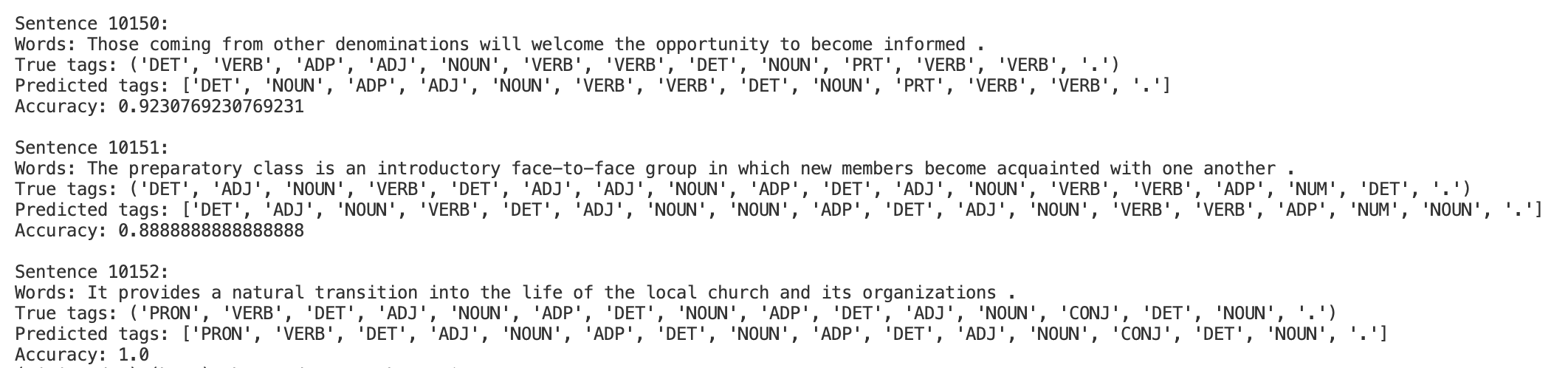
Su Zhang

POS Tagging Explanation

**Compare against the truth. Explain why your POS tagger does or does not produce the correct tags.**



The accuracy results between the output from my POS tagger model and the true tags are between 0.8 and 1.0, meaning that my POS tagger model did not catch all the correct tags.

I think there are a few reasons why:

1. The training data is limited (only 10,000 sentences), and cannot capture the accurate distribution of possibilities of tags
2. Some words have multiple POS tags and my model might have less common usages
3. HMM only considers the previous state when simulating the next state / tag, and thus my model is limited when dealing with more complex sentence structure
4. Add-1 smoothing prevents zero probabilities, but at the same time led to possibly overestimate the probability of uncommon occurrence