Iterative Human Coding and Computational Text Analysis: Assessing the Effects of Public Pressure on Policy

Devin Judge-Lord | Harvard University | DevinJudgeLord@FAS.Harvard.edu

Human coding and computational text analysis are more powerful when combined in an iterative workflow.

1. Text analysis tools can strategically **select texts for human coders**—texts representing larger samples and outlier texts of high inferential value.
2. Preprocessing can **speed up hand-coding** by extracting features like names and key sentences.
3. Humans and computers can iteratively **tag entities** using regex tables and **group texts by key features** (e.g., identify lobbying coalitions by common policy demands)

Applying simple search and text-reuse methods to public comments on all U.S. federal agency rules, a **sample of 10,894 hand-coded comments** yields **41 million as-good-as-hand-coded comments** regarding both the organizations that mobilized them and the extent to which policy changed in the direction they sought.

Hand-coding dynamic data

Workflow: `googlesheets4` allows analysis and improving data in real-time. For example, in Fig. 1:

- The "org_name" column is populated with a guess from automated methods. As humans identify new organizations and aliases, other documents with the same entity strings are auto-coded to match human coding.
- As humans identify each organization's policy "ask," other texts with the same ask are put in their coalition.
- If the organization and coalition become known, it no longer needs hand coding.

Fig. 1: Coded Public Comments in a Google Sheet

url	txt	summary	org_name	coalition	coalition _type	org_type	ask	success
https:/	https:/	When initial fracturing operations are over, the	American Society of Civil Engineers	Sierra Club	public	ngo; advocacy	Mandate full public disclosure of all chemicals and other propping agents in the fracturing fluid.	1
https:/	https:/	Rodrigues ma aug president barack	Food & Water Watch	Sierra Club	public	ngo; advocacy	proposed rules for regulating hydraulic fracturing on Federal and Indian lands are not only weak, they do not	1
https:/	https:/	Our comments are as follows we	Kashia Band of Pomo Indians	Sierra Club	public	gov; tribe	we do not support hydraulic fracturing but we do support the BLM's endeavors to create oversight and	1
https:/	https:/	However, natural gas methane does leak	Global Change Consulting Consortium	Global Change Consulting Consortiu	public	ngo	No coverage of leakage to the atmosphere of methane as a potent greenhouse gas	1
https:/	https:/	While I understand blm's concern for	Montana	Energy Citizens	public	gov; state; governor	the proposed rules impose a redundant regulatory process	-1
https:/	https:/	Sweetwater county strongly objects to	Sweetwater County	Energy Citizens	public	gov; local; county	it is imperative that the BLM defer to the State	-2
https:/	https:/	The ndic and dmr are in the best position to	John HOEVEN; Heidi HEITCAMP; Kevin CRAMER	Energy Citizens	private	senate; house	request North Dakota and our tribes, and similar states, be exempted from the final rule	-1
https:/	https:/	This will ensure the public has adequate	Western Energy Alliance	Western Energy Alliance	private	corp group	an extension to the comment period	2
https:/	https:/	Txoga comment submission re oil and	Texas Oil and Gas Association	Western Energy Alliance	private	corp group	an extension of at least 120 days	2

Regex tables to tag entities

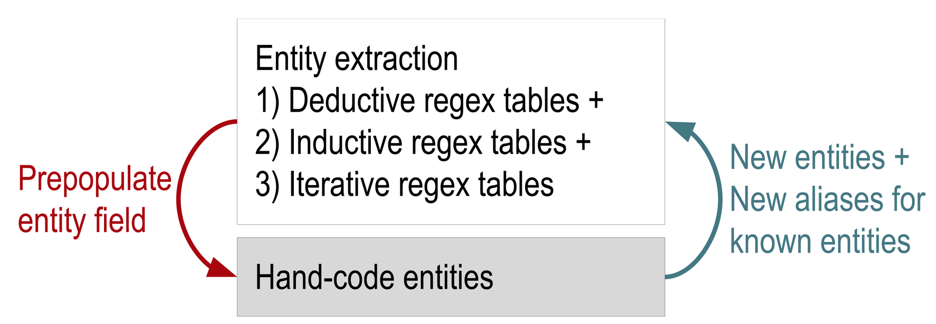
- **Deductive:** Start with databases of known entities.

Table 1: Lookup Table Deduced from Center for Responsive Politics Lobbying Data, Collapsed into an Initial Regular Expression Table

Entity	Pattern
3M Co	3M Co|3M Health Information Systems|Ceradyne|Cogent Systems|Hybrivet Systems
Teamsters Union	Brotherhood of Locomotive Engineers (and|&) Trainmen|Brotherhood of Maint[a-z]* of Way Employ|Teamsters

- **Inductive:** Add entity strings that frequently appear in the data to regex tables.
- **Iterative:** Add to regex tables as humans identify new entities or new aliases for known entities. Update data (Google Sheets) to speed hand coding.

Fig 2: Iteratively Building Regex Tables



For example, the `legislators` package uses a regex table, adding variants (e.g., "AOC") to standard legislator names to detect them in messy text.

Results: Who mobilizes public comments?

Of 58 million public comments on proposed agency rules, the top 100 organizations mobilized 43,938,811. The top ten organizations mobilized 25,947,612.

Table 2: The Top 5 Organizations Mobilized 20 Million Public Comments

Organization	Rules Lobbied On	Pressure Campaigns	Percent (Campaigns /Rules)	Comments	Average per Campaign
NRDC	530	62	11.7%	5,939,264	95,795
Sierra Club	591	110	18.6%	5,111,922	46,472
CREDO	90	41	45.6%	3,019,150	73,638
Environmental Defense Fund	111	31	27.9%	2,849,517	91,920
Center For Biological Diversity	572	86	15.0%	2,815,509	32,738
Earthjustice	235	59	25.1%	2,080,583	35,264

Grouping with text reuse

Fig. 3: Iteratively Group Documents

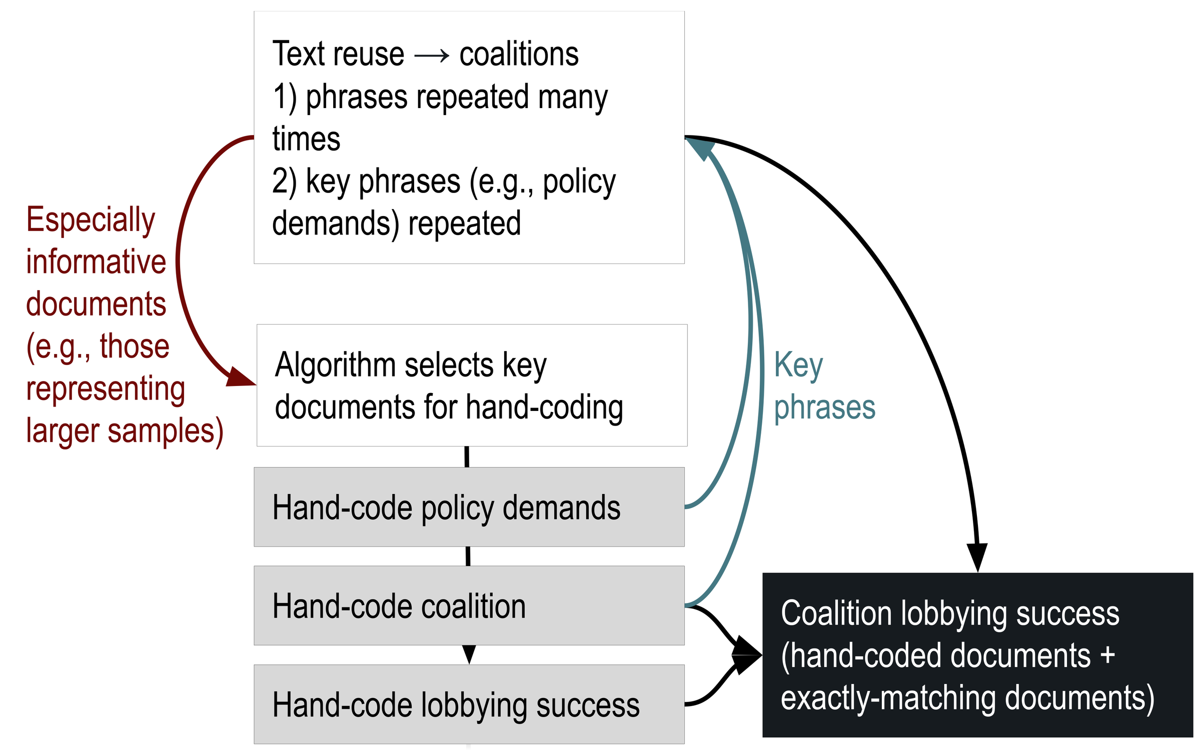
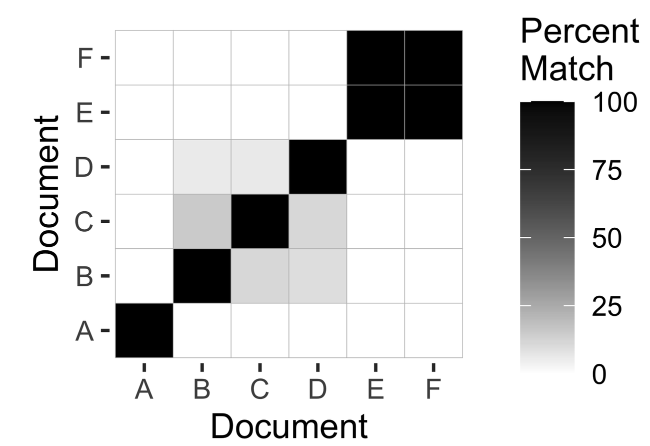


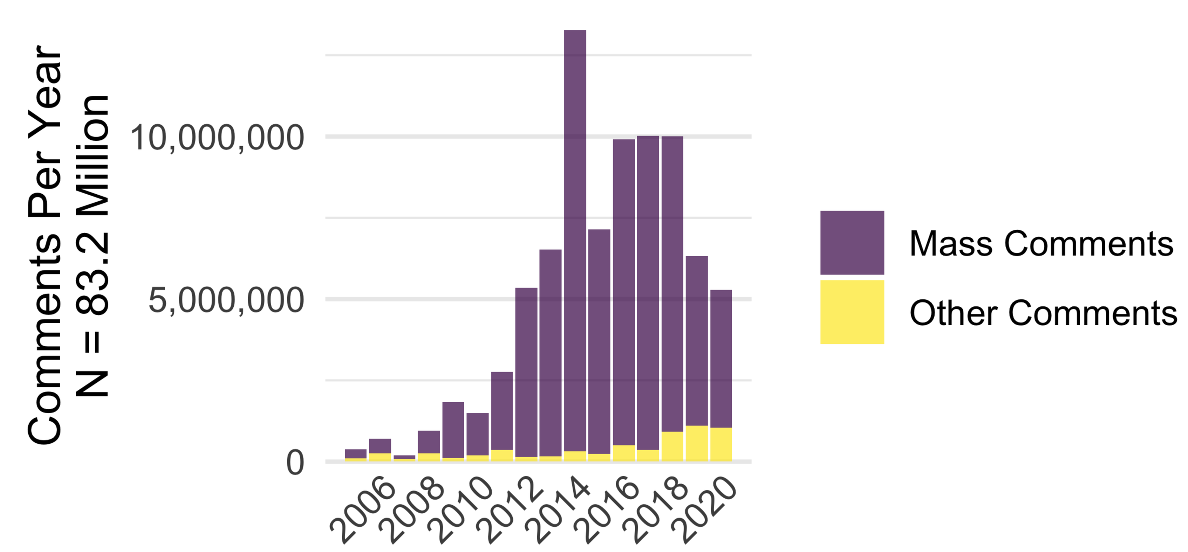
Fig 4: Identifying Groups of Linked Documents with Text Reuse (a 10-gram Window Function)



- Document A shares no 10-word phrases with the others
- B, C, and D share some text (they are part of an organized mass comment campaign)
- E and F are the same text that was submitted twice

Results: Most public comments result from organized pressure campaigns

Fig. 5: Public Comments on Regulations.gov, 2005-2020



Comments that share a 10-gram with 99 or more others are part of a mass comment campaign.

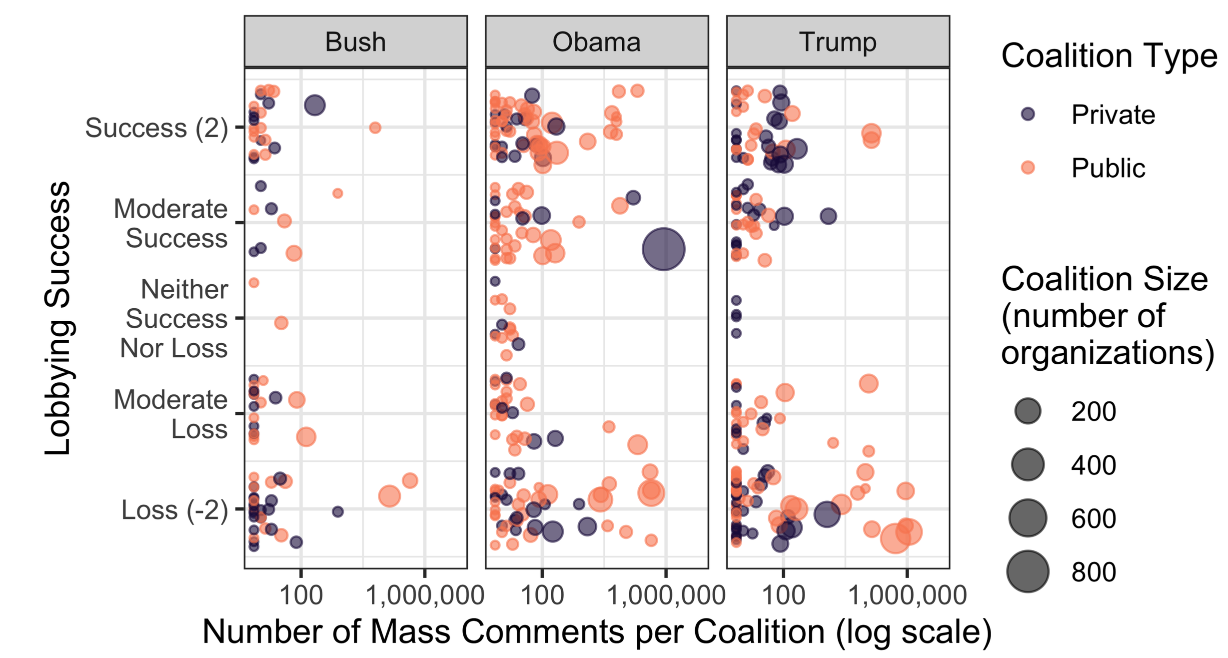
Grouping with key phrases

1. Humans identify groups of selected documents (e.g., lobbying coalitions)
2. Humans copy and paste key phrases
3. Computer puts other documents containing those phrases in the same group (coalition)

Preprocessing tip: **Summaries** speed hand-coding (e.g., use `textrank` to select representative sentences).

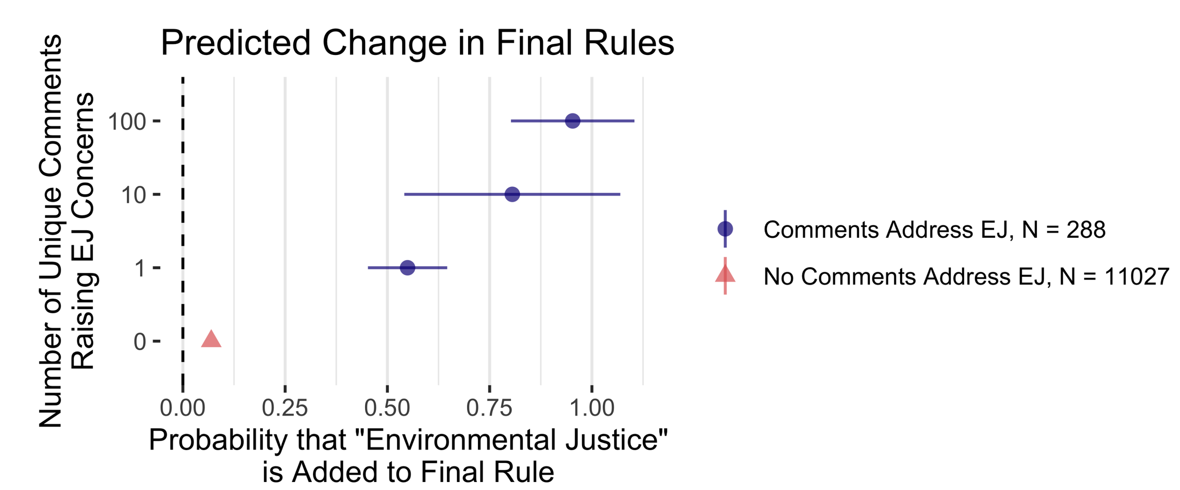
Results: Larger coalitions → more likely to win

Fig. 6: Lobbying Success by Campaign Size



Public pressure on climate and environmental justice greatly affected policy documents (Fig. 7), but a few organizations dominate lobbying coalitions (Table 2). When tribal governments or local groups lobby without the support of national advocacy organizations, policymakers typically ignore them.

Fig. 7: Policy Text Change by Coalition Size



Next steps

- Compare exact entity linking (regex tables) to probabilistic methods (`linkit`, `fastlink`, supervised classified with hand-coded training set)
- Compare exact grouping (e.g., by policy demands) to supervised probabilistic classifiers/clustering