CBE 1 Final Draft

Betsy Fridman

Introduction

For this experiment, I aim to investigate how language differs in editorial and reportage papers. My question of interest is, "How do the frequency of evaluative and stance-related words differ between the editorial and reportage genres?". This is particularly interesting because editorials are expected to convey more opinionated, subjective statements, whereas reportage generally aims to present unbiased, factual stories. By comparing the relative frequencies and keyness of evaluative and stance-related words across the two subcorpora,

Data

The corpus I used was the Brown Corpus, and was compiled in the 1960s. Containing 500 texts, it was the first major structured corpus of American English text that included different genres. For this study, I chose to compare the press editorial and press reportage subcorpora, made up of 27 and 44 texts respectively. The texts were tokenized and converted to lowercase to keep the words standardized.

Table 1: Overview of Subcorpora

Subcorpus	Press..Editorial	Press..Reportage
Tokens (words)	54515	88636
Files	27	44

Methods

To find the differences in language use, I focused on evaluative words, which convey judgement or feelings (e.g., *good*, *bad*, *significant*), and stance words, which express the author's certainty or perspective (e.g., modal verbs like *may*, *might*, *must*, or hedge words like *possibly* or *likely*). I selected these categories because editorials tend to be opinionated, while reportages tend to present factual information. By comparing the frequency and keyness of evaluative and stance words, I am able to highlight the linguistic difference between texts written based on opinion or fact. I created a document-feature matrix (DFM) for each subcorpus and computed their relative frequencies per 1,000 tokens, doing this separately for evaluative words and stance words. The log-likelihood keyness test was used to identify words that are statistically key to one subcorpus, relative to the other. Words with higher log-likelihood (G^2) values have a stronger association with a subcorpus, and are considered key.
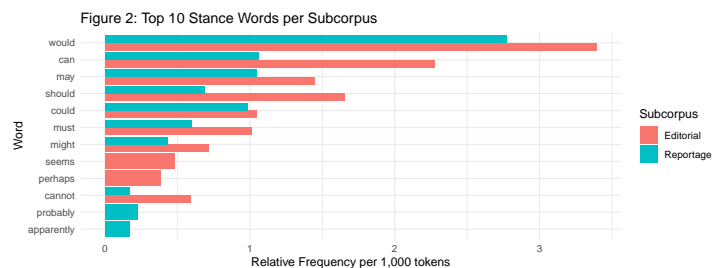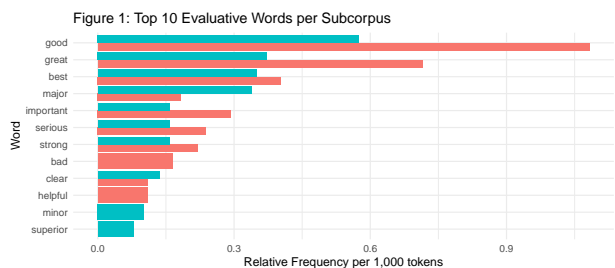
Results

Figure 1: Top 10 Evaluative Words per Subcorpus

Figure 2: Top 10 Stance Words per Subcorpus

Table 2: Editorial vs Reportage Evaluative Word Frequencies Side by Side

feature_ed	frequency_ed	rank_ed	docfreq_ed	per_1000_words_ed	feature_rep	frequency_rep	rank_rep	docfreq_rep	per_1000_words_rep
good	59	1	20	1.0822709	good	51	1	25	0.5753870
great	39	2	20	0.7153994	great	33	2	19	0.3723092
best	22	3	17	0.4035587	best	31	3	22	0.3497450
important	16	4	11	0.2934972	major	30	4	21	0.3384629
serious	13	5	9	0.2384665	strong	14	5	12	0.1579494
strong	12	6	9	0.2201229	serious	14	5	12	0.1579494
major	10	7	9	0.1834358	important	14	5	11	0.1579494
bad	9	8	7	0.1650922	clear	12	8	11	0.1353852
helpful	6	9	5	0.1100615	minor	9	9	7	0.1015389
clear	6	9	5	0.1100615	superior	7	10	6	0.0789747

Table 3: Editorial vs Reportage Stance-Related Word Frequencies Side by Side

feature_ed	frequency_ed	rank_ed	docfreq_ed	per_1000_words_ed	feature_rep	frequency_rep	rank_rep	docfreq_rep	per_1000_words_rep
would	185	1	27	3.3935614	would	246	1	40	2.7753960
can	124	2	25	2.2746033	can	94	2	34	1.0605172
should	90	3	24	1.6509218	may	93	3	31	1.0492351
may	79	4	21	1.4491424	could	87	4	34	0.9815425
could	57	5	24	1.0455838	should	61	5	24	0.6882080
must	55	6	20	1.0088966	must	53	6	22	0.5979512
might	39	7	21	0.7153994	might	38	7	19	0.4287197
cannot	32	8	14	0.5869944	probably	20	8	17	0.2256420
seems	26	9	13	0.4769330	apparently	15	9	11	0.1692315
perhaps	21	10	12	0.3852151	cannot	15	9	11	0.1692315

Editorials have higher relative frequencies of evaluative words (e.g., *good*, *great*, *best*, *important*) and of stance words (e.g., *would*, *can*, *should*), whereas reportage uses them less frequently. Log-likelihood analysis is used to indicate proportionally how much more often a word is used, with the editorials as my target corpus and the reportages as my reference corpus.

Table 4: Top 10 Evaluative and Stance Words by Log-Likelihood (G^2)

feature_eval	G2_eval	p_eval	n_target_eval	n_reference_eval	feature_stance	G2_stance	p_stance	n_target_stance	n_reference_stance
helpful	5.5239062	0.0187583	6	0	seems	5.831082	0.0157454	26	11
good	2.3566733	0.1247480	59	51	cannot	5.824297	0.0158063	32	15
bad	1.7087903	0.1911424	9	5	certainly	5.390626	0.0202452	19	7
great	1.6523803	0.1986353	39	33	should	5.007715	0.0252346	90	61
desirable	0.8428972	0.3585695	5	2	can	3.541885	0.0598374	124	94
harmful	0.6468323	0.4212475	2	0	tends	3.540815	0.0598760	5	0
unreasonable	0.6468323	0.4212475	2	0	suggest	2.445235	0.1178822	8	2
important	0.4892631	0.4842562	16	14	suggests	2.424528	0.1194492	4	0
appropriate	0.3833490	0.5358164	3	1	appears	2.187530	0.1391323	8	3
negative	0.3052726	0.5805957	4	2	perhaps	1.182490	0.2768496	21	14

For evaluative words, *helpful* is signficant (G^2=5.52, $p = 0.019$), and *good*, *bad*, and *great* appear more frequent in editorials but their frequencies aren't different with statistical significance. For stance words, *seems* (G^2=5.83, $p = 0.016$), *cannot* (G^2=5.82, $p = 0.016$), *certainly* (G^2=5.39, $p = 0.020$), and *should* (G^2=5.01, $p = 0.025$) have statistically significant higher frequencies in editorials. These results indicate that editorial writers use more modal verbs or hedging words.

Table 5: Bottom 10 Evaluative and Stance Words by Log-Likelihood (G^2)

feature_eval	G2_eval	p_eval	n_target_eval	n_reference_eval	feature_stance	G2_stance	p_stance	n_target_stance	n_reference_stance
major	-8.9530934	0.0027700	10	30	would	-14.8817987	0.0001145	185	246
minor	-4.7197927	0.0298173	1	9	could	-8.1647209	0.0042713	57	87
superior	-2.8623327	0.0906759	1	7	believes	-2.5664989	0.1091491	3	8
critical	-2.0195423	0.1552859	1	6	probably	-2.3783118	0.1230304	12	20
clear	-1.4590598	0.2270798	6	12	may	-1.9234374	0.1654782	79	93
inappropriate	-1.2766543	0.2585220	0	1	clearly	-0.6020728	0.4377883	8	11
inferior	-1.2766543	0.2585220	0	1	tend	-0.5622462	0.4533566	0	2
notable	-1.2766543	0.2585220	0	1	apparently	-0.0108261	0.9171310	15	15
unsatisfactory	-1.2766543	0.2585220	0	1	appear	-0.0035683	0.9523668	5	5
best	-0.7538534	0.3852590	22	31	might	-0.0027516	0.9581653	39	38

For evaluative words, only *major* (G^2=-8.95, $p = 0.003$) and *minor* (G^2=-4.72, $p = 0.030$) occur more often in reportage texts with statistical significance. For stance words, the difference in frequency of *would* (G^2=-14.88, $p = 0.0001$) and *could*

(G^2=-8.16, $p = 0.004$) has statistical significance. In both cases, most evaluative or stance-related words don't appear more common in reportages with statistical significance.

Discussion

The results support the expectation that the editorials use both evaluative and stance-related language more frequently than reportages do. My analysis suggests that evaluative words are more present in the editorial subcorpus, but they may not be able to distinguish the two genres. On the other hand, the results for stance words showed a statistically significant difference between the two subcorpora. In editorials, stance markers like *seems*, *cannot*, *certainly*, and *should* were significantly more likely to be used; this reflects how editorial writing can have more difference in certainty or possibility well, when reportage tends to remain more neutral. From this analysis, stance-related words are more reliable for distinguishing genres.

This study has many limitations. The Brown Corpus was put together in the 1960s, and may not accurately represent current news writing practices. The two subcorpora that I chose from the Brown Corpus, editorial press and reportage press, were of size 27 and 44 respectively. Not only are the subcorpora different sizes, but they were of small size, both of which can largely influence the analysis. Another limitation is the methods of choosing evaluative and stance words. Although I chose commonly used words from these categories, I don't have a comprehensive set of words that includes all of the possible evaluative and stance words, and some relevant ones in the corpus may have been left out of the analysis.