Computerized Text Analysis: Classwork 7
Dictionary Construction

Kenneth Benoit

In today's lab we will conduct a dictionary-based analysis of the UK manifesto and Irish Budget texts. For the UK manifestos, we will use a dictionary described in Laver and Garry (2000), and available through QDA Miner. For the Irish budget texts, we will define a custom dictionary.

Instructions

1. Open the UK Manifesto texts project from previous classes.

2. Open Wordstat, and select the Dictionary tab.

3. Select the Laver and Garry Dictionary from the Categorization Dictionary dropdown, and tick the check box.

4. Explore the distribution of words among dictionary categories in the Frequency tab. Experiment by altering the level of hierarchy used (an option on the dictionary panel). You can also tabulate the dictionary category frequencies with variables such as Party and Year using the Crosstab panel (you must select 'analyse in relation with other variables' or 'assigned codings' when you start WordStat to do this).

5. Open the Irish budget speeches project. If you haven't already done so, assign a variable to categorize each speech as either pro-budget (FF, Green) or anti-budget (Lab, FG, SF). A quick way to assign multiple variables is to use the spreadsheet view in QDA Miner - you can copy and paste variables to multiple cells, like in Excel (`Variables -> Spreadsheet Editor`).

6. Examine the budget speeches using KWIC and cross-tabulation by the pro/anti budget variable. Which words are most likely to indicate a pro or anti budget position? Discover dictionary words first by reading the text, and then confirm or alter your selections using the statistics available in the crosstab panel.

7. Create your own custom dictionary to group words as Pro-Budget or Anti-Budget:

   (a) Select 'new' from beside the Categorization dictionary option in WordStat.
   (b) Save the new dictionary file in a location where you have write permissions.
   (c) Select the file from the dropdown and choose 'add' categories to add the desired categories (pro budget and anti-budget)
   (d) Select 'add' words, and enter the words that you want to group under each category.

8. For more detailed instructions on creating dictionaries and lemmatization rules, choose the help option in WordStat and see the section under 'Dictionaries page'.