Clarisse Taboy
HW 6 Write Up, Methods II
April 17, 2021

1. I was stuck for a while on figuring out how to decompress the file
with gzip.GzipFile. I initially wrote "with
gzip.GzipFile("news.2007.en.shuffled.gz") as source:" and also tried
gzip.open(), but because I got a bunch of errors that mentioned the
nltk.tokenize part, I thought that I had a problem with how the file
was being opened. I tried a few more things out, but after no success,
I ultimately asked Kyle who then nudged me in the right direction.

2. Running my program for the first time took a very long time since I
didn't pick a sample of sentences from the file. I looked up how to do
so and found a strategy that involved importing itertools and
specifying the number of lines from 'source' to iterate over.

3. On the stretch goal: I was unclear about how subprocess worked even
with the python docs, so I fortunately found a helpful video by Corey
Schafer on YouTube (https://www.youtube.com/watch?v=2Fp1N6dof0Y) that
demonstrated it in action. I used f strings for the the subprocess
args, which seemed to work fine. Initially I tried using the list
format, but it didn't work for some reason. I had to specify
'shell=True' for each subprocess arg because Terminal would give me an
error saying something like 'Invalid path'. I've heard from the video
about 'shell=True' making your computer susceptible to security
issues, but I'm hoping that for this hw, nothing like that will
happen.