

## Notes:

- 1) Our group decided to count punctuation marks as a word type and word token. However, during preprocessing, we decided to remove punctuation from the beginning and the end of sentences. This is because punctuation does not make sense at the beginning of a sentence, and we replace the punctuation at the end of the sentence with a </s> token.
- 2) For now, we hardcoded the Unigram sentences to be 10 tokens long, but we can change this if necessary. Bigram sentences are at most 10 tokens. Sentences generated by seeding are at most 10 tokens plus the initial starting tokens.
- 3) For the seeding sentences, we always start with the three starting phrases:
  - The movie was
  - I am
  - The film
- 4) We assume that we do not have to perform seeding for unigram sentences, because it doesn't make much sense.

## Sample Sentences:

### Positive Unigram Sentences:

of about writer\director offering as is and trust years Bound  
ambitious were : parable between sentimental who entertaining anymore this  
8 whimsical Shyamalan about her insane horror Ash While with

### Negative Unigram Sentences:

see 's for suffering madcap were Shreve the your a  
a races sporting make ca a disingenuous there does appears  
drudgery In the genuinely the The , writer-director , the

### **Positive Bigram Sentences:**

<s> A rip-roaring comedy film work may be , powerful </s>  
<s> A colorful , but wanes in this gentle humor of </s>  
<s> A big-budget\all-star movie that keep the line , like those </s>

### **Negative Bigram Sentences:**

<s> Make Chan and you put together and entirely improvised </s>  
<s> A culture clash between not-very-funny comedy that was the duration </s>  
<s> An excruciating demonstration of believability of Love does n't Stop </s>

### **Positive Bigram Sentences with seeding:**

The movie was so that freshly considers arguments the film pantheon </s>  
I am to be a questioning the project in reality into modern </s>  
The film of Alexander Payne constructs a movie , touching </s>

### **Negative Bigram Sentences with seeding:**

The movie was a relatively short of wreckage that , it also looking </s>  
I am baffled by far too weak script and a toddler </s>  
The film lacks the most visually bland animated sequel that would play </s>

## **Analysis:**

The sentences generated by the bigram model are more realistic and natural compared to those generated by the unigram model. The sentences generated by the unigram model have an unnatural sentence structure. For example, the bigram model sentences always correctly start with a capitalized word. This is not always the case for the unigram model sentences. In addition, the unigram sentences may contain capitalized words in the middle of a sentence. Also, the bigram sentences use adjectives more effectively and naturally. For example the bigram sentences have phrases like 'rip-roaring comedy film' and 'excruciating demonstration'. In each case, it is clear what the adjective is modifying. However, in the unigram sentences, adjectives are often out of place and do not modify a specific noun.

The sentences generated by the bigram model generally convey positive and negative sentiments better than those generated by the unigram model. For example, the sample positive sentences generated by the unigram model are all very vague. It is not clear that these are positive reviews. On the other hand, the third positive review produced by the bigram model includes the phrase 'A big-budget\all-star movie', which clearly indicates that this is a positive review. Similarly, the negative reviews generated by the unigram model are vague and unintelligible. None of the three sample reviews generated by the unigram model are clearly

negative. In contrast, the second negative review generated by the bigram model contains the phrase 'not-very-funny comedy', which is clearly negative. Therefore, the bigram model sentences are more effective at conveying negative or positive opinions.

The seeding model chose phrases that make sense in the context of the starting phrase. For example, in the first negative bigram sentence generated by seeding, the starting phrase 'The movie' was followed by 'was a relatively short'. In the second negative sentence generated by seeding, 'I am' is followed by 'baffled by'. In the third negative sentence generated by seeding, 'The film' is followed by 'lacks'. In all three instances, the phrase chosen by the seeding model makes sense in conjunction with the previous phrase.