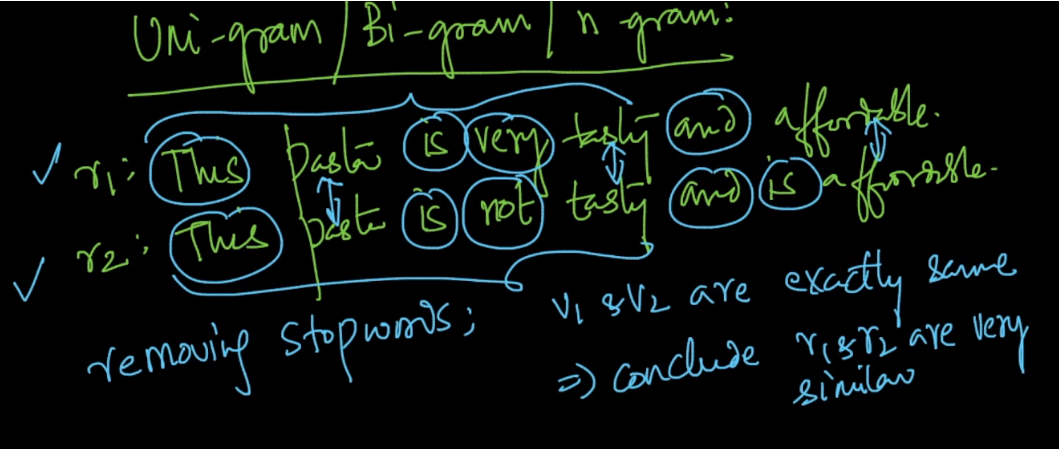
**N-gram:**

In the fields of [computational linguistics](https://en.wikipedia.org/wiki/Computational_linguistics) and [probability](https://en.wikipedia.org/wiki/Probability), an ***n*-gram** is a contiguous sequence of *n* items from a given [sample](https://en.wikipedia.org/wiki/Sample_(statistics)) of text or speech

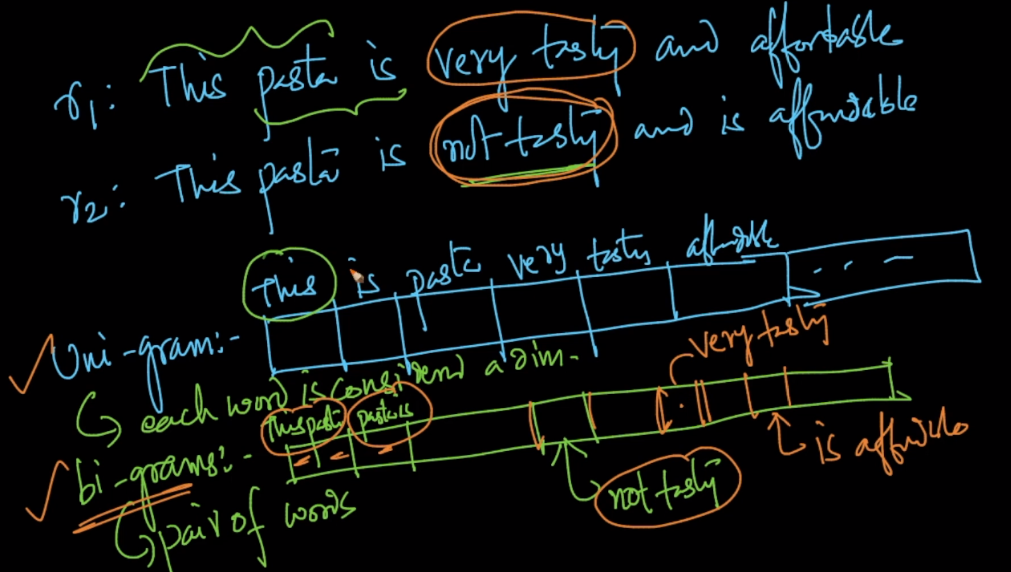


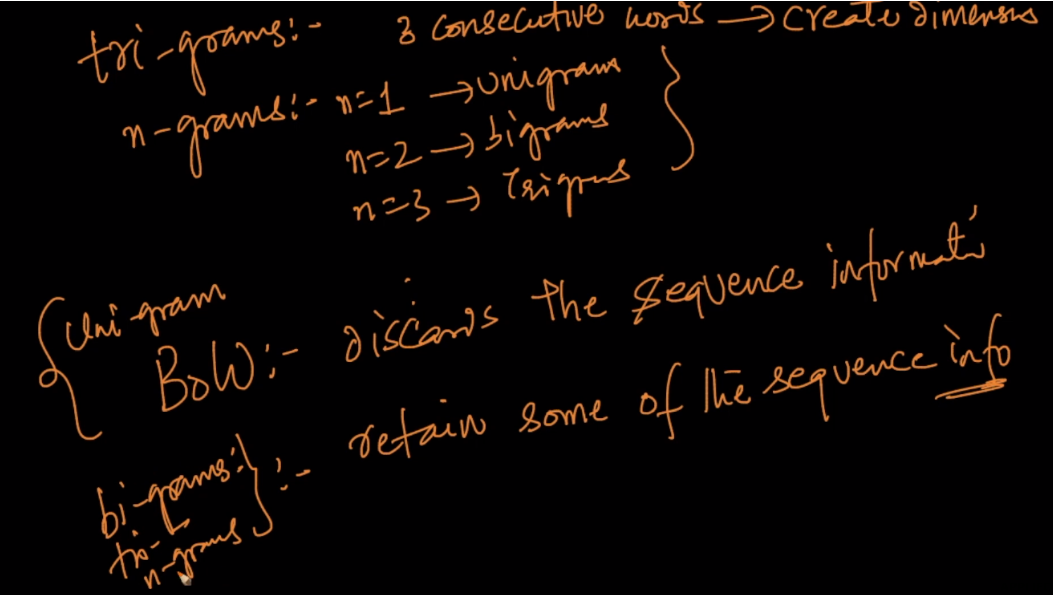
Example for unigram we take a single word for each dimension

For bigram we take two consecutive word for each dimension, therefore for r1, v1 will be

**This pasta, pasta is, is very, very tasty, tasty and, and affordable.**

Similarly for N-gram we take n consecutive words for each dimension.





After creating N-grams if any dimension contains only stop words, then remove that dimension.

Ex: if any dimension contain “is very”, then this dimension will be removed.

