

# Unsupervised Methods for Scaling Texts: Lab Exercise

Kenneth Benoit

This exercise involves using the automatic document scaling using correspondence analysis in WordStat, using texts from Evans et al. (2007) *amicus curiae* briefs ([files here](#)) and from the 2010 Irish Budget debates ([files here](#)).

## Instructions

1. Reopen the project for the amicus briefs from the QTA 2 session.
2. Open these in Wordstat, and choose the Analyze tab and the Correspondence Analysis button.
3. Explore the results and test using different options.
4. Import the Irish budget speeches. Include the file [2010\\_BUDGET\\_11\\_John\\_Gormley\\_Green\\_ENTIRE.txt](#).
5. Use Wordstat to do a correspondence analysis on all texts, and inspect the two Gormley positions. Now remove one of the Gormley positions and reanalyze.
6. To analyze the results in R using the Poisson scaling model, try the following steps:
  - (a) Export the sudget corpus as a .csv object, call it “budget2010.csv”.
  - (b) Start an R session.
  - (c) Type the following:

```
# required to install quanteda from Github
install.packages("devtools")
library(devtools)
# needed by quanteda - Will Lowe's austin package
install.packages("austin", repos="http://r-forge.r-project.org",
                 type="source", dependencies=TRUE)
# install the latest version quanteda from Github
install_github("quanteda", username="kbenoit")
# read in the exported QDAMiner .csv budget project
wsc <- getWordStatCSV()
rownames(wsc$attribs) <- wsc$attribs$NAME
wsc$attribs$PARTY <- factor(wsc$attribs$PARTY,
                          labels=c("FF", "FG", "LAB", "SF", "Greens"))
# summary stats
summary(wsc)
# create the word frequency matrix
wfm <- create.fvm.corpus(wsc)
# run the wordfish model and plot the results
require(austin)
wfish <- wordfish(as.wfm(wfm), c(6,5))
summary(wfish)
plot(wfish)
```

## References

Evans, Michael, Wayne McIntosh, Jimmy Lin and Cynthia Cates. 2007. "Recounting the Courts? Applying Automated Content Analysis to Enhance Empirical Legal Research." *Journal of Empirical Legal Studies* 4(4, December):1007–1039.