Computerized Text Analysis: Classwork 5
Wordscores

Kenneth Benoit

The objective of this class exercise is to better understand the Wordscores text scaling algorithm, using examples from Laver, Benoit and Garry (2003).

You will need some software to implement this example. Choices are Stata, using the Wordscores library (see http://www.politics.tcd.ie/wordscores/), or R using the Austin library (see http://www.williamlowe.net/software/#austin). Alternatively, you can program the algorithm yourself in a spreadsheet such as Microsoft Excel.

We will use two sets of files for this:

- The example from Table 1 of LBG (2003), available as LBGexample.csv. This file is in .csv (comma separated value) format and can be loaded directly into Stata, R, or your spreadsheet.

- The UK manifestos from Day 4.

Instructions:

1. Load the LBGexample.csv into your numerical analysis program. Inspect the "word" frequency matrix.

2. Compute the wordscores using the procedure from LBG (2003). The reference scores should be set at -1.50, -0.75, 0.00, 0.75, and 1.50 for reference texts r_1 through r_5 respectively.

3. Score the virgin text and compare your results to LBG (2003) Table 1.

4. Now score the words with just two reference texts, the extremes, using -1.50 and 1.50 respectively. Compare the scores of the words from the previous scoring. Score the Virgin text and compare to the earlier result.

5. Generate a word frequency matrix from the Labour, Lib Dem, and Conservative parties from 1992 and 1997, using whatever method you prefer.

6. Load the UK manifestos for the Labour, Lib Dem, and Conservative parties from 1992, and set their economic reference scores at 5.35, 8.21, and 17.21 respectively. Score the words, and then score the three corresponding 1997 manifesto texts. Compare the results to those from LBG (2003) Table 2.