15/07

* BNC – replicate frequency for BNC
* Write summaries 🡪 upload onto github (everything)
* CLEAN UP STRUCTURE OF FILES
* Get rid of names/non-choice words 🡪 manually re-input into comp?
* Parse sentences in text

🡪now look at making it into a tree

16/07

Run the BNC freq thing (will take a while)

🡪analyse results and how to pursue

Syntax and see if a measurement of “embedding” is possible

*I used number of verbs (see graph below)*

THINK ABOUT GETTING RID OF QUOTES?

Think about questions [search for question marks]

[look at 1971][look at 89]

Notes

* BNC corpus has taken a lot longer than expected 🡪 heavy amounts of data, and having to build bottom-up (brown corpus was all automated in nltk… had to download BNC and collate text, words, etc)
* Heavy computer demands – my computer couldn’t always keep up [solution?]
* Brown corpus: American English

Wordnet corpus:

(no explicit information as to where the information comes from).

According to [the book Speech and Language Processing](https://web.stanford.edu/~jurafsky/slp3/) from Daniel Jurafsky and James H. Martin, the sense frequencies come from the [SemCor](https://www.sketchengine.eu/semcor-annotated-corpus/) corpus which is a subset of the already small and outdated Brown Corpus.

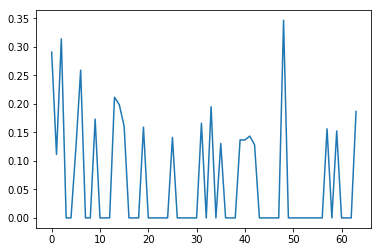
* NOTE FOR FINAL REPORT:

TWO-FOLD RESISTENCE TO CHANGE:

1. Conservative language of queen
2. Written text is more resistant to change, and this speech will be based on written text

Look at this for parsing sentences (not nltk… stnford ):

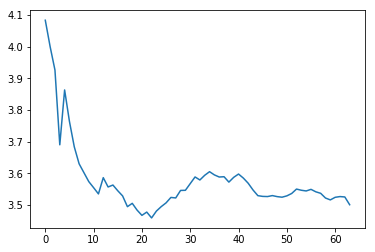
<https://stackoverflow.com/questions/42322902/how-to-get-parse-tree-using-python-nltk>



Under 10 frequency proportion from BNC…

Need to try a higher cut-off point

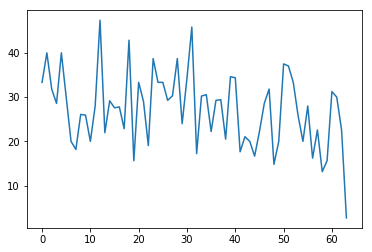
Other than the spike – general trend downwards! 50s-60s again?



Number of verbs per sentence average, across all speeches

(number of verbs is a ROUGH estimate for amount of embedding)

🡪embedding is clause within clause. Clause is subject and verb. Subject can be implicit



Proportion of sentences in speech which have more than 4 verbs

Slight trend downwards

These two graphs show us: generally lower average verb count per sentence... and the higher end is lowering too.