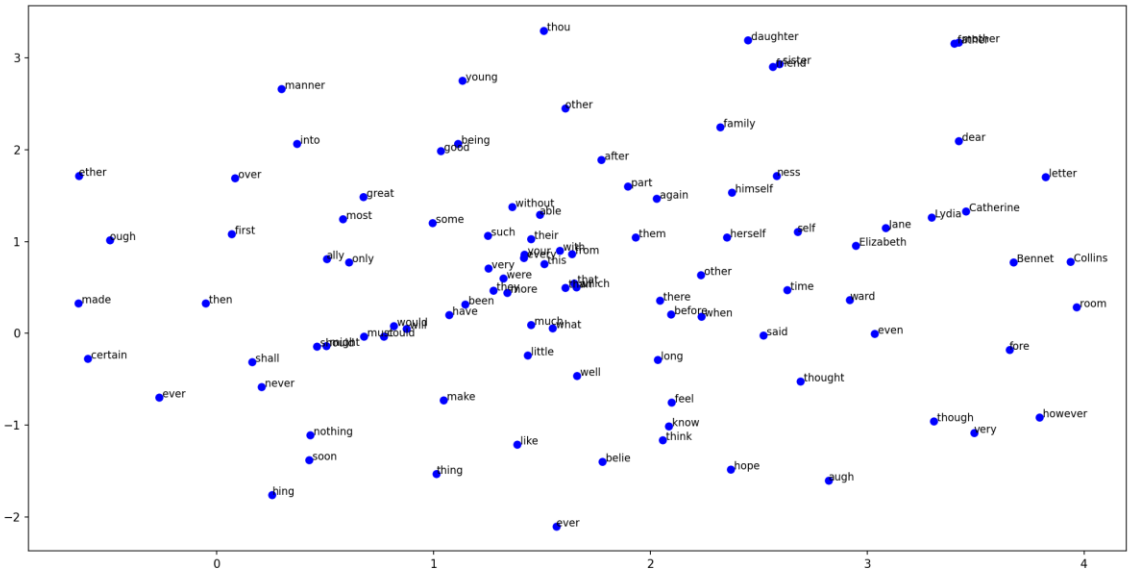
The TSNE graphs show clustering of words, indicating that the Word2Vec model is learning meaningful relationships. Words with similar contexts, like names and auxiliary verbs, are grouped together. Although the visual difference between 10 and 100 epochs is subtle, the loss reduction indicates improved performance with more epochs.

Another experiment I did was to see how well it performed with different number of tokens These are experiments between 400 and 600 as below 400 would be not enough tokens to plot and above 600 caused memory constraints. Below are the TSNE graphs of vocab sized 400 and 600 all trained to 100 epochs.

400



600



It seems like the 600 vocab_size performed very as not only does it have more tokens, but all the tokens seem to be grouped up well. As we can see in the 600-vocab graph. The top right area being more human focused such as names, relationships to people, and ways to address people. Bottom right being things that's humans think such as hope, L(augh), think, feel etc. and the middle to be things auxiliary verbs such as would, could, should etc. With the left-side being things that don't have connections to people such as certain, over, first. Overall, the extra sized vocab doesn't seem to negatively affect performance but instead increases the number of words and their relationship to each other. Although it is hard to tell much of a difference between these 2 plots, as the clusters seem to be similar. But the higher tokens does add more information.

There are multiple different factors that we can test out but due to time constraint, I am just going to keep this 600 token word2vec as it performs properly in establishing relationships between the different words. Where we can see that it clearly does group words with similar meaning together and words that are different a further away.

Task 2.

In task 2, we are to build a language model and train it on the text pride and prejudice. For this task I have followed a GPT style Architecture starting with a embeddings layer, positional encoding, and an x amount of transformer layers, ending with a dense layer. I have trained 3 models with each of the different embedding layers and below are some results. One thing to Note is that the one hot encoding needs to use the vocab_size as the amount of vector dimensions and causes the training time to be a lot longer. Each of the models have been trained on 6 transformer layers, 6 attention heads (Half the size of GPT), 600 vocab, 200 dimensions (word2vec shape and 600 dimensions for the 1hot encoding) and 75 epochs (time constraint) with 15 sequence length (Slightly above the minimum of 10 hoping to make it perform better on longer sentences). These parameters were chosen from intuition and was chosen as half the size of ChatGPT.

Embedding Layer	Accuracy on Train	Loss on Train
Fixed Embedding	81.94	0.6976
One Hot	84.98	0.5545
Trainable Embeddings	85.05	0.5762

From the accuracy and loss above it seems like the fixed embeddings performed the worse. Surprisingly worse than the one hot. But I suspect that to be because of the extra dimensions that one hot has as one hot's dimensions will end up being the same as the vocab_size. This was very prevalent in the training as the one hot took close to double the time to train compared to the fixed embedding's model. As well as the fact that the Fixed Embeddings is trained on a limited amount of data and potentially may not capture much meaningful relationships. The trainable embeddings on the other hand also performed well and better than the fixed embeddings model. This probably means that the trainable embeddings managed to pick up more meaningful relationships compared to the fixed embeddings as it was trained for a lot longer in comparison to the word2Vec Model. This probably means that the trainable embeddings picked up a lot more relationships that were specific towards the text throughout the training.

What I did next was test these models on some actual prompts. So, we can see the outputs of these models. Below are the prompts that were given. These are all predicting 300 tokens which was chosen as we have enough information to try deciphering what area of the book it is and how they perform with a "medium" sized text

Prompts

"It is a truth universally acknowledged" （quote about the book）

"" (nothing to see what it generates)

"Interpret Mr. Darcy's letter to Elizabeth" (prompt which has some context on the book)

Prompt1

Fixed Embedding

It is a truth universally acknowledged, that a single man in possession of a good fortune must be in want of a wife. However little known the feelings or views of such a man may be on his first entering a neighbourhood, this truth is so well fixed in the minds of the surrounding families, that he is considered as the rightful property of some one or other of their daughters. “ My dear Mr. Bennet, ” replied his wife, “ how can you be so teasing? ” “ I honour your circumspection. A fortnight’ s acquaintance is certainly very little. One cannot know what a man really is by the end of a fortnight. But if we do not venture, somebody else will; and after all, Mrs. Long and her nieces must stand their chance; and, therefore, as she will think it an act of kindness, if you decline the office, I will t

1Hot

It is a truth universally acknowledged, that a single man in possession of a good fortune must be in want of a wife. However little known the feelings or views of such a man may be on his first entering a neighbourhood, this truth is so well fixed in the minds of the surrounding families, that he is considered as the rightful property of some one or other of their daughters. “ My dear Mr. Bennet, ” replied his wife, “ how can you be so tiresome? You must know that I am thinking of his marrying one of them. ” “ Is that his design in settling here? ” “ Design? Nonsense, how can you talk so! But it is very likely that he may fall in love with one of them, and therefore you must visit him as soon as he comes. ” “ I see no occasion for that. You and the girls may go or you may send them by themselves, which per

Trainable Embeddings

It is a truth universally acknowledged, that a single man in possession of a good fortune must be in want of a wife. However little known the feelings or views of such a man may be on his first entering a neighbourhood, this truth is so well fixed in the minds of the surrounding families, that he is considered as the rightful property of some one or other of their daughters. “ My dear Mr. Bennet, ” replied his wife, “ how can you be so tiresome? You must know that I am thinking of his marrying one of them. ” “ Is that his desigent not solead, laughingly.

“ Oh, Jane! take care. ” “ My dear Lizzy, where can you have been walking to? ” was a question which Elizabeth received from Jane as soon as she entered the room, and from all the others when they sat down to table. “ What do you think? It is excellent news

Analysis

What was interesting was that all 3 models deviated from the book at the same time up to that point following the book word for word. Each model changing after “My dear Mr. Bennet”. All 3 models provide readable text mirroring the style of each other, although all go through their own tangent. Where the 1 hot and fixed embeds ended up on a similar context, trainable embeddings seemed to diverge quite a bit and it even had 2 made up words at “desigent” and “solead”. However, the other 2 performed fine when it came to spelling and generating real words. On the grammar and sentence structure side the models tend to generate text that is grammatically correct with proper sentence structure but occasionally errors do occur with missing punctuation or awkward phrasing. What was surprising was how all 3 ended up using a lot of quotation marks diving into what seems like conversations between people.

Prompt2

Fixed Embedding

away. When the tea things were removed, and the card tables placed, the ladies all rose; and Elizabeth was then hoping to be soon joined by him, when all her views were overthrown, by seeing him fall a victim to her mother' s rapacity for whist players, and in a few moments after seated with the rest of the party. She now lost every expectation of pleasure. They were confined for the evening at different tables; and she had nothing to hope, but that his eyes were so often turned towards her side of the room, as to make him play as unsuccessfully as herself. Mrs. Bennet had designed to keep the two Netherfield gentlemen to supper; but their carriage was, unluckily, ordered before any of the others, and she had no opportunity of detaining them. " Well, girls, " said she

1Hot

, and the carriage stopped at the small gate, which led by a short gravel walk to the house, amidst the nods and smiles of the whole party. In a moment they were all out of the chaise, rejoicing at the sight of each other. Mrs. Collins welcomed her friend with the liveliest pleasure, and Elizabeth was more and more satisfied with coming, when she found herself so affectionately received. She saw instantly that her cousin' s manners were not altered by his marriage: his formal civility was just what it had been; and he detained her some minutes at the gate to hear and satisfy his inquiries after all her family. They were then, with no other delay than his pointing out the neatness of the entrance, taken into the house; and as soon as they were in the parlour, he welcomed them a second time, with ostentatious

TrainableEmbeddings

, and the gentlemen did nothing but eat and admire. When the ladies returned to the drawing - room, there was little to be done but to hear Lady Catherine talk, which she did without any intermission till coffee came in, delivering her opinion on every subject in so decisive a manner as proved that she was not used to have her judgment controverted. She inquired into Charlotte' s domestic concerns familiarly and minutely, and gave her a great deal of advice as to the management of them all; told her how everything ought to be regulated in so small a family as hers, and instructed her as to the care of her cows and her poultry. Elizabeth found that nothing was beneath this great lady' s attention which could furnish her with an occasion for dictating to others. In the intervals of her discourse with Mrs. Collins

Analysis

When given nothing as the prompt, each model seems to take from an area in the book and build around it. Interestingly the fixed embedding started with a word while the other 2 models start with a comma. All 3 models seem to generate sentences that are readable and maybe represents a part of the book? (Haven't read the book) and what seems like maintaining the author's style. They all adhere to English grammar and sentence structure with occasional errors. What was interesting is that all 3 models have used semi colons in their generation although not commonly used in the real world, could be potentially used in the book? Overall, Fixed Embedding provides a seamless continuation, 1 Hot excels in descriptive scenes, and Trainable Embeddings maintains a detailed and instructive narrative. All 3 models were readable although sometimes with an extra space or an awkward grammar. All the sentences make sense to some point.

Prompt 3

Fixed Embeddings

Interpret Mr. Darcy's letter to Elizabeth? Is it not so, Mrs. Bennet? ” “ Yes, or I will never see her again. ” “ An unhappy alternative is before you, Elizabeth. From this day you must be a stranger to one of your parents. Your mother will never see you again if you do not marry Mr. Collins, and I will never see you again if you do. ” Elizabeth could not but smile at such a conclusion of such a beginning; but Mrs. Bennet, who quarrelled with no compliments, answered most readily, “ You are very kind, sir, I am sure; and I wish with all my heart she were well settled. But with such a father and mother, and such low connections, I am afraid there is no chance of it. ” “ I think I have heard you say that their uncle is an attorney in Meryton? ” “ Yes; and they have another, who lives somewhere near Che

1 Hot

Interpret Mr. Darcy's letter to Elizabeth, and, after listening for a few minutes, said to Darcy, “ Miss Bennet would not play at all amiss if she practised more, and could have the advantage of a London master. She has a very good notion of fingering, though her taste is not equal to Anne’ s. Anne would have been a delightful performer, had her health allowed her to learn. ” Elizabeth looked at Darcy, to see how cordially he assented to his cousin’ s praise: but neither at that moment nor at any other could she discern any symptom of love; and from the whole of his behaviour to Miss De Bourgh she derived this comfort for Miss Bingley, that he might have been just as likely to marry her, had she been his relation. Lady Catherine continued her remarks on Elizabeth’ s performance, mixing with them many instructions on execution and taste.

Trainable Embeddings

Interpret Mr. Darcy's letter to Elizabeth, that your ambition almost unwilling, but I must leave you. I must find Mr. Gardiner this moment on business that cannot be delayed; I have not an instant to lose. ” “ Good God! what is the matter? ” cried he, with more feeling than politeness; then recollecting himself, “ I will not detain you a minute; but let me, or let the servant, go after Mr. and Mrs. Gardiner. You are not well enough; you cannot go yourself. ” Elizabeth hesitated; but her knees trembled under her, and she felt how little would be gained by her attempting to pursue them. Calling back the servant, therefore, she commissioned him, though in so breathless an accent as made her almost unintelligible, to fetch his master and mistress home instantly.

Analysis

For the prompt to interpret Mr. Darcy’s letter, according to information from GPT（understands the book more than I do） the Fixed Embedding model includes relevant dialogue but strays from directly interpreting the letter, focusing more on interactions between characters. The sentence structures are mostly dialogue-based and maintain a natural flow. The 1 Hot model provides a scene critique that, while well-written, diverges from the task of interpreting the letter, with clear and varied sentence structures. Trainable Embeddings generates a dramatic and emotionally intense scene, but it lacks focus on the interpretation of the letter, with complex and dramatic sentences. Overall, none of the models directly address the prompt to interpret the letter, with Fixed Embedding and 1 Hot providing engaging dialogue and scenes, and Trainable Embeddings offering high emotional intensity. Unfortunately, there isn’t enough data for the model to provide analysis, so it just starts spouting words from parts of the book.

Overall, all 3 models performed similarly. With what seems like sentences that mirrored the author's style but also suffer from some issues regarding punctuality and formatting. They seem to take the input and select some scene in the book to imitate. Most of the time creating outputs that are readable and make some sort of sense that a reader that that has read the book could potentially understand what area of the book the model is referring to.

Something to not that wasn't included was that in poorer performing models (smaller models trained on low epochs), they would often start repeating the same sentences/few words. Notably this did not happen in these models after I trained to 10,000 tokens. Even after generating this amount of text, the quality of text seems to remain constant. Below is a snippet of the end of the Trainable Embeddings models when predicting 10,000 tokens.

"The letter which she wrote on this occasion to her sister will prove what she felt: " My dearest Lizzy will, I am sure, be incapable of triumphing in her better judgment, at my expense, when I confess myself to have been entirely deceived in Miss Bingley' s regard for me. But, my dear sister, though the"

As seen above, it keeps its sentence structure even during large amounts of texts.

When it comes to which one is the best performer, from my personal reading it seems like they are all similar with small differences such as made up words by the trainable embeddings model, but when it comes to chat GPT they claim that 1 Hot performs the best. According to the below reasons

- The 1Hot model consistently maintains the original tone and narrative style of Jane Austen's "Pride and Prejudice" across all prompts.
- It provides a coherent narrative flow and character interactions, staying true to the source material.
- While it also ends abruptly in some instances, it does so in a less jarring manner compared to the other models.

While the **Fixed Embedding** model also performs well but tends to truncate more noticeably, while the **Trainable Embeddings** model diverges more significantly from the original style and coherence.

In conclusion, the text that the models generated had a lot of resemblance to the original text which I am quite happy with. Unfortunately, due to time constraints I could not try maximizing the model as training some of these models took hours on hours. There would have been a lot of different parameters that I would have liked to test but couldn't. Furthermore, all the models that I trained all fairly well, producing sentences which have grammar that matches the original text (I think) and I don't think I could of told the difference between them without analysis by ChatGPT.