08-07

* Look at adjective frequency
* Think about how to “clean” data to what matters

Perhaps about how to define “upper-class” language?

Because the infrequent words appearing aren’t infrequent in this way

THIS HIGHLIGHTS NEED FOR HUMAN EYE / INSPECTION FROM LINGUIST

Intuitively can tell that earlier words are more upper-class

Need to find a way of assessing this

* First step: remove punctuation and anything with a number in it

List of adjectives with least frequency

Remove those that have no option – for example “ninetieth”

* Element of choice ?
* Look back at BNC british language corpus

To avoid 0 frequency of british spellings