| Experiment No.5 |
| --- |
| Implement Bi-Gram model for the given Text input |
| Date of Performance: |
| Date of Submission: |

**Aim:**  Implement Bi-Gram model for the given Text input

**Objective:** To study and implement N-gram Language Model.

**Theory:**

A language model supports predicting the completion of a sentence.

Eg:

* Please turn off your cell \_\_\_\_\_
* Your program does not \_\_\_\_\_\_

Predictive text input systems can guess what you are typing and give choices on how to complete it.

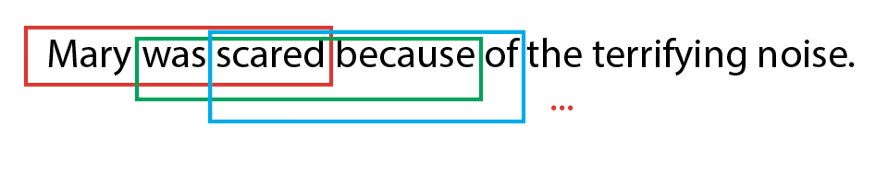
**N-gram Models:**

Estimate probability of each word given prior context.

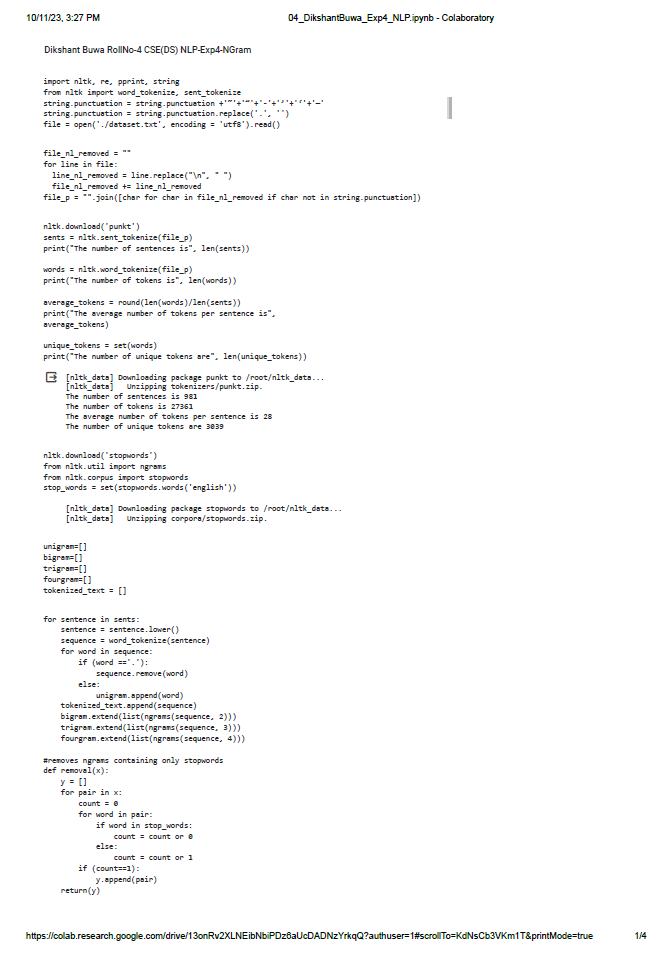
P(phone | Please turn off your cell)

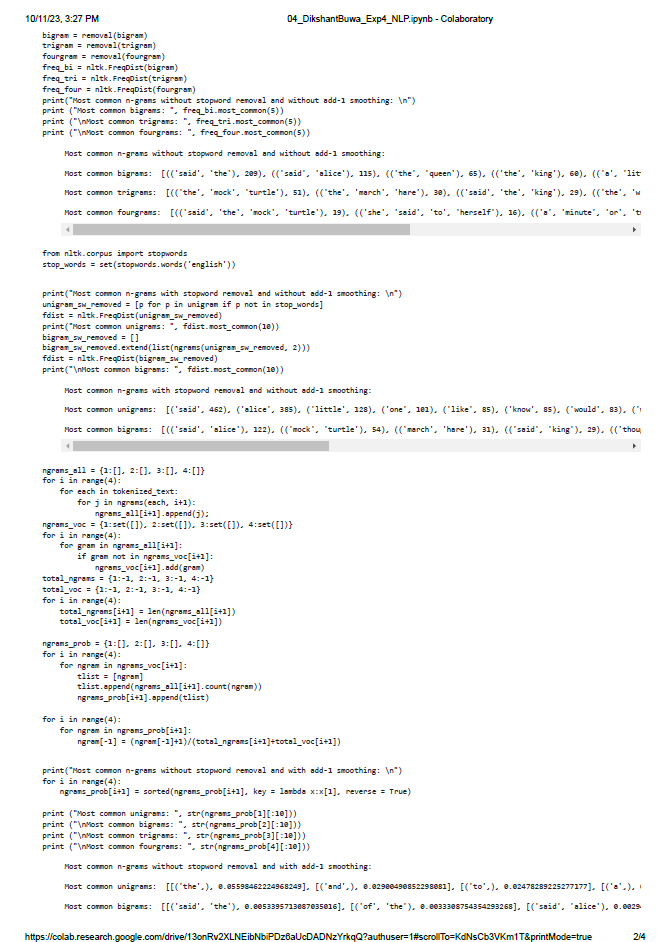
* Number of parameters required grows exponentially with the number of words of prior context.
* An N-gram model uses only N1 words of prior context.
  + Unigram: P(phone)
  + Bigram: P(phone | cell)
  + Trigram: P(phone | your cell)
* The Markov assumption is the presumption that the future behavior of a dynamical system only depends on its recent history. In particular, in a kth-order Markov model, the next state only depends on the k most recent states, therefore an N-gram model is a (N1)-order Markov model.

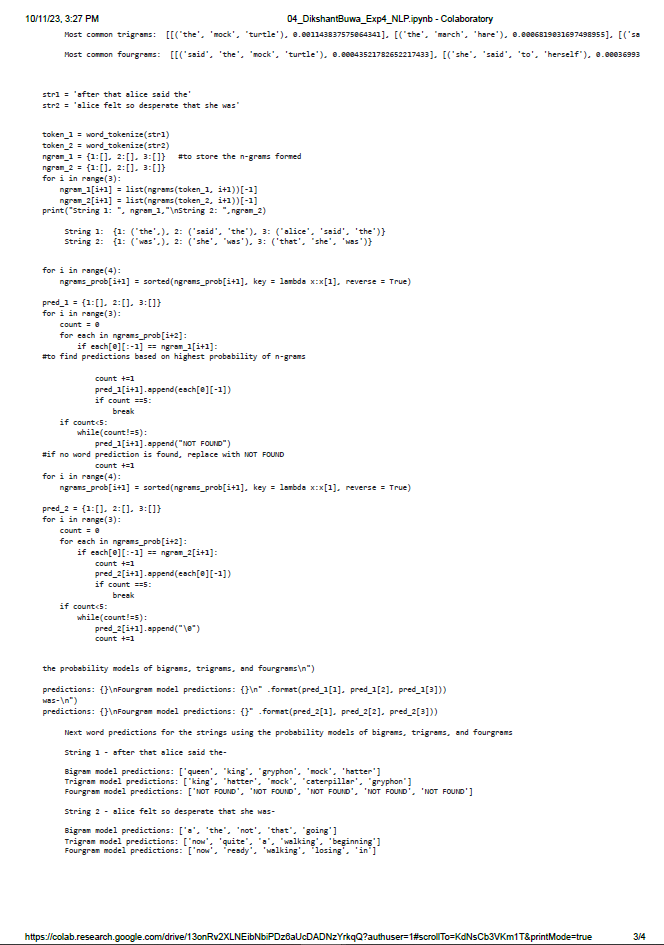
**N-grams**: a contiguous sequence of n tokens from a given piece of text



**Fig**. Example of Trigrams in a sentence







**Conclusion:**

The N-gram language model is a simple and widely used approach for natural language processing tasks, such as text generation and speech recognition. It operates by analyzing the statistical relationships between words in a given text, with "N" representing the number of preceding words considered for prediction. While N-gram models are easy to implement and computationally efficient, they have limitations in capturing long-range dependencies and understanding context. As a result, they may struggle with handling more complex language tasks compared to more advanced models like recurrent neural networks or transformer-based models. In conclusion, N-gram language models are a valuable tool for certain applications but may fall short in tasks that require a deeper understanding of language and context.