June Summary.

* Began by following the metrics outline in Savoy paper

Didn’t anticipate significant changes, as Savoy compared between people 🡪 likely to find greater changes. Changes between same speaker likely much more subtle (especially for a conservative speaker)

MSL; Lexical diversity; lexical richness; big words

Trend downwards of MSL (mean sentence length).

Past few decades see fewer big words than first few – big dip around 1985

(other than spike in year 1999)

Trend upwards for lexical diversity (wider range of words) and density (more lexical than functional). This hints at more descriptive language… which actually would encourage more big words. Therefore, in spite of more lexical words, word size still reduces

* POS\_distribution

Of interest:

Pronoun use has gone down

Recent (past decade) dip in conjunctions (this can reflect less complexity syntax 🡪 shorter sentences)

* Pronoun dip analysis:

First decade (50s) saw much higher use of personal posession pronouns ‘my’ ‘mine’

🡪’my people’

* Frequency relative to that of the general population

Brown corpus 🡪skewed by American English vs british English

(consider time of creation of corpus / years from which the text come)

Next steps:

* Starting to look into parsing of sentences
* Pronoun use / sense of posession
* Same topics / different expression over years (e.g. commonwealth/women)
* Looking into the “prediction” of text