PROBLEM 8 – Markov Assumption

In this homework, the Markov assumption is implicitly applied when we calculate the probability of a word in a sequence depending only on the previous word. This corresponds to a first-order Markov assumption, where a word's probability depends on the immediately preceding word.

If we were to apply a second-order Markov assumption, using trigrams, it means that a word's probability would depend not only on the previous word but on the previous two words. This would introduce a higher level of context and would be more restrictive because it would consider the last two words in the sequence.

This is how accuracy results might be affected by using trigrams

- Data Sparsity. Trigrams would require a larger dataset to have sufficient examples of each unique trigram. This could be challenging, especially for languages with complex grammatical structures.
- Increased Complexity. The model would be more complex computationally and in terms of implementation. The time and resources required for training and inference would also increase.
- Risk of Overfitting. With more parameters, there's a higher risk of overfitting the training data, especially if the dataset is not sufficiently large.
- Loss of Generalization. The model might become too specific and have difficulty generalizing to new and unseen data.