

Computerized Text Analysis: Classwork 3

Exploring and Coding Texts in MaxQDA

Kenneth Benoit

In today's lab we will explore some more descriptive statistics of the UK Manifesto texts that we loaded and annotated yesterday, and also load and annotate a new set of texts.

Instructions:

1. Start QDAMiner and open the project containing the UK Manifesto texts that we worked with in yesterday's lab. Check that the variables attached to each case (party and year) have been saved correctly.
2. Using WordStat (Analyse, Content Analysis), explore some of the basic concepts covered in today's lecture, for example:
 - (a) word frequency (frequency tab)
 - (b) lemmas and stemming (look at the difference between using the lemma dictionary and the Porter stemmer in the dictionary tab)
 - (c) Concordance (keyword-in-context tab)
 - (d) Stopwords (Exclusion list in Dictionary tab)
 - (e) Type/token ratio, mean statistics (Frequency tab, blue information button)
 - (f) Hapax (reverse sort frequency list)
 - (g) (Optional) - Zipf's Law states that in a corpus of natural language, the frequency of a word is inversely proportional to its rank in the sorted frequency list. Test this on the UK Manifesto texts by exporting the frequency list to a file, importing the file with Stata or Excel, and plotting the log of the rank against the log of the frequency. A simple way to import the data into Excel or OpenOffice is to right click on the table shown in the Frequency tab and select 'copy to clipboard', and then paste this directly into a new spreadsheet.
3. In the S drive, you will find Budget speeches from the Irish austerity budget debate in 2010. Create a new project in QDA Miner and annotate these speeches with appropriate variables.
or
4. If you have any texts related to your own research that you would like to analyse with WordStat, you can load them and annotate them now.