- 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…

- Any sentence can be decomposed into **n-grams**.

- For instance, consider the sentence "The quick fox jumped."
  - The **unigrams** are: (1) ***the***, (2) ***quick***, (3) ***fox***, and (4) ***jumped*** (four unigrams)
  - The **bigrams** are: (1) ***the quick***, (2) ***quick fox***, and (3) ***fox jumped*** (three bigrams)
  - The **trigrams** are: (1) ***the quick fox***, and (2) ***quick fox jumped*** (two trigrams)

- **N-grams** help to resolve **polysemy** and **homonymy**, as they capture context with additional surrounding words, like adjectives and adverbs.

and adverbs.

capture context with additional surrounding words, like adjectives

- **N-grams** help to resolve **polysemy** and **homonymy**, as they

  (two trigrams)

  - The trigrams are: (1) **the quick fox**, and (2) **quick fox jumped**
  - Jumped (three-grams)