All Women Shortlists Methodology

library(knitr)  
opts\_chunk$set(echo = FALSE, eval=FALSE, include=FALSE)

# Methodology

To account for the possible effects of age, parliamentary experience and cohort, and in order to compare women selected through all women shortlists to women who were not (but theoretically had the possibility to contest all-women shortlists), speech analysis has been restricted only to Labour MPs elected during or after the 1997 General Election, and before the 2017 General Election. Words contained in parentheses were removed, as they are added by Hansard to provide additional information not actually spoken by the MP.[[1]](#footnote-21) Speeches and MP data is from a dataset previously assembled by Odell ([2018](#ref-odell2018)). Information on all women shortlists is from House of Commons Library (Kelly [2016](#ref-kelly2016)). Unsuccessful General Election candidates selected through all women shortlists who were subsequently elected in a byelection are classified as having been selected on an all women shortlist.

Word classification used the Linguistic Inquiry and Word Count 2015 (LIWC) dictionary (Pennebaker et al. [2015](#ref-pennebaker2015)) and tokenising tools from the Quanteda R package (Benoit [2018](#ref-benoit2018)). Word counts and words-per-sentence were calculated using stringi, a wrapper to the ICU regex library. [NEED CITATIONS!!!]

Previous research has found gender differences in some LIWC categories and in parts of speech usage in members of the US House of Representatives (Yu [2014](#ref-yu2014)). Following YU ([2014](#ref-yu2014)), we used the following categories

* All Pronouns (pronoun)
* First person singular pronouns (i)
* Verbs (verb)
* Auxiliary verbs (auxverb)
* Social processes (social)
* Positive emotions (posemo)
* Negative emotions (negemo)
* Tentative words (tentat)
* Words longer than six letters (Sixltr)
* First person plural pronouns (we)
* Articles (article)
* Prepositions (preps)
* Anger words (anger)
* Swear words (swear)
* Cognitive processes (cogproc)

We also included words-per-sentence (WPS), total word count (WC) and Flesch–Kincaid grade level (FK) (Kincaid et al. [1975](#ref-kincaid1975)), calculated using Quanteda (Benoit [2018](#ref-benoit2018))[, and some measure of LEXICAL DIVERSITY???]

!CHECK UP ON NAZ SHAH, MAKE SURE SHE ALWAYS COUNTS AS LABOUR

## Corpus creation

# Descriptive Statistics

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| General Election | Total MPs | Total Labour MPs | Total Female Labour MPs | Newly elected MPs | Intake Women | Percentage Intake Women | Intake Shortlist | Nominated Shortlist |  |
| 1997 | 659 | 418 | 101 | 177 | 64 | 36% | 35 | 38 |  |
| 2001 | 659 | 412 | 95 | 38 | 4 | 11% | 0 | 0 |  |
| 2005 | 646 | 355 | 98 | 40 | 26 | 65% | 23 | 30 |  |
| 2010 | 650 | 258 | 81 | 64 | 32 | 50% | 28 | 63 |  |
| 2015 | 650 | 232 | 99 | 49 | 31 | 63% | 31 | 77 |  |

Data in this table is take from (Kelly [2016](#ref-kelly2016)) and (Audickas, Hawkins, and Cracknell [2017](#ref-audickas2017)). All women shortlists were not used by Labour during the 2001 General Election.

Hypothesis: - Women selected through all women shortlists use more “feminine” language – pronouns, particular personal pronouns, passive voice, hedges, etc – than women selected in open selections.

## LIWC creation

## Women vs Men

## Shortlists vs Non-Shortlists

Effect size was calculated using Cohen’s *d* (Cohen [1988](#ref-cohen1988))

### Experiment in plotting

## Spacy

## POS Analysis

STILL OUTSTANDING: - topic choice

## Tokenising / Keyness

The most commonly used words by both men and women would be protocol boilerplate expressions. With that in mind, we calculate the keyness of words.

### Men vs Women

### Shortlists vs Non-Shortlists

Audickas, Lukas, Oliver Hawkins, and Richard Cracknell. 2017. “UK Election Statistics: 1918-2017.” Briefing Paper CBP7529. London: House of Commons Library. <http://researchbriefings.parliament.uk/ResearchBriefing/Summary/CBP-7529>.

Benoit, Kenneth. 2018. *Quanteda: Quantitative Analysis of Textual Data*. <https://doi.org/10.5281/zenodo.1004683>.

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, N.J: L. Erlbaum Associates.

Kelly, Richard. 2016. “All-Women Shortlists.” Briefing Paper 5057. London: House of Commons Library. <https://researchbriefings.parliament.uk/ResearchBriefing/Summary/SN05057>.

Kincaid, J. Peter, Robert P. Fishburne, Richard L. Rogers, and Brad S. Chissom. 1975. “Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel:” Fort Belvoir, VA: Defense Technical Information Center. <https://doi.org/10.21236/ADA006655>.

Odell, Evan. 2018. “Hansard Speeches and Sentiment V2.5.1 [Dataset],” July. <https://doi.org/10.5281/zenodo.1306964>.

Pennebaker, James W, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. “The Development and Psychometric Properties of LIWC2015,” 26. <https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf>.

Yu, B. 2014. “Language and Gender in Congressional Speech.” *Literary and Linguistic Computing* 29 (1): 118–32. <https://doi.org/10.1093/llc/fqs073>.

1. e.g. a reference to “the member for Bethnal Green and Bow” in keeping with Parliamentary convention of identifying MPs by their seat rather than their name would be followed by “(Rushnara Ali)”. [↑](#footnote-ref-21)