**Explanatory File**

Jones, Jennifer J. 2016. “‘Talk like a man.’ The Linguistic Styles of Hillary Clinton, 1992–2013.” *Perspectives on Politics* 14(3).

1. Data Appendix
	1. Summary statistics for the Hillary Clinton transcript corpus
	2. Procedures for preparing and processing the transcripts for analysis
	3. Transcript source citations
	4. Additional notes on transcripts uploaded to the Cambridge Digital Archive
2. HRC corpus metadata
	1. Identifying information for all transcripts used in the analysis
3. HRC LIWC output
	1. Numeric data generated by Linguistic Inquiry and Word Count (LIWC). containing all LIWC linguistic categories for each individual transcript.
	2. Values (except for word count) are given as a percent of the total words per transcript.
4. HRC analysis
	1. R code to generate analysis, models, and graphs.
	2. Requires HRC LIWC output
5. Original transcripts
	1. Pre-processed, fully intact interviews and debates—with transcript metadata, interviewer/moderator questions, etc. See Data Appendix for processing procedures.
	2. Transcripts sourced from the Clinton Library are formatted as OCR-optimized PDFs. See Data Appendix for more information on OCR processing.
	3. Transcripts sourced from the U.S. State Department are organized by the year the interview took place and formatted as a single tab-separated value (.tsv) file per year.
	4. Transcripts sourced from Lexis Nexis, Factiva, or EBSCOhost are formatted as text files. Note that in a few cases (post 2000) multiple interviews are contained on a single text file. These are labeled accordingly.