

# THE N-GRAM

- The most fundamental unit in NLP is the **n-gram**.
- An **n-gram** is simply a sequence of **N** words. For instance:
  - Dog (is a unigram,  $N=1$ )
  - San Francisco (is a bigram,  $N=2$ )
  - The Three Musketeers (is a trigram,  $N=3$ )
  - She stood up slowly (is a quadgram,  $N=4$ , which is rarely used)
  - Et cetera...

