Quantitative Text Analysis
Exercise 5: Dictionary Analysis

22nd July 2014, Essex Summer School

Kenneth Benoit and Paul Nulty

In today's lab we will use R to apply dictionary-based measures.

Instructions

1. For the UK Manifestos, we will use a dictionary from an article concerning populist language in party manifestos.

   (a) Open RStudio and load the `quanteda` library. The UK Manifestos corpus object is a built-in data object in `quanteda`. these commands will load it and make a subset from the texts we are interested in:

   ```
   data(UKManifestos)
   corpus <- subset(UKManifestos, (year %in% c(1992, 2001, 2005)
   & (party %in% c("Lab", "LD", "Con", "BNP", "UKIP"))))
   summary(corpus)
   ```

   (b) The dictionary argument to the dfm function in quanteda requires a list of terms to group under the key we are interested in — in this case a list of words that are considered to represent populist language. The command below will create a new list containing words used in the original article. (Rooduijn, M., & Pauwels, T. (2011). Measuring populism: Comparing two methods of content analysis. West European Politics, 34(6), 1272-1283.))

   ```
   dictionaryPopulismEN <-
     list(populism=c("elit*", "consensus*", "undemocratic*", "referend*",
                     "corrupt*", "propagand", "politici*", "*deceit*",
                     "*deceiv*", "*betray*", "shame*", "scandal*", "truth*",
                     "dishonest*", "establishm*", "ruling*"))
   ```

   (c) Look at the documentation for the `dfm` function, and use it to create a matrix of documents and counts of populist language. Use R to construct some of the measures discussed in Wednesday's class: the difference or ratio of the normalized populist and non-populist language.

2. A dictionary with multiple categories for political language was defined and applied in Laver. M. & Garry, J. (2000). Estimating Policy Positions from Political Texts. *American Journal of Political Science*, 44 (3) The Laver and Garry dictionary is available here.

   (a) This dictionary file is in a format designed to be used with the WordStat software, but there is a textttquanteda function to import this format — `readWStatDict`. Use it to load the dictionary and apply it the UK manifesto corpus.

   (b) Summarize the economic policy positions of the three main parties for 1992 and 1997, by constructing an index of left v. right. We are not picky about what format you choose, but try to combine them to construct some measure.