Vermin's Victory, by Jasmine Kwok and Betsy Fridman

Statement of Problem

Overdispersed Strategies was recently commissioned by the 2028 presidential candidate Vermin Supreme to develop effective campaign messages that will reach voters, specifically those in swing states. This project aims to analyze linguistic patterns in past Congressional policy statements to uncover similarities and differences in writing styles of Democratic and Republican candidates, as well as between winning and losing candidates. Using Congressional campaign texts from 2018, 2020, and 2022 from CampaignView, this analysis uses both collocation and keyword analysis to understand how phrasing and modality shape persuasion throughout different candidates. This study will provide insights to support Vermin Supreme in tailoring his messages to appeal to a more diverse voter base and maximize his campaign effectiveness.

Summary of Findings

Collocation analysis is used to explore how Democrats and Republicans speak of key policy topics differently. The token *health* was chosen as the main node for the collocation analysis due to its high relative frequency in both the democratic and republican policy statement subcorpora, as seen in Table 1 of the Appendix. Additionally, health is a universal concern that all Americans are affected by, and people on both parties care deeply about issues relating to health. Next, keyword analysis on modal verbs was used to identify differences in the certainty of statements between winning and losing candidates. By looking at the log-likelihood ratio to compare the usage of modal verbs, we can identify which words are associated with better or worse results, and whether using more modal verbs can help candidates win. Together, these analyses reveal how word choice and the confidence in language can contribute to be persuasive and effective in political messages.

The collocation analysis on the noun health shows distinct linguistic stylistic differences between the Democratic and Republican policy statements. The following items were identified among the top ten collocates of the noun *health* in the Democrat policy statements using MI statistics (MI(5.5), L5-R5, NC15): *behavioral*, *substance*, *maternal*, *centers*, and *parenthood*, as seen in Table 2 of the Appendix. These associations suggest that democratic policy statements often include socially and ethically driven terms when speaking about health, which reflects a focus on public health and human welfare. On the other hand, Republican policy statements emphasize economic terms, framing *health* through a market-based lens by focusing on costs, transparency, and efficiency. Some examples of health collocates in Republican policy statements using are *affordable*, *savings*, *purchase*, *providers*, and *plans* (MI(5.5), L5-R5, NC15), as seen in Table 2 of the Appendix. Some shared *health* collocations present in both parties include *care*, *mental*, *women*, *coverage*, and *insurance*, as shown in the collocation network in Figure 1 of the Appendix; these highlight main concerns of the general public.

Analysis of modal verbs in policy speeches from swing states (Arizona, Nevada, Georgia, Michigan, North Carolina, Pennsylvania, Wisconsin, New Hampshire, and Minnesota) highlights clear distinctions between winning and losing campaigns. In Table 3 of the Appendix, the modals *can* (LL = −71.09), *should* (LL = −67.60), and *will* (LL = −58.02) stand out as statistically most distinctive, appearing with greater relative frequency in losing candidate speeches than in winning ones. In practical terms, this suggests that losing policy statements overuse modal verbs that convey potential and uncertainty, whereas winning campaigns favor more assertive and decisive sentences. These linguistic choices seem to influence how effectively candidates can persuade their audience in the swing states.

Recommendations

To align messaging with American values and voter expectations, Vermin Supreme should adopt a tone that emphasizes empathy, equity, and collective well-being. Language that highlights compassion, care, and accessibility will resonate with Democratic audiences, and speaking of integrating economic and structural framing helps engage the Republican audiences. For instance, discussing oral hygiene as he advocates for mandatory tooth brushing laws in terms of affordability, infrastructure, and preventive care, can appeal to voters concerned with both fiscal responsibility and systemic reform. Blending these two perspectives would allow Vermin to reach a broader audience, positioning health and hygiene as both a right and a responsibility. Framing messages around shared concerns, such as *care*, *insurance*, and *coverage*, can unify the diverse audiences and make Vermin's campaign's platform appear pragmatic and inclusive.

An analysis of modal verb use in swing-state policy statements suggests that winning candidates favor more confident, concrete language than uncertainty. Terms such as *can*, *should*, *will*, *must*, and *could* appear significantly more often in losing campaigns, indicating a tendency toward tentative or overly prescriptive phrasing. To strengthen his appeal, Vermin should minimize reliance on such modals and instead use direct, decisive statements that convey capability and commitment. For example, "We will ensure every American has a free pony and access to zombie-powered renewable energy" would sound much more confident and convincing, rather than "We should work toward getting zombie-powered renewable energy, and we may be able to get everyone a free pony". Framing policies in actionable, confident language will help project leadership, inspire trust, and communicate a sense of certainty and follow-through, which are traits that voters associate with successful candidates.

Appendix

Table 1 - Frequency and dispersion measures for top 12 most frequent tokens in Democratic Policy Statements (on left) and Republican Policy Statements (on right)

	Token	AF	Per_10.6	ARF	DP
1	must	2909	6311.468	1509.1234	0.5480505
2	can	2531	5491.346	1477.9805	0.4948113
3	need	2391	5187.597	1267.6700	0.5585371
4	support	2319	5031.384	1234.4773	0.5695486
5	people	2259	4901.206	1207.7328	0.5794187
8	health	2203	4779.706	965.4092	0.6846067
10	public	1778	3857.611	877.5242	0.6567440
7	work	1744	3783.844	989.2863	0.6006347
6	also	1693	3673.192	996.6986	0.5659547
9	make	1648	3575.559	953.0006	0.6005089
13	care	1607	3486.604	716.8844	0.7488527
18	education	1434	3111.257	598.6122	0.7780499

	Token	AF	Per_10.6	ARF	DP
1	must	1921	6047.403	995.2304	0.6376637
2	government	1715	5398.905	901.5821	0.6632071
4	need	1540	4847.996	801.8114	0.6519861
3	can	1470	4627.633	867.5659	0.6004990
6	federal	1376	4331.716	711.7762	0.6910310
5	support	1371	4315.976	718.6870	0.6849961
7	people	1279	4026.355	661.5828	0.6860158
8	american	1254	3947.654	655.3751	0.6942771
19	tax	1083	3409.338	425.0185	0.8166848
9	also	1071	3371.561	623.3844	0.6577928
21	health	1025	3226.751	395.5684	0.8121653
10	work	969	3050.460	549.9897	0.7280561

Table 2 - The top 12 most frequent tokens collocated with the word health in Democratic Policy Statements (on left) and Republican Policy Statements (on right)

token	col_freq	total_freq	PMI
behavioral	32	25	8.06
mental	348	423	7.43
substance	37	74	6.71
care	798	1607	6.70
insurance	289	585	6.69
maternal	15	33	6.57
centers	46	106	6.50
reproductive	72	201	6.23
parenthood	20	56	6.22
professionals	17	55	6.01
planned	31	101	6.00
coverage	97	317	6.00

token	col_freq	total_freq	PMI
mental	130	169	7.90
care	507	923	7.41
savings	34	65	7.34
womens	16	34	7.19
coverage	50	114	7.09
insurance	142	332	7.05
association	16	43	6.85
purchase	28	86	6.66
plans	30	95	6.61
affordable	64	210	6.56
providers	25	89	6.44
expanding	18	81	6.11

Figure 1 - Health Collocation Network of Democratic vs Republican Collocates with Health

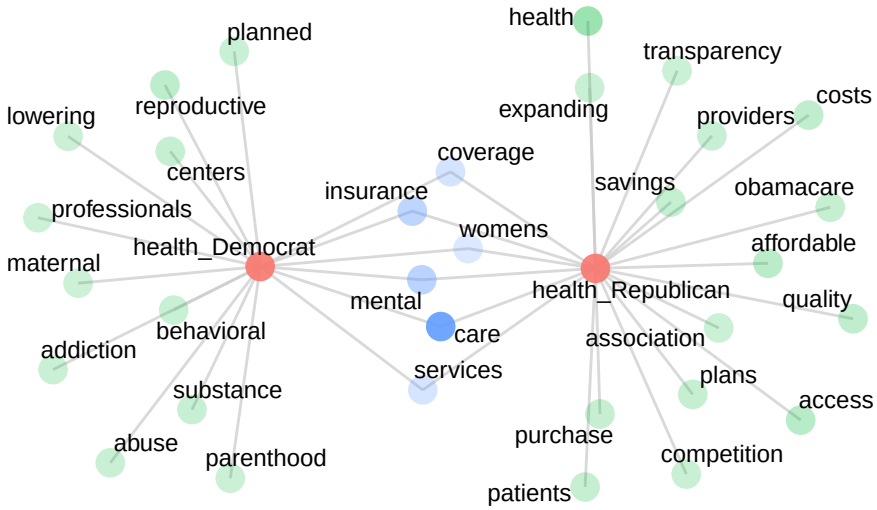

Table 3 - Keyness statistics for modal verbs from swing states with AF Tar > 1 or AF Ref > 1

Token	Tag	LL	LR	PV	AF_Tar	AF_Ref	Per_10.6_Tar	Per_10.6_Ref	DP_Tar	DP_Ref
ought	md	-0.71	-0.74	0.39879	5	6	11.32	18.85	0.996	0.996
would	md	-1.43	-0.10	0.23163	658	509	1489.83	1598.92	0.642	0.695
shall	md	-1.60	-0.47	0.20575	30	30	67.93	94.24	0.985	0.984
might	md	-2.62	-0.92	0.10543	11	15	24.91	47.12	0.993	0.978
may	md	-12.47	-0.77	0.00041	80	98	181.13	307.85	0.922	0.879
could	md	-15.93	-0.72	0.00007	117	139	264.91	436.64	0.887	0.850
must	md	-34.79	-0.34	0.00000	1293	1182	2927.59	3713.02	0.586	0.530
will	md	-58.02	-0.34	0.00000	2132	1951	4827.24	6128.69	0.425	0.440
should	md	-67.60	-0.56	0.00000	875	929	1981.16	2918.27	0.647	0.572
can	md	-71.09	-0.50	0.00000	1157	1182	2619.66	3713.02	0.521	0.477