WK6 Exercise sheet

Lab Problem: Ham-Bot

Have you ever seen one of those bots that can generate genuine-sounding words?

Like, someone has written bots to generate songs that could have been written by Metallica, tweets that could have been written by President Trump, and research papers that could have been written by a scientist.

They're pretty fun. Today, we're going to do our own: to generate Hamilton lyrics.

The heart of this algorithm will need 2-grams. A 2-gram is all the contiguous phrases of two words each that occur throughout a test. For example, if the text is *I am not giving away my shot*, then the 2-grams are *I am, am not, not giving, giving away, away my, my shot*.

We're going to use 2-grams to do our best impersonation of Lin-Manuel Miranda.

Here's the idea:

- Read in the Hamilton lyric file. We've provided starter code that does a bit of cleanup, but make sure you also remove punctuation, as well as the lines that indicate who's talking: "[HAMILTON], [BURR], [WHOLE COMPANY]", etc. As far as we know, they are the only instances of square brackets, so you can look for those.
- Generate a list of words that begin each line.
- Generate a list of words that end each line.
- Generate a dictionary of 2-grams. The key of the dictionary is a word, and the value is a list of all the words that follow it. For example, one of the entries in this dictionary would be something like 'home':['lord', 'begins', 'sweet', 'lord']. Don't worry about repeating words in the value. We have lord here twice, but that's both OK and expected.

Once you've got all that, it's time to generate some lyrics! The starter code contains a function that randomly builds a Hamilton line using your start, ending, and 2-gram inputs. Try calling the function 5 or 6 or 10 times, and you'll have yourself a little song. It might be gibberish, or HEY it might be great!

Done Early? Keep going...

If you've finished the lab and still have time, try rewriting your program to work with 3-grams instead of 2-grams. (Or, generate your own input file; there are lots of musical lyrics out there!)