10/11-07

* Clean up code
* (add functions… e.g. to “prep” text)
* Write little summaries
* Upload to github
* BNC (look into)
* Change frequency stuff to see the comparison between her frequency of words and the corpus’

Miguel’s suggestion of looking into a word very frequent in one speech but not another and why

11/12-07:

* BNC
* Do same again, but then go over the output
* Get rid of names/non-choice words
* Try verbs
* Input into computer and see results
* Check 50s – 60s (and check what was written in my report about this… phonetics)
* Spike in commonwealth – why? Threat?

Notes:

* Frequency analysis:

Need to replicate, for a british english corpus (because of spelling issues)

different reasons for infrequency [think on this]

think on some way of determining “choice” words and non-choice

e.g. ninetieth

a lot of "zero-frequency" words are non-words, due to lemmatisation issues

OR not… e.g. ‘gb’ – go through data and replace gb with great british