

“Large-scale Computerized Text Analysis in Political Science: Opportunities and Challenges,” *Annual Review of Political Science* (forthcoming)

On-line Appendix

John Wilkerson & Andreu Casas

September 1, 2016

This online appendix contains the following supplemental materials for the article “Large-scale Computerized Text Analysis in Political Science: Opportunities and Challenges” by John Wilkerson and Andreu Casas at the *Annual Review of Political Science*:

- 1) List of top terms of the 50 clusters used in Section 4 of the paper
- 2) Validation of the *Agriculture* cluster in Section 4.

In addition, the data and the code used in Section 4 of the article can be found in:
<https://github.com/CasAndreu/wilkerson-casas.2016-TAD>.

1 List of top terms of the 50 clusters in Section 4

#	Metatopic	issue
1	African Am Community	american, african, right, black, civil, king, first, alabama, nation, martin, histori, equal, histor, jone, movement
2	Agruculture	farm, agricultur, kentucki, farmer, mr, food, industri, famili, counti, produc, year, today, state, bureau, land
3	Arts	art, music, perform, mr, band, artist, news, show, festiv, arizona, mani, honor, cultur, talent, award
4	Budget	feder, state, govern, bank, requir, act, tax, financi, would, regul, report, law, new, credit, secur
5	Business	busi, compani, small, industri, chamber, manufactur, bank, commun, commerc, product, job, year, econom, success, famili
6	Congrats	busi, commun, forti, leader, iowa, record, great, mr, state, honore, de, career, year, class, moin
7	Congrats	award, work, dedic, hard, youth, receiv, ridg, servic, ambassador, wheat, arvada, applaud, extend, congratul, charact

8	Congrats	scout, eagl, rank, boy, project, achiev, mr, earn, award, work, space, girl, congratul, commun, well
9	Congrats	iowa, hero, red, american, cross, mr, help, tom, serv, state, heartland, year, indian, greater, great
10	Congrats	team, coach, state, championship, school, high, win, athlet, game, season, congratul, footbal, basketbal, titl, year
11	Congrats	texa, mrs, john, district, fort, north, behalf, mr, congratul, congression, today, second, rise, twenty, commun
12	Congrats	award, state, mr, iowa, nation, honor, program, repres, unit, congratul, commun, recogn, receiv, effort, servic
13	Congrats	th, celebr, anniversari, new, year, today, town, commun, citi, mr, histori, park, continu, histor, home
14	Congrats	award, commun, chamber, mr, busi, recogn, year, honor, commerc, recipi, receiv, counti, main, west, congratul
15	Congrats	serv, board, presid, member, commun, associ, counti, committe, director, elect, also, servic, state, council, award
16	Congress	bill, support, legisl, act, r, h, mr, would, congress, colleagu, today, hous, vote, committe, fund
17	Defense	servic, armi, medal, air, serv, unit, militari, forc, command, colonel, state, nation, marin, offic, corp
18	Defense	veteran, war, mr, servic, honor, nation, serv, militari, countri, american, today, world, men, memori, women
19	Defense	marin, honor, sergeant, action, medal, enem, unit, mr, armi, soldier, corp, wound, first, today, fire
20	Defense	air, forc, gener, militari, colonel, nation, command, aviat, base, unit, pilot, train, fli, oper, navi
21	Defense	war, mr, world, american, bob, serv, us, island, servic, immigr, countri, honor, guam, armi
22	Education	school, student, high, educ, teacher, nation, district, young, mr, princip, ms, elementari, colleg, recogn, program
23	Education	commun, program, servic, provid, educ, organ, center, develop, help, famili, children, need, hous, assist, support
24	Education	dr, univers, colleg, educ, state, degre, graduat, student, award, school, research, serv, presid, honor, year

25	Family	famili, life, love, friend, wife, pass, mr, grandchildren, children, commun, great, year, school, son, honor
26	Family	famili, life, love, children, friend, mrs, year, commun, mr, school, pass, wife, grandchildren, live, great
27	Gender	women, ms, woman, work, equal, girl, men, lo, angel, honor, make, histori, nation, lgbt, toni
28	Government	govern, american, freedom, citizen, mr, constitut, today, right, liberti, polit, individu, democraci, support, religi, countri
29	Healthcare	health, care, medic, hospit, center, servic, patient, provid, nurs, physician, clinic, access, healthcar, insur, mental
30	Healthcare	cancer, research, diseases, awar, children, autism, treatment, medic, health, today, nation, american, mr, support, year
31	Healthcare	job, health, american, care, work, insur, worker, republican, fund, hous, law, budget, obamacar, cut, plan
32	Human Rights	victim, human, violenc, women, children, traffick, abus, sexual, crime, mr, prevent, nation, justic, child, domest
33	Latino Community	florida, hispan, central, commun, month, counti, orlando, honor, south, puerto, recogn, heritag, mr, northwest, work
34	Law	law, court, judg, legal, justic, attorney, state, rabbi, practic, bar, patent, case, associ, lawyer, suprem
35	Law Enforcement	servic, mr, year, retir, work, career, serv, chief, offic, dedic, public, wish, director, manag, posit
36	Law Enforcement	offic, polic, depart, fire, enforc, sheriff, chief, law, counti, commun, year, serv, servic, safeti, deputi
37	Religion	church, pastor, baptist, reverend, commun, serv, congreg, year, rev, st, ministri, member, faith, christian, first
38	Religion	st, john, illinoi, chicago, cathol, joe, commun, year, parish, father, th, sister, paul, church, today
39	Resources	water, energi, michigan, nation, kentucki, museum, project, river, lake, land, protect, park, conserv, state, histor
40	Security	israel, nuclear, jewish, iran, state, unit, world, peopl, rabbi, intern, presid, mr, weapon, must, today

41	Trade	us, state, unit, taiwan, intern, american, presid, trade, countri, world, island, foreign, china, nation
42	<i>Unclear</i>	william, john, david, robert, jame, daniel, davi, jone, virginia, rabbi, smith, volunt, jack, honor, princ
43	<i>Unclear</i>	mr, mississippi, mrs, jackson, school, ask, black, join, state, african, commun, colleagu, jone, honor, high
44	<i>Unclear</i>	peopl, make, get, want, would, one, said, work, say, know, year, american, time
45	<i>Unclear</i>	california, san, valley, mr, counti, state, lo, fresno, year, commun, angel, tom, district, riversid, program
46	<i>Unclear</i>	citi, mayor, john, commun, council, colorado, public, serv, servic, park, resid, town, villag, year, tom
47	<i>Unclear</i>	counti, mrs, commun, servic, virginia, serv, board, west, year, joe, public, ms, work, mr, district
48	<i>Unclear</i>	new, york, john, jersey, mr, state, librari, main, joe, work, local, buffalo, year, jack, member
49	<i>Unclear</i>	washington, d, c, virginia, district, bob, mike, west, davi, mr, state, columbia, wisconsin, ohio, hous
50	<i>Melting Pot</i>	us veteran, award, marin, john, women, dr, children, ms, indiana, north, new, fire, state

2 Validation of the *Agriculture* cluster in Section 4

Are the metatopics in the article capturing actual substantive political issues? As we discussed in the paper, validation is a key step of any text as data project. As Grimmer and Steward (2013) point out, researchers using unsupervised classifiers often use a semantic and/or a predictive approach to validation. Semantic validation focuses on showing that the most predictive features of a topic are words that researchers would associate with that topic, and that topics that are substantially similar (e.g. Defense and Security) are also semantically similar (e.g. Quinn et al. 2010; Grimmer and King 2011). Predictive validation focuses on showing that topics predict real-life events that researchers would expect them to predict (e.g. Grimmer 2013).

In this case we present an example of predictive validation. We would expect Representatives from districts with larger populations working in the Agricultural sector to talk more about the meta-topic Agriculture. To test this, we collected district level data from the U.S. Census reporting the percentage of the total population that works in Agriculture, Forestry, Fishing, Hunting and Mining. The specific source is the 2014 American Community Survey 1-Year Estimates, and the 2014 County Business Patterns: <http://www.census.gov/mycd/>.

