Quantitative Text Analysis: Assignment 4
Dictionary Construction

Kenneth Benoit

In today's lab we will conduct a dictionary-based analysis of the budget debates and of the amicus brief texts. For the budget debates, we will use the wordcat sentiment dictionary available through QDA Miner. For the amicus briefs, we will construct a custom dictionary, using the search and keyword in context (KWIC) tools as aids to dictionary construction. The purpose of this lab is to allow you to explore two new sets of documents using the tools that you learned last week. Examine the texts first in an exploratory fashion — read a few completely, and use the retrieval tool to get a feel for important terms. Consider what level of stemming or use of stopwords is appropriate.

Instructions

1. Open QDA Miner and your project containing the Irish budget debates. If you have not done so already, assign a variable for government/opposition. (Fianna Fail and Greens were in government in 2010.)

2. Open Wordstat, and select the Dictionary tab. Select the Wordcat sentiment dictionary.

3. 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 examine the KWIC tool in Wordstat. How well does the dictionary handle negation? Can you find cases where it fails?

4. Examine the budget speeches using KWIC and cross-tabulation by the pro/anti budget variable. Which words are most likely to indicate a Petitioner or Respondent position? Discover dictionary words first by reading the text, and then confirm or alter your selections using the statistics available in the crosstab panel — use the chi-square statistic to examine terms that differ significantly across variables.

5. Now open your project with the Amicus briefs. Make sure that you have a variable assigned to distinguish the petitioner from the respondent briefs.

6. Create your own custom dictionary to group words as Petitioner or Respondent. You can examine which words discriminate among the texts using the Cross-Tab and looking at the highest values (and seeing which side with which they are associated).

   (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 (petitioner and respondent).
   (d) Select 'add' words, and enter the words that you want to group under each category.

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

7. Apply the dictionary to the petitioner and respondent briefs. How well does it score each set of texts?