

- Index for words
  - two bytes allow a vocabulary of  $2^{16} = 65,536$  words, typically more needed
  - Huffman coding to use fewer bits for frequent words.
- Probabilities
  - typically stored in log format as floats (4 or 8 bytes)
  - quantization of probabilities to use even less memory, maybe just 4-8 bits