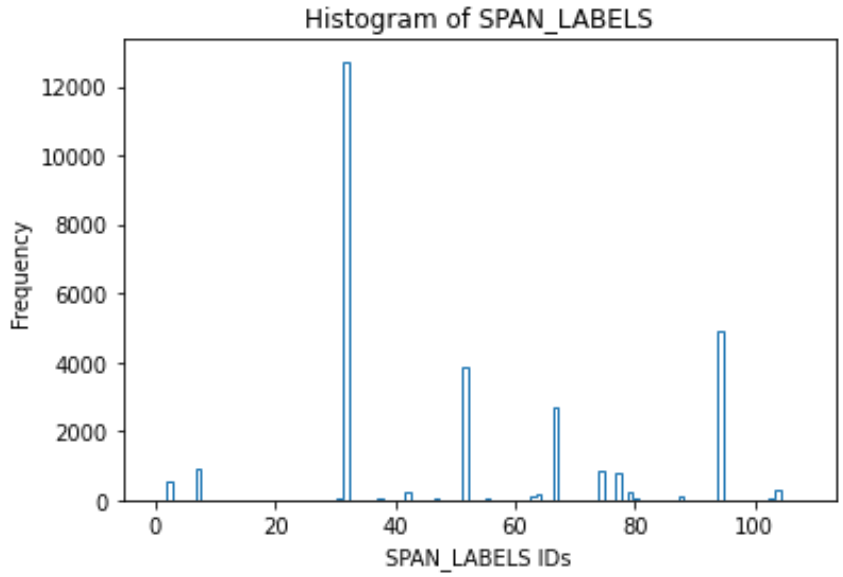
Homework 3
Yorick Chern
COMPSCI 288

Error Analysis

In this report, I will be focusing on the Parsing Model's performance across different sentence categories. To more effectively look for relevant sentence categories, I created a histogram that maps out the frequency distribution of all the sentence categories.



With some code, I have found the most frequent sentence categories, as can be seen in this table below. Here, I show the top 5 most relevant sentence categories.

Category	Frequency
NP	12719
VP	4913
PP	3893
S	2709
ADVP	889

Then, based on these specific tags, I calculated the recall for each of the tags on the dev set. In this experiment and report, I was particularly interested in the model's ability to correctly classify these relevant tags.

Category	Recall
NP	0.8949
VP	0.8915
PP	0.8402
S	0.9077
ADVP	0.7862

As can be expected, the noun phrases and verb phrases (NP and VP, respectively) are the most frequent phrases as a sentence is mainly composed of a noun and a verb. Having sufficient data with the labels of NP and VP and the distinctive nature between nouns and verbs, it is expected for NP and VP to both have high recall. The next most relevant tag is the prepositional phrase (PP), with slightly lower performance. PP's are known to begin with and contain very specific words such as "to", "of", "about", "at", etc. Thus, it is not hard to imagine that the model can pick up these words using the multihead attention mechanism. Then, we have the entire sentence, denoting S. I think this was slightly surprising to me because I thought the model may very well confuse the entire sentence as just yet another sentence category. However, with the model seeing many different categories grouping together to form a sentence, the model definitely picked up patterns that indicated the composition of a whole sentence. Last but not least, we have ADVP, which is the adverb phrase. In my opinion, it makes sense that ADVP has the lowest recall. In a sentence, ADVP is used to modify other expressions, such as the phrase "she ate the breakfast quickly." That said, it makes sense to me that recognizing that a phrase is used to modify another expression instead of just another phrase to describe a different event could be difficult to model. It is also worth noting that a lot of the adverbs share subwords with other adjectives and phrases, so this confusion only makes the classification task more difficult.