+ Word based encoding

ASKi => Number value to letter.
No semantics do.

LiالسTEN vs SILENT

value per word. incead is more usefull.

+ Using API's

TOKENiZER -> Strip punctuation
-> Adds numerical value.
-> We make a corpus.

+ Text to sequence..

Text to sequences => encodes from a word index

↳ Unseen words => OOV token
OOV => Out of Vocabulary

+ Padding: We pad the sentences so that they are a uniform size

+ Sarcasm:

Itr a labeled dataset. To Train a model To look for sarcasm.

1. Import json

2. Turn it into a df. (lists) for each point of interest

3. Working with the Tokenizer.
   - Fit Tokenizer after initializing

4. Much greater variety of words.