# 2019-01-31 Lab 3

**Contents**



## Exercise: Babel

- [ ] **Discuss** how you would build a simple Babel (Baby Babel) library in code.
  - A simple library is one with a small number of characters and small messages, e.g. 5 characters and messages of length 6.
- [ ] Observe example of how to generate Baby Babel as a permutation space (sample space).
- [ ] **Discuss** how you would create a Babel library with full vocabulary (Big Babel).
  - Remember – the permutation space is a function of the alphabet size and message length.
- [ ] Observe example of how Big Babel might be built.
- [ ] **Discuss** how you would improve the performance of Big Babel by adding a language model.
  - By performance, we mean the ability to generate English looking messages.
- [ ] Observe example of applying language model to Big Babel.

## Exercise: Entropy in Moby Dick

- [ ] How might we use entropy to detect stop-words in Moby Dick?
- [ ] Observe example of how we create a vocabulary table from the tokens table.
- [ ] Observe example of how we compute an estimate of the entropy of the words in the text.

## Homework

- [ ] Using either Spyder or Jupyter, write a script following the Pandas method to generate a Babel library permutation space with the following parameters:
  - Use the Simple Polynesian alphabet (as found in the Schütz and Manning reading.)
  - Use a message length of 7.
- [ ] Using the unigram term frequencies in that reading as weights, write a simple text generator from the Babel library you just created.
- [ ] Submit your source code to the appropriate Assignment in Collab.

## Files

- babel.ipynb
- moby3.ipynb