LITTLE WORDS CAN
MAKE A BIG DIFFERENCE

SYRACUSE UNIVERSITY
School of Information Studies

LITTLE WORDS CAN MAKE A BIG DIFFERENCE

Function words are useful for certain text mining tasks:

Genre classification
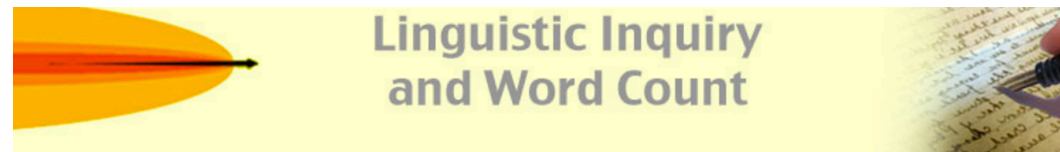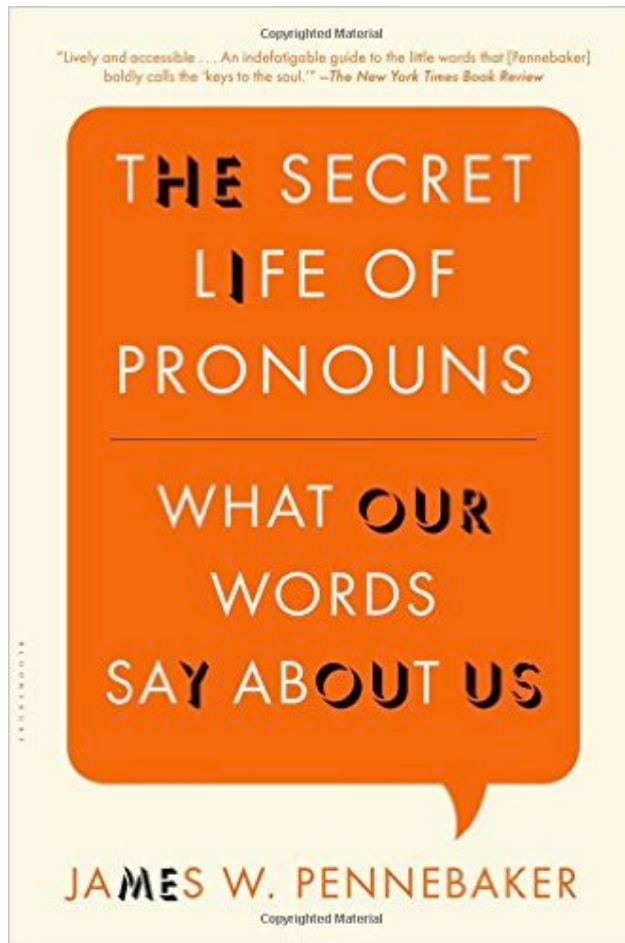
Authorship attribution

Gender detection

GENRE CLASSIFICATION

Personal home page identification (Riloff, 1995)

Top features "I" and "my"

Riloff, E. (1995, July). Little words can make a big difference for text classification. In *Proceedings of the 18th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 130–36). Association for Computing Machinery.

SYRACUSE UNIVERSITY
School of Information Studies

PERSONAL PRONOUNS

Linguistic Inquiry and Word Count

Testing LIWC Online

We understand completely. You are a student, a poor faculty member, or a researcher who wants to analyze a few cases without having to buy the LIWC program for almost $100. We've been there, and, because we know your plight, this page is for you. This is a no-frills page whereby you can enter text (by typing it directly or copying it from some other place and pasting it here) and get the basic LIWC output. All you have to do is enter the text file you want to analyze, press SUBMIT, and voila, we will give you feedback on some of the LIWC dimensions. That's the kind of people we are.
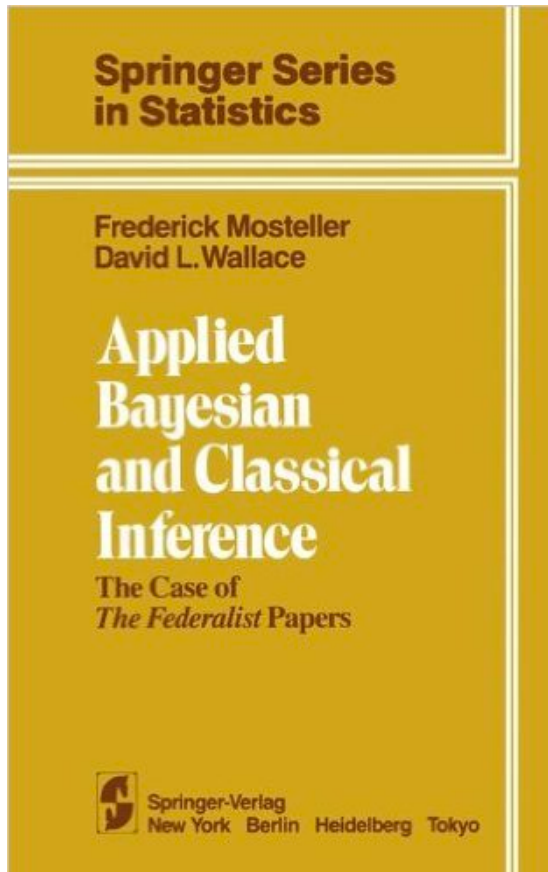
OK, we admit it. We aren't completely altruistic. We would like to keep a copy of your text files to add to our growing archive of 50,000+ files. To help us with our data, could you enter the age and gender of the author of the text (if you know it). If you don't know them or don't want to enter them, then choose 'No details' from the 'Gender of text author' selector.

Gender of author: [?] *Age of author:* [?]

Type or paste the text you want analysed into the box below and then hit the submit button.

[Submit text for online LIWC analysis]

FUNCTION WORDS FOR AUTHORSHIP ATTRIBUTION

Springer Series in Statistics

Frederick Mosteller
David L. Wallace

Applied Bayesian and Classical Inference

The Case of *The Federalist* Papers
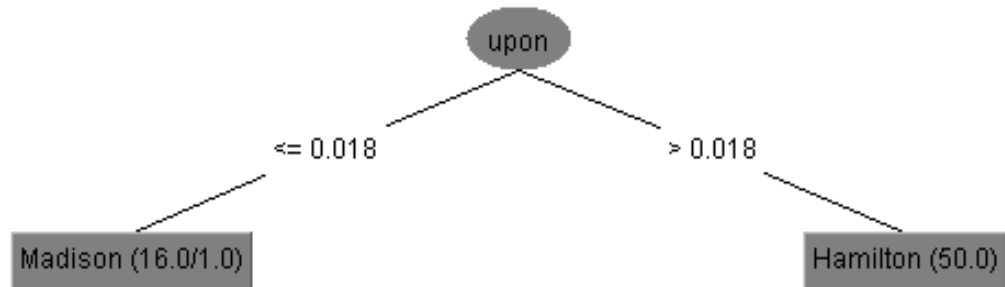
Springer-Verlag
New York Berlin Heidelberg Tokyo

The Federalist Papers

⚙ Too

This web-friendly presentation of the original text of the Federalist Papers (also known as The Federalist) was obtained from the e-text archives of Project Gutenberg. For more information, see About the Federalist Papers.

No.	Title	Author	Publication	Date
1	General Introduction	Hamilton	For the Independent Journal	- -
2	Concerning Dangers from Foreign Force and Influence	Jay	For the Independent Journal	- -
3	The Same Subject Continued: Concerning Dangers from Foreign Force and Influence	Jay	For the Independent Journal	- -
4	The Same Subject Continued: Concerning Dangers from Foreign Force and Influence	Jay	For the Independent Journal	- -
5	The Same Subject Continued: Concerning Dangers from Foreign Force and Influence	Jay	For the Independent Journal	- -
6	Concerning Dangers from Dissensions Between the States	Hamilton	For the Independent Journal	- -

upon

<= 0.018   > 0.018

Madison (16.0/1.0)   Hamilton (50.0)

GENDER CLASSIFICATION IN GENERAL TEXTS

TABLE 1 (*Continued*)

LIWC Dimension	Examples	Female M	Female SD	Male M	Male SD	Effect Size (d)
Pronouns		14.24	4.06	12.69	4.63	0.36
First-person singular	I, me, my	7.15	4.66	6.37	4.66	0.17
First-person plural	we, us, our	1.17	2.15	1.07	2.12	*ns*
Second person	you, you're	0.59	1.05	0.65	1.15	−0.06
Third person	she, their, them	3.41	3.45	2.74	3.01	0.20

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), 211–36.

SYRACUSE UNIVERSITY
School of Information Studies

GENDER CLASSIFICATION IN CONGRESS

Table 4 Gender differences in selected LIWC categories

LIWC dimension	Corpus	Female		Male		Effect size (d)	Result
		Mean	SD	Mean	SD		
Pronoun	NGHP	14.24	4.06	12.69	4.63	0.36	Disagree
	HS	7.55	0.01	7.69	0.01	−0.1	

Table 6 Gender differences in pronoun case use

Pronoun cases		Female		Male		Effect size (d)
		Mean	SD	Mean	SD	
Subjective	We	1.18	0.40	1.37	0.51	−0.39
	I	1.48	0.32	1.57	0.43	−0.21
Possessive	Our	0.76	0.30	0.58	0.28	0.64
	My	0.46	0.15	0.40	0.17	0.36
Objective	Us	0.22	0.10	0.22	0.10	0.00
	Me	0.15	0.07	0.15	0.08	−0.09

Congresswomen	Congressmen
"Our community"	"Our enemy"
"Our workforce"	"Our side"
"We honor"	"We ought"
"We share"	"We gave"

SYRACUSE UNIVERSITY
School of Information Studies