Token indices sequence length is longer than the specified maximum sequence length for this model (1311 > 1024). Running this sequence through the model will result in indexing errors

Using device: cuda

Checkpoint 2.1:
Chunked Tensors for "Enchanted(Taylor's Version)": tensor([[50256, 1858,   314,   373,   757,
9975,   198, 1890, 2259, 20263,
         11,   277,   868, 21845,   198, 30556, 1468, 10032, 11, 21757,
       1295,   198,   54, 5691,   286, 1035, 40310,   198, 2484, 13309,
       2951,   290, 39836,   198, 25298, 1348,   618,   314, 2497,   534,
       1986,   198, 3237,   314,   460,   910,   318,   340,   373, 23260,
        278,   284, 1826,   345,   198,   198, 7120, 2951, 25029,   11,
        366, 11980,   356, 50256],
        [50256, 1138, 1701,   198, 40553,   262, 2119,   11,   534, 41834,
        198, 1273, 5889,   284,   787,   663,   835,   284,   502,   198,
        464, 34264, 5273, 4940,   198, 31694,   477,   534, 2068, 10252,
        198, 7594, 6427, 4710,   287, 22780,   198, 1870,   340,   373,
      23260,   278,   284, 1826,   345,   198, 3237,   314,   460,   910,
        318,   314,   373, 42282,   284, 1826,   345,   198,   198, 1212,
       1755,   318, 9009, 50256],
        [50256, 2815, 3256,   836,   470,   345, 1309,   340,   467,   198,
         40, 1101, 4240, 19554,   694,   11, 37854,   259,   6,   477,
        262,   835, 1363,   198,   40, 1183, 4341, 8097, 4240,   259,
          6,   611,   345, 2993,   198,   40,   373, 42282,   284, 1826,
        345,   198,   198,   464, 28528, 1808, 4030,   502,   510,   198,
         17,   257,   13,   76, 1539,   508,   466,   345, 1842, 30,
        198,   40, 4240, 50256],
        [50256,   705, 47163,   314, 1101, 3094, 21693,   198, 1870,   783,
         11,   314, 1101, 37572,   736,   290, 6071,   198,   54, 3929,
        345,   547,   379,   616, 3420,   198, 40, 1549, 1280,   510,
        290,   345,   561,   910,   198,   1, 10814,   11,   340,   373,
      23260,   278,   284, 1826,   345,   1,   198, 3237,   314,   760,
        318,   314,   373, 42282,   284, 1826,   345,   198,   198, 1212,
       1755,   318, 9009, 50256],
        [50256, 2815, 3256,   836,   470,   345, 1309,   340,   467,   198,
         40, 1101, 4240, 19554,   694,   11, 37854,   259,   6,   477,
        262,   835, 1363,   198,   40, 1183, 4341, 8097, 4240,   259,
          6,   611,   345, 2993,   198, 1212, 1755,   318, 38870,   11,
        836,   470,   345, 1309,   340,   467,   198,   40, 1101, 4240,
      19554,   694, 11, 15360, 1088,   477, 3436,   198,40, 1183,
       4341, 8097, 4240, 50256],
        [50256,   259, 6,   611,   345, 2993,   198,   40,   373, 42282,
        284, 1826,   345,   198,   198, 1212,   318,   502, 26002,   326,

```
    198, 1212,   373,   262,   845,   717, 2443,   198, 3673,   810,
    262, 22992, 5645,   198, 3666, 6066,   481, 9809,   534, 1438,
    198, 18273,   314,   766,   345,   757,   198, 4711,   389,   262,
   2456,   314, 2714,   736,   198, 1722,   314,   373, 4305, 1165,
   2582,   198,    40, 50256],
  [50256,   373, 42282,   284, 1826,   345,   198, 5492,     11,   836,
    470,   307,   287, 1842,   351, 2130, 2073,   198, 5492,   11,
    836,   470,   423, 8276, 4043,   259,         6,   319,   345,   198,
   5492,   11,   836,   470,   307,   287, 1842,   351, 2130, 2073,
    357,    46, 1219,      8,   198, 5492,        11,   836,   470,   423,
   8276, 4043,   259,    6,   319,   345,   357, 5812,   12, 1219,
      8,   198,   198, 50256],
  [50256, 1212, 1755,   318, 9009, 2815, 3256,   836,   470,   345,
   1309,   340,   467,   198,      40, 1101, 4240, 19554,   694,        11,
  37854,   259, 6,   477,   262,   835, 1363,   198,    40, 1183,
   4341, 8097, 4240,   259,      6,   611,   345, 2993,   198, 1212,
   1755,   318, 38870,   357, 5492,     11,   836,   470,   307,   287,
   1842,   351, 2130, 2073,      8,   198, 3987,   470,   345, 1309,
    340,   467,   198, 50256],
  [50256,        40, 1101, 4240, 19554,   694,   357, 5492, 11,   836,
    470,   423, 8276, 4043,   259,         6,   319,   345,          8,   198,
    35, 5077, 1088,   477, 3436,   198, 40, 1183, 4341, 8097,
    357, 5492,     11,   836,   470,   307,   287, 1842,   351, 2130,
   2073, 8,   198, 42337,   259,            6,   611,   345, 2993,   198,
    40,   373, 42282,   284, 1826,   345,   198, 5492,   11,   836,
    470,   307,   287, 50256],
  [50256, 1842,   351, 2130, 2073,   198, 5492,     11,   836,   470,
    423, 8276, 4043,   259,      6,   319,   345, 50256, 50256, 50256,
  50256, 50256, 50256, 50256, 50256, 50256, 50256, 50256, 50256, 50256,
  50256, 50256, 50256, 50256, 50256, 50256, 50256, 50256, 50256, 50256,
  50256, 50256, 50256, 50256, 50256, 50256, 50256, 50256, 50256, 50256,
  50256, 50256, 50256, 50256, 50256, 50256, 50256, 50256, 50256, 50256,
  50256, 50256, 50256, 50256]])
```

Checkpoint 2.2:
logits shape: (batch_size, seq_len, vocab_size)
hidden_state shape: (1, batch_size, rnn_hidden_dim)

Checkpoint 2.3:

Training Set Loss Curves
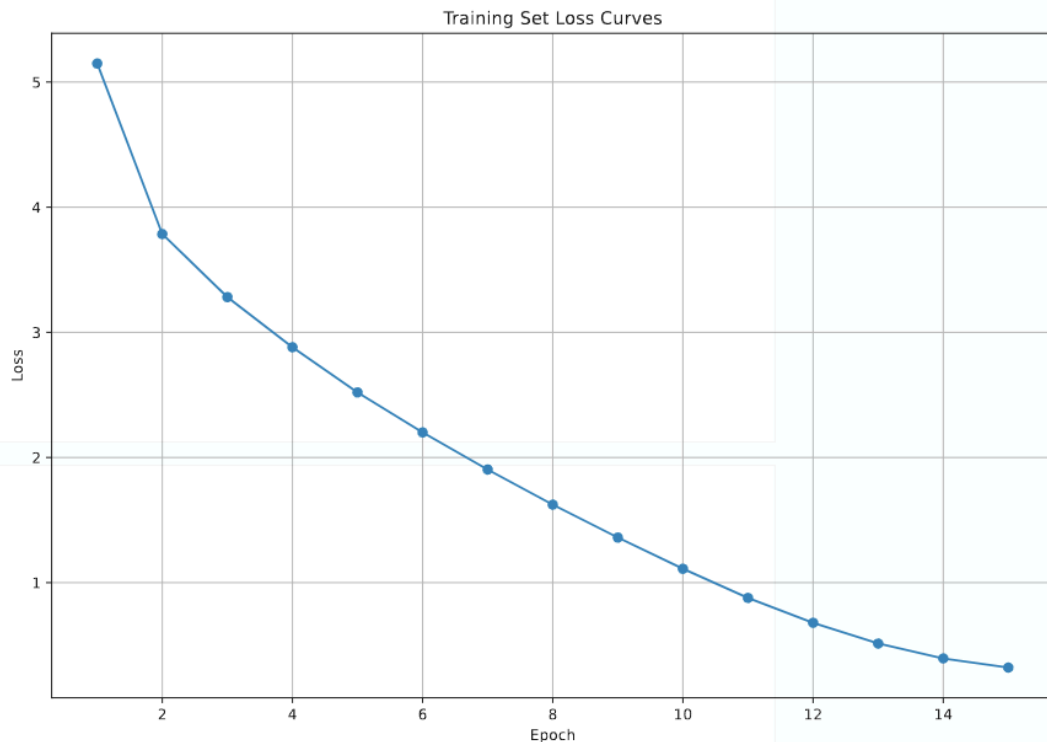
Sample: "And you gotta live with the bad blood now"
Perplexity: 10699.0966

Sample: "Sit quiet by my side in the shade"
Perplexity: 1355.9899

Sample: "And I'm not even sorry, nights are so starry"
Perplexity: 1761.7063

Sample: "You make me crazier, crazier, crazier, oh"
Perplexity: 1530.3083

Sample: "When time stood still and I had you"
Perplexity: 5009.6391

OBSERVATIONS:
The model here does indeed perform better than the TrigramLM from Part1, in terms of
perplexity and different orders of magnitude depending on the sample. I believe it works better
due to a much more thoroughly trained model than just a basic TrigramLM(which essentially just
stores counts and guesses). The RNN LM has multiple separate stages, is trained multiple

separate times, and optimizes based on loss(with batch gradient descent as we used here). This allows it to be much more accurate than before and as a result have a lower perplexity with the same samples(bar the G/space symbol of course).

Checkpoint 2.4:
Start phrase "<s>Are we":
 we out of the woods yet? Are we out of the woods yet?
Are we out of the woods yet? Are we out of the woods yet?
Are we out of the woods yet? Are we out of the woods yet?
Are we out of the woods yet? Are we out of the woods

Start phrase "<s>Like we're made of starlight, starlight":
light

"I want you, bless my soul (He got my heartbeat)
I want you, bless my (Skipping down 16th Avenue, baby)
I want you, bless my soul (He got my heartbeat)
I want you, bless my (Skipping down 16th Avenue

Start phrase "<s>Why can't I":
 I showed
The highway don't care only you
But I don't know

You're not sorry
No, no, no

You're not sorry
No, no, no

No, no, no, no
No, no, no
No, no, no
No,