All Women Short lists Methodology

Contents

TO DO:

2. Male MP topic models matching

Methodology

Previous research on gender differences in political speech patterns has focused on differences between male and female politicians (Yu, 2014) or on variations in Hilary Clinton's speech patterns (Bligh, Merolla, Schroedel, & Gonzalez, 2010; Jones, 2016). This paper focuses on differences in speech patterns between female Labour MPs nominated through All Women Shortlists (AWS) and female Labour MPs nominated through open short lists. We examined differences in speaking styles using the `Linguistic Inquiry and Word Count 2015` (LIWC) dictionary (Pennebaker, Boyd, Jordan, & Blackburn, 2015) and the `spaCy` (Honnibal & Montani, 2017) Parts-of-Speech (POS) tagger. We examined differences in the topics discussed by AWS and non-AWS MPs, using χ^2 tests for individual words and for bigrams. We experimented with unsupervised Latent Dirichlet Allocation (Blei, Ng, & Jordan, 2003), and trained a Naive Bayes classifier to distinguish AWS and non-AWS speeches.

To account for the possible effects of age, parliamentary experience and cohort, and in order to compare women selected through all women short lists to women who were not (but who theoretically had the opportunity to contest all-women short lists), our analysis is been restricted only to Labour MPs first elected to the House of Commons in the 1997 General Election, up to but excluding the 2017 General Election. Comparisons between MPs of different parties are also restricted to MPs first elected in the 1997 General Election, and before the 2017 General Election. Speeches made by the Speaker, including Deputy Speakers, were also excluded. Words contained in parentheses were removed, as they are added by Hansard to provide additional information not actually spoken by the MP.[1] Speeches and data on MPs' gender and party affiliation are from a previously assembled dataset (Odell, 2018). Information on candidates selected through all women short lists is from the House of Commons Library (Kelly, 2016). Unsuccessful General Election candidates

[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)".

Table 1: Labour MPs and Intakes

General Election	Total MPs	Labour MPs	Female Labour MPs	Labour MPs Intake	Intake Women	Intake Short list	Nominated Short list
1997	659	418	101 (24%)	177	64 (36%)	35	38
2001	659	412	95 (23%)	38	4 (11%)	0	0
2005	646	355	98 (28%)	40	26 (65%)	23	30
2010	650	258	81 (31%)	64	32 (50%)	28	63
2015	650	232	99 (43%)	49	31 (63%)	31	77

selected through all women short lists who were subsequently elected in a byelection are classified as having been selected on an all women short list.

Word classification used the `Linguistic Inquiry and Word Count 2015` (LIWC) dictionary (Pennebaker et al., 2015) and tokenising tools from the `Quanteda` R package (Benoit, 2018). Word counts and words-per-sentence were calculated using `stringi` (Gagolewski, 2018), a wrapper to the ICU regex library.

Following Yu (2014) drawing on (Newman, Groom, Handelman, & Pennebaker, 2008) we used the following LIWC categories:

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

We also included mean words-per-sentence (WPS), total speach word count (WC) and Flesch–Kincaid grade level (FK) (Kincaid, Fishburne, Rogers, & Chissom, 1975), calculated using `Quanteda` (Benoit, 2018) and `stringi` (Gagolewski, 2018).

Descriptive Statistics

Data in this table is from House of Commons library reports (Audickas, Hawkins, & Cracknell, 2017; Kelly, 2016). All women short lists were not used by Labour during the 2001 General Election.

Women vs Men

There are no categories where gender differences meet the effect size threshold of $|0.2|$ suggested by Cohen (1988, pp. 25–26) to indicate a small effect. 4 categories – words with more than six letters, prepositions, words-per-sentence and Flesh-Kincaid grade level – exceeded the $|0.1|$ threshold suggested by Newman et al (2008).

Table 2: Number of Speeches and Words in Dataset

Gender	Speeches	Words
All	656412	111180398
Female	148702	26231034
Male	507710	84949364
Conservatives		
All	285291	44800169
Female	48768	7363031
Male	236523	37437138
Labour		
All	261942	46494850
Female	84569	15897929
Non-All Women Shortlists	28695	5422776
All Women Shortlists	55874	10475153
Male	177373	30596921
Liberal Democrat		
All	72716	13485902
Female	7552	1503459
Male	65164	11982443
Other		
All	36463	6399477
Female	7813	1466615
Male	28650	4932862

Table 3: Effect Sizes for Male and Female Labour MPs

	Women		Men		Effect Size	
	Mean	SD	Mean	SD	Cohen's D	Magnitude
All Pronouns	10.07	4.60	10.15	4.99	0.02	negligible
First person singular pronouns	1.89	2.41	2.02	2.55	0.05	negligible
First person plural pronouns	0.97	1.42	0.99	1.51	0.01	negligible
Verbs	12.82	5.00	12.67	5.36	-0.03	negligible
Auxiliary verbs	7.91	3.45	7.93	3.69	0.01	negligible
Social processes	8.47	4.82	8.18	5.11	-0.06	negligible
Positive emotions	2.73	2.49	2.57	2.54	-0.06	negligible
Negative emotions	1.15	1.68	1.07	1.77	-0.05	negligible
Tentative words	1.48	1.74	1.58	1.90	0.05	negligible
More than six letters	10.58	3.68	10.22	3.92	-0.10	negligible
Articles	7.65	3.30	7.96	3.55	0.09	negligible
Prepositions	12.58	4.42	12.14	4.73	-0.10	negligible
Anger words	0.23	0.81	0.24	0.77	0.01	negligible
Swear words	0.00	0.06	0.00	0.09	0.01	negligible
Cognitive processes	8.68	4.83	8.82	5.15	0.03	negligible
Words per Sentence	43.63	19.68	41.15	20.04	-0.13	negligible
Total Word Count	402.79	691.27	370.18	647.36	-0.05	negligible
Flesh-Kincaid Grade Level	10.81	7.68	9.78	7.87	-0.13	negligible

Table 4: Effect Sizes for Female Labour MPs by selection process

	All Women Short lists		Open Shorlists		Effect Size	
	Mean	SD	Mean	SD	Cohen's D	Magnitude
All Pronouns	10.01	4.67	10.19	4.48	-0.04	negligible
First person singular pronouns	1.86	2.41	1.95	2.42	-0.04	negligible
First person plural pronouns	0.88	1.36	1.16	1.51	-0.19	negligible
Verbs	12.88	5.10	12.69	4.80	0.04	negligible
Auxiliary verbs	7.94	3.49	7.86	3.38	0.02	negligible
Social processes	8.48	4.94	8.46	4.59	0.00	negligible
Positive emotions	2.69	2.52	2.81	2.42	-0.05	negligible
Negative emotions	1.16	1.69	1.13	1.67	0.02	negligible
Tentative words	1.48	1.75	1.49	1.73	0.00	negligible
More than six letters	10.52	3.73	10.70	3.58	-0.05	negligible
Articles	7.69	3.38	7.55	3.15	0.04	negligible
Prepositions	12.55	4.54	12.63	4.15	-0.02	negligible
Anger words	0.23	0.78	0.24	0.88	-0.01	negligible
Swear words	0.00	0.06	0.00	0.05	0.01	negligible
Cognitive processes	8.59	4.90	8.86	4.69	-0.06	negligible
Words per Sentence	44.02	20.45	42.85	18.05	0.06	negligible
Total Word Count	401.77	704.40	404.78	664.97	0.00	negligible
Flesh-Kincaid Grade Level	10.97	7.97	10.48	7.06	0.07	negligible

Short lists vs Non-Short lists

The following plots show changes in the occurences of selected LIWC terms, words-per-sentence, total word count and Flesch–Kincaid grade level, over the course of an MP's career. There do not appear to be any notable changes in speaking style over the course of female Labour MPs' careers.

There are no categories among female Labour MPs by selection process meeting the $|0.2|$ threshold. Only one category – first person plural pronouns, d=0.19 – exceeded $|0.1|$.

Conservatives vs Labour

There are no categories with effect sizes exceeding $|0.2|$ between Labour and Conservative MPs, like inter-Labour differences.

All MPs Gender Differences

POS Analysis

Part-of-speech (POS) tagging was done using `spaCy` (Honnibal & Montani, 2017) and the `spacyr` package (Benoit & Matsuo, 2018).

Tokenising / Keyness

The most commonly used words by both men and women would be protocol decorum expressions, so we calculate the keyness of words to identify gender differences in the choices of topics raised by men and women, and by short-list and non-short list women.

Table 5: Effect Sizes for All Labour and Conservative MPs

	Labour		Conservatives		Effect Size	
	Mean	SD	Mean	SD	Cohen's D	Magnitude
All Pronouns	10.12	4.87	10.62	4.84	0.10	negligible
First person singular pronouns	1.98	2.51	2.15	2.56	0.07	negligible
First person plural pronouns	0.98	1.48	1.22	1.70	0.15	negligible
Verbs	12.72	5.24	12.93	5.13	0.04	negligible
Auxiliary verbs	7.92	3.61	8.17	3.58	0.07	negligible
Social processes	8.28	5.02	8.13	4.80	-0.03	negligible
Positive emotions	2.62	2.53	2.85	2.66	0.09	negligible
Negative emotions	1.10	1.74	1.05	1.78	-0.03	negligible
Tentative words	1.55	1.85	1.57	1.88	0.01	negligible
More than six letters	10.34	3.85	10.28	3.76	-0.02	negligible
Articles	7.86	3.48	7.82	3.45	-0.01	negligible
Prepositions	12.28	4.64	12.38	4.49	0.02	negligible
Anger words	0.24	0.78	0.24	0.82	0.01	negligible
Swear words	0.00	0.08	0.00	0.10	0.00	negligible
Cognitive processes	8.77	5.05	8.86	5.06	0.02	negligible
Words per Sentence	41.95	19.96	42.76	20.16	0.04	negligible
Total Word Count	380.71	662.03	335.54	592.41	-0.07	negligible
Flesh-Kincaid Grade Level	10.12	7.82	10.41	7.91	0.04	negligible

Table 6: Effect Sizes for Male and Female MPs, All Parties

	Women		Men		Effect Size	
	Mean	SD	Mean	SD	Cohen's D	Magnitude
All Pronouns	10.31	4.65	10.26	4.90	-0.01	negligible
First person singular pronouns	1.99	2.45	2.00	2.52	0.00	negligible
First person plural pronouns	1.11	1.57	1.08	1.59	-0.02	negligible
Verbs	12.88	4.97	12.80	5.26	-0.02	negligible
Auxiliary verbs	8.00	3.45	8.08	3.64	0.02	negligible
Social processes	8.45	4.77	8.00	4.93	-0.09	negligible
Positive emotions	2.84	2.53	2.69	2.58	-0.06	negligible
Negative emotions	1.10	1.65	1.08	1.78	-0.01	negligible
Tentative words	1.47	1.73	1.61	1.91	0.08	negligible
More than six letters	19.73	6.94	19.25	7.18	-0.07	negligible
Articles	7.62	3.31	8.00	3.51	0.11	negligible
Prepositions	12.58	4.36	12.22	4.62	-0.08	negligible
Anger words	0.23	0.78	0.25	0.82	0.02	negligible
Swear words	0.00	0.05	0.00	0.10	0.01	negligible
Cognitive processes	8.67	4.79	8.93	5.12	0.05	negligible
Words per Sentence	43.25	19.45	42.06	20.12	-0.06	negligible
Total Word Count	377.31	648.92	358.13	623.49	-0.03	negligible
Flesh-Kincaid Grade Level	10.63	7.61	10.16	7.89	-0.06	negligible

Table 7: Part-of-Speech Effect Sizes for Male and Female Labour MPs

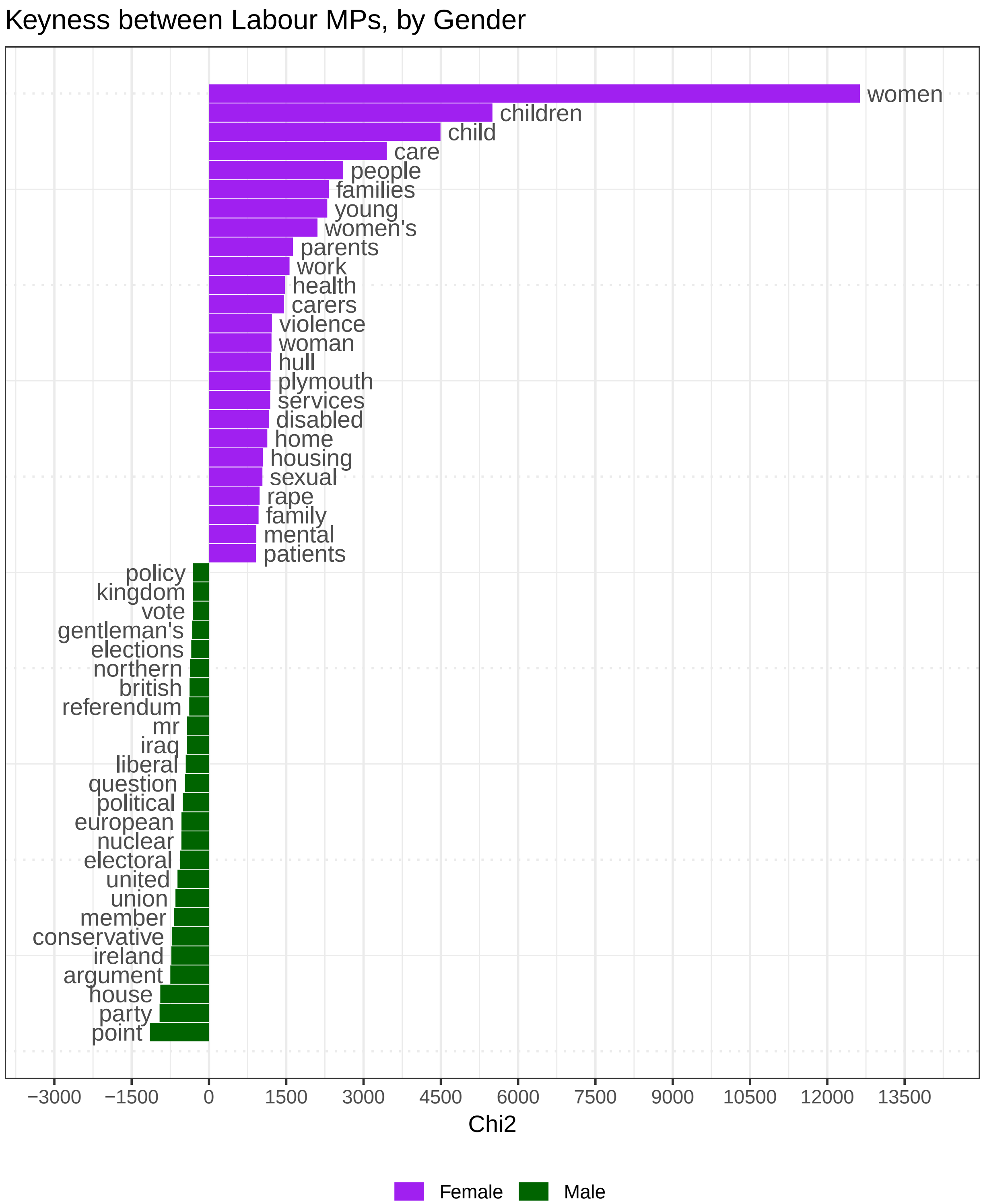
	Women		Men		Effect Size	
Word Type	Mean	SD	Mean	SD	Cohen's D	Magnitude
All Nouns	22.18	9.60	21.66	10.96	-0.04	negligible
Plural Nouns	5.85	3.72	5.03	3.79	-0.16	negligible
Singular Nouns	15.62	9.84	16.01	11.19	0.02	negligible
Adjectives	9.58	4.78	9.28	5.29	-0.02	negligible
Adverbs	4.91	4.26	5.07	4.91	0.03	negligible
Verbs	20.94	9.52	20.78	10.28	-0.02	negligible

Table 8: Part-of-Speech Effect Sizes for AWS and non-AWS Labour MPs

	All Women Short lists		Open Shorlists		Effect Size	
Word Type	Mean	SD	Mean	SD	Cohen's D	Magnitude
All Nouns	22.16	8.78	22.18	10.00	-0.04	negligible
Plural Nouns	6.03	3.60	5.76	3.77	-0.16	negligible
Singular Nouns	15.51	8.97	15.67	10.26	0.02	negligible
Adjectives	9.83	4.59	9.45	4.86	-0.02	negligible
Adverbs	4.95	3.78	4.89	4.49	0.03	negligible
Verbs	20.88	9.04	20.97	9.76	-0.02	negligible

Men vs Women

Keyness – a linguistic measure of the frequency of different words in two groups of texts – reveals clear gender differences in the most disproportionately common words used by female and male Labour MPs. Unsurprisingly, despite male MPs saying almost twice as many words (30601887 vs 15898845) as their female colleagues, female Labour MPs were more than two-and-a-half (2.61) times as likely to say "women". They were also much more likely to refer to "women's" and "woman". Female Labour MPs also appear much more likely to discuss "children", "people", "care", "families", "home", "parents", "work" and social policy areas such as "services", "disabled [people]" and "housing" than their male colleagues. Male MPs were more likely to refer to military topics ("Iraq", "nuclear"), and to parliamentary process and protocol – "question", "political", "conservative", "electoral", "house", "party", "argument" "liberal" and "point" are far more common in speeches by male Labour MPs than by female ones. This could suggest that male MPs are more comfortable using the traditional language of House of Commons debate.

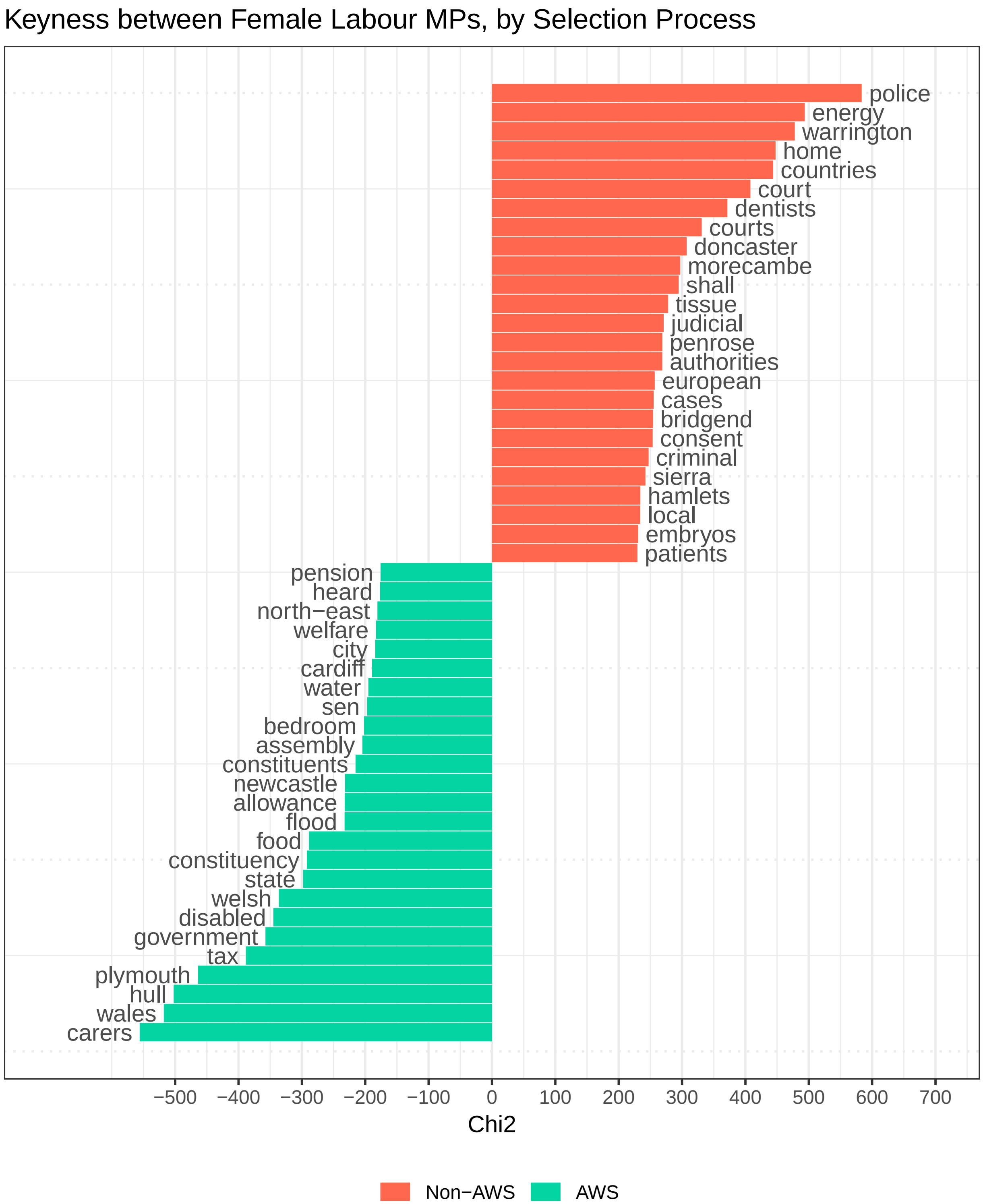


Short lists vs Non-Short lists

Keyness differences by selection process are not as obviously stereotypical. Nonetheless, the most common words amongst AWS MPs included “carers”, “disabled”, “bedroom” and “sen”[2]. Also of note is AWS MPs

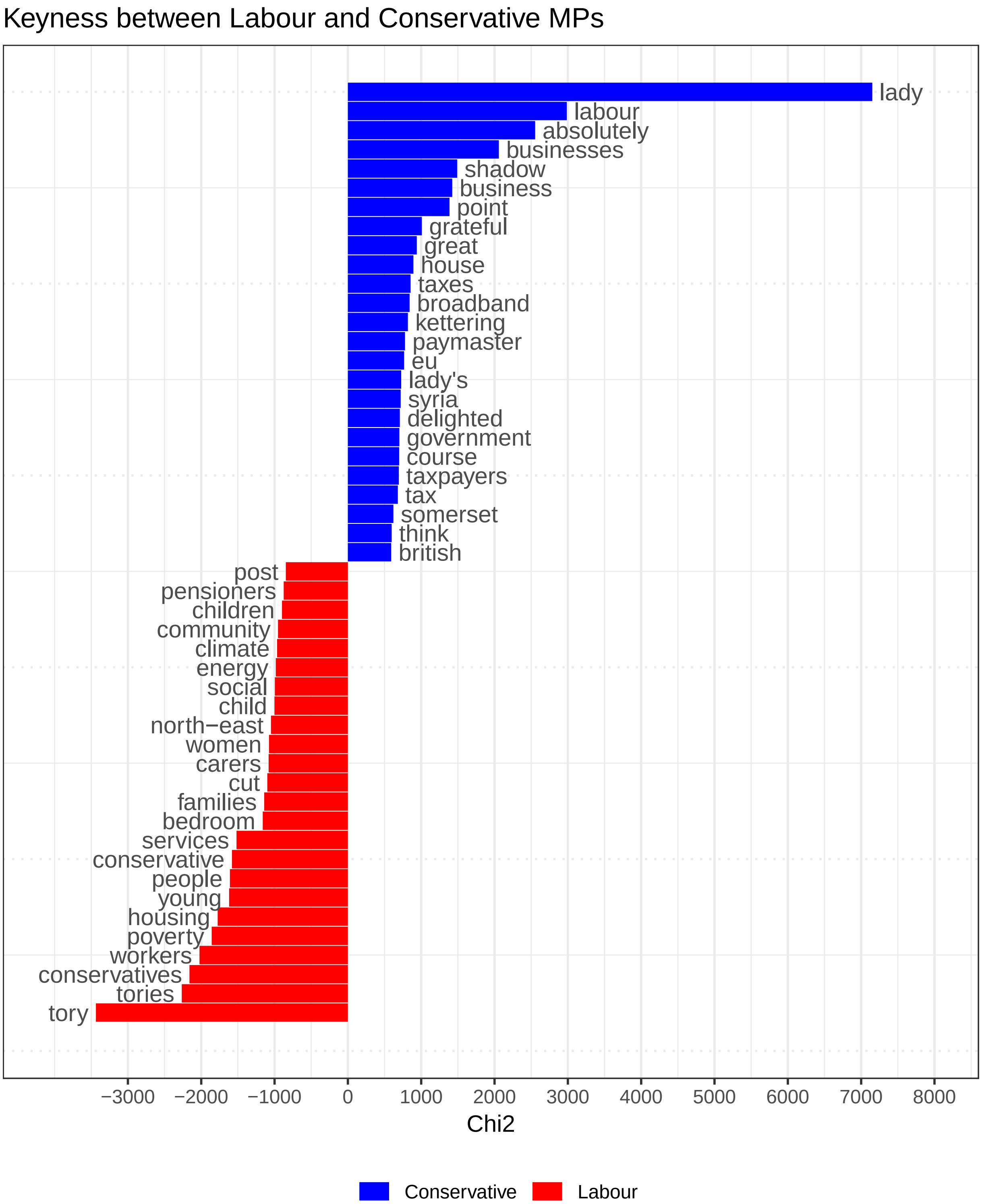
[2]Special Educational Needs

making more references to their "constituency" and its "constituents", [drawing on representation of]



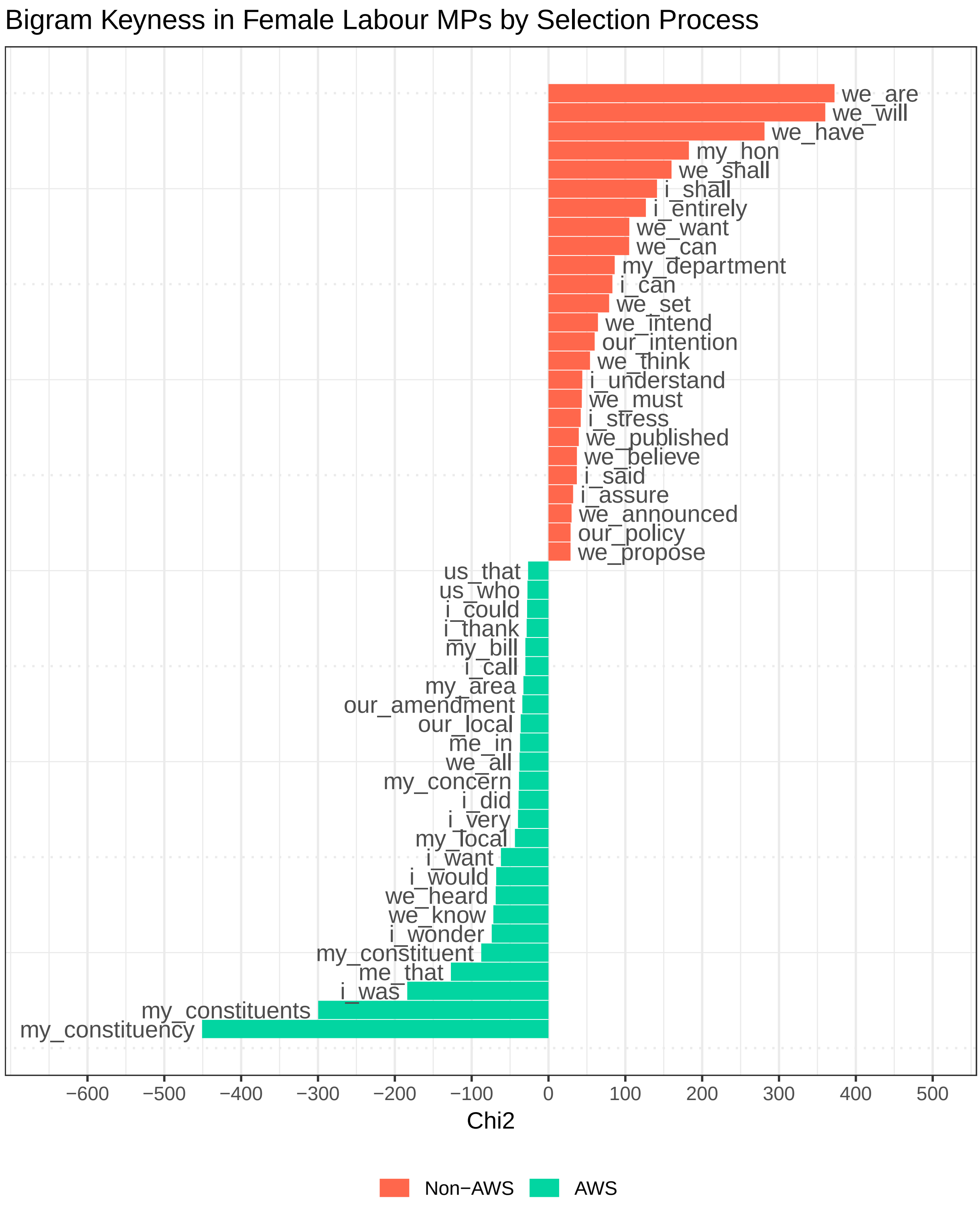
Labour vs Conservative

The keyness differences between Labour and Conservative MPs are much greater than gender differences within Labour. The very high use of "Lady" by Conservative MPs is reflective of the greater proportion of female MPs in other parties, as it is often used to refer to comments by other members of the house.



Bigrams

We created bigrams of all first person plural and singular pronouns for female Labour MPs.



Topic Models

Using topic models to classify text is widely used in social sciences, as, when combined with the large volume of plain text data available, it allows for a rapid and consistent method of analysis. Topic modelling and other statistic methods of textual analysis are not a substitute for reading the texts themselves, but can augment other analysis or – as in this case – analyse and classify larger amounts of text than would be feasible using human coders (Grimmer & Stewart, 2013). Topic models classify a series of documents (in this case individual speeches) into one of a given number of topics, identifying terms that are common in some documents but rare in others. When developing topic models, there is a trade-off between high precision in the classification of each document with broader topics when using smaller numbers of topics, or lower precision in individual speech classification with more finely-grained topics when using larger numbers of topics. Grimmer and Stewart also highlight the importance of validating unsurpervised topic models when applied to new sets of texts.

The R package `stm` (Roberts, Stewart, & Tingley, 2018) implements a structured topic model (Arora et al., 2013; Roberts, Stewart, & Airoldi, 2016)

Short lists vs Non-Short lists

```r
library(readr)
library(stm)
```

```
## stm v1.3.3 (2018-1-26) successfully loaded. See ?stm for help.
##  Papers, resources, and other materials at structuraltopicmodel.com
```

```r
topic_model2 <- read_rds("stm-topic-model2.rds")

#x <- labelTopics(topic_model2)
```

Naive Bayes classification

We trained a Naive Bayes classifier with document-frequency priors and a multinomial distribution to predict the gender of speakers when given speeches by all Labour MPs in our dataset, and the selection process when only given female Labour MPs. The accuracy of both models were roughly equivalent, 70.55% accuracy when predicting gender and 70.84% when predicting short lists. By contrast, the classifier could distinguish between Labour and Conservative speeches with 74.23% accuracy.

Discussion

There do not appear to be substantial or meaningful differences in the speaking styles or topic choices of female Labour MPs selected through all women short lists when compared to their female colleagues selected through open short lists. The few small differences between male and female Labour MPs were not replicated when comparing female Labour MPs by how they were selected.

There is more gender distinction in selected terms and topics.

References

Arora, S., Ge, R., Halpern, Y., Mimno, D., Moitra, A., Sontag, D., … Zhu, M. (2013). A Practical Algorithm for Topic Modeling with Provable Guarantees. In S. Dasgupta & D. McAllester (Eds.), *Proceedings of the 30th International Conference on Machine Learning* (Vol. 28, pp. 280–288). Atlanta, Georgia, USA: PMLR. Retrieved from http://proceedings.mlr.press/v28/arora13.pdf

Audickas, L., Hawkins, O., & Cracknell, R. (2017). *UK Election Statistics: 1918-2017* (Briefing Paper No. CBP7529) (p. 89). London: House of Commons Library. Retrieved from http://researchbriefings.parliament.uk/ResearchBriefing/Summary/CBP-7529

Benoit, K. (2018). *Quanteda: Quantitative Analysis of Textual Data.* https://doi.org/10.5281/zenodo.1004683

Benoit, K., & Matsuo, A. (2018). *Spacyr: Wrapper to the 'spaCy' 'NLP' Library.* Retrieved from http://github.com/quanteda/spacyr

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022.

Bligh, M., Merolla, J., Schroedel, J. R., & Gonzalez, R. (2010). Finding Her Voice: Hillary Clinton's Rhetoric in the 2008 Presidential Campaign. *Women's Studies*, *39*(8), 823–850. https://doi.org/10.1080/00497878.2010.513316

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Hillsdale, N.J: L. Erlbaum Associates.

Gagolewski, M. (2018). R package stringi: Character string processing facilities. https://doi.org/10.5281/zenodo.1292492

Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, *21*(03), 267–297. https://doi.org/10.1093/pan/mps028

Honnibal, M., & Montani, I. (2017). spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. *To Appear.* Retrieved from https://spacy.io

Jones, J. J. (2016). Talk "Like a Man": The Linguistic Styles of Hillary Clinton, 1992-2013. *Perspectives on Politics*, *14*(03), 625–642. https://doi.org/10.1017/S1537592716001092

Kelly, R. (2016). *All-women shortlists* (Briefing Paper No. 5057) (p. 34). London: House of Commons Library. Retrieved from https://researchbriefings.parliament.uk/ResearchBriefing/Summary/SN05057

Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (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

Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender Differences in Language Use: An Analysis of 14,000 Text Samples. *Discourse Processes*, *45*(3), 211–236. https://doi.org/10.1080/01638530802073712

Odell, E. (2018). Hansard Speeches and Sentiment V2.5.1 [dataset]. https://doi.org/10.5281/zenodo.1306964

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). The Development and Psychometric Properties of LIWC2015, 26. Retrieved from https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf

Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A Model of Text for Experimentation in the Social Sciences. *Journal of the American Statistical Association*, *111*(515), 988–1003. https://doi.org/10.1080/01621459.2016.1141684

Roberts, M. E., Stewart, B. M., & Tingley, D. (2018). *Stm: R Package for Structural Topic Models.* Retrieved from http://www.structuraltopicmodel.com

Yu, B. (2014). Language and gender in Congressional speech. *Literary and Linguistic Computing*, *29*(1), 118–132. https://doi.org/10.1093/llc/fqs073