

Using data science to estimate issue salience among elite journalists and opinion formers

Paul Heden, 2021

Introduction

[Issue salience is a concept used to denote the importance of issues](#). Broadly speaking, the higher the salience of an issue in the media, the more likely it is that it will feature prominently on the political agendas of people with power. ¹

Issue salience doesn't exist in a vacuum. The way issues are framed in the media and the authority of the "salient actors" who push and comment on those issues can influence the type and range of political actions that are taken in response.

Can we use natural language processing techniques to understand issue salience in the media?

I wanted to find out what I could learn about "issue salience" and "salient actors" by analysing the output of a number of elite UK economics journalists.

Using the data science programming language "R" I developed code to:

- Scrape, download and clean a corpus of online news content;
- Use text mining techniques to perform basic word frequency analysis; and
- Produce visual materials (graphs, word clouds, bar charts) to illustrate a set of actionable insights.

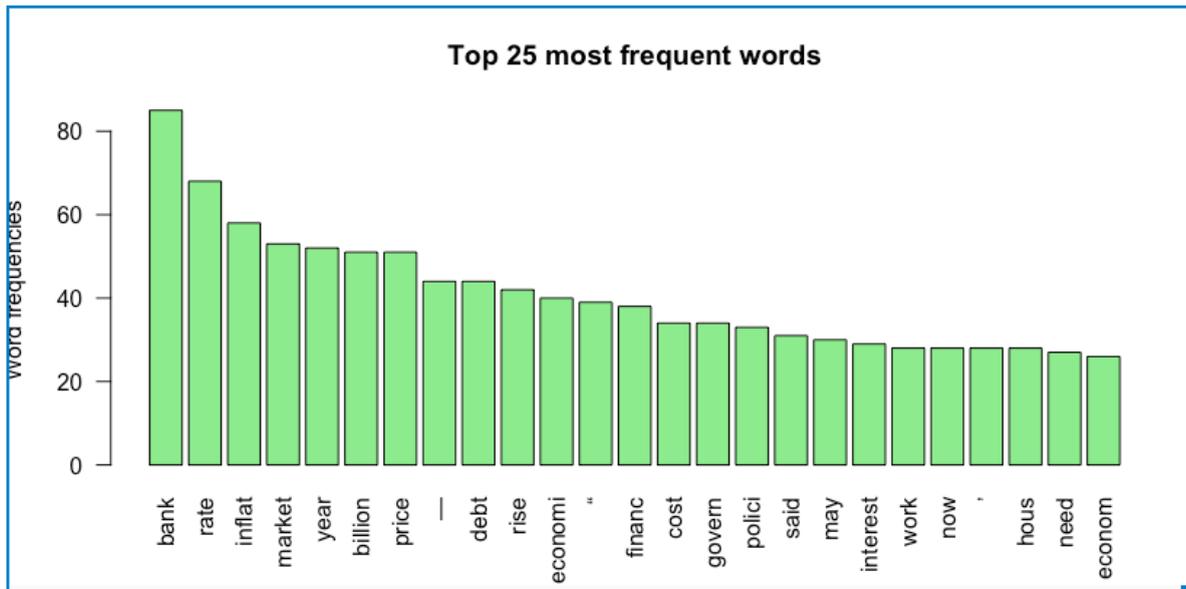
Method: Developed a corpus of words based on two months (Jul to August 2021) of comment pieces by x3 influential economics editors at UK national print newspaper titles.

(NB - The findings below are from a study of one of the three journalists under consideration as an example of the kind of findings that can be uncovered by this technique.)

Findings:

Word frequency: "*Word frequency*" gives an initial marker of the number of times a particular word is used and can be helpful in drawing conclusions about the issues that are "*salient*" for the journalist under scrutiny.

¹ Media is just one advocacy tool open to people or campaigns looking to make political change. Of at least equal importance is political and movement-led advocacy and research.



- The research ranked the top 300 words most frequently used during the study period.
- “Bank” was one of the highest frequency words (80+ mentions).
- “Inflation” was mentioned 85 times, by contrast “deflation” was mentioned 8 times.
- “Interest” and “rise” are also relatively high frequency words.
- “Austerity” appears just six times.
- No mention of “green” “climate” or “net” (unlike LE’s copy) which suggests climate change policy economics is less salient as an issue for this journalist (PA)
- No mention of “union” or “labour”. By contrast “labour” was the second most frequent term used by LE.

Sourcing strategy: what can word frequency analysis tell us about who is an “authoritative source” for this journalist?

All stories require sources and the sources that journalists reference are a strong marker of who they (or their media publication) believes are an “authority” in relation to a particular issue.

The role of a source within a news story is to take a “point of view” on issues and to “frame” those issues in a particular way.

Salient actor findings:

- “Sunak” 19 mentions.
- “Chancellor” 16 mentions.
- “Haldan” (a stemmed version of “Haldane”, ex Bank of England Governor) and “Governor” 15 mentions.
- No other actors were mentioned in the corpus of words under study.

This reflects something about the influence that the Treasury and Bank of England have over this journalist’s work. It is a marker of the authoritative status and the footprint these salient actors have on the issue space this reporter works in.

Conclusion

By applying data science techniques to a corpus of journalistic work I was able to:

- Draw conclusions about the salient issues that interest this journalist;
- Understand which salient actors matter to this journalist;
- Draw conclusions about which issues and actors are missing from this journalist's coverage.

This type of analysis can provide useful intelligence for campaigners and media professionals who need to understand which issues and actors are driving elite media debates.

It can help campaigners understand the issues the media is concerned with. It can also help campaigners draw reliable assumptions about who is influencing the debate and who is absent from it. Such knowledge could be helpful when thinking about the types of interventions needed to shift or influence the media discourse

This analysis was applied to media content. However, the same techniques can be applied to any digital corpus of words: for example, it could be used to analyse speeches by MPs, policy documents, blogs or digital booklets.

Appendix

	word	freq						
bank	bank	85	furlough	furlough	22	level	level	14
rate	rate	68	get	get	22	come	come	14
inflat	inflat	58	new	new	22	sector	sector	14
market	market	53	time	time	22	say	say	14
year	year	52	last	last	21	mean	mean	14
billion	billion	51	public	public	21	borrow	borrow	14
price	price	51	yet	yet	21	cut	cut	14
-	-	44	make	make	21	competit	competit	14
debt	debt	44	like	like	20	ear	ear	13
rise	rise	42	back	back	19	month	month	13
economi	economi	40	peopl	peopl	19	obr	obr	13
"	"	39	gdp	gdp	19	compani	compani	13
financ	financ	38	monetari	monetari	19	start	start	13
cost	cost	34	chang	chang	19	leav	leav	13
govern	govern	34	fund	fund	19	covid	covid	13
polici	polici	33	sunak	sunak	19	effect	effect	13
said	said	31	money	money	19	today	today	13
may	may	30	unemploy	unemploy	18	product	product	13
interest	interest	29	increa	increa	18	england	england	13
work	work	28	labour	labour	18	countri	countri	12
now	now	28	treasuri	treasuri	18	point	point	12
,	,	28	spend	spend	18	take	take	12
hous	hous	28	even	even	17	target	target	12
need	need	27	low	low	17	everi	everi	12
econom	econom	26	power	power	17	though	though	12
million	million	26	sinc	sinc	16	consum	consum	12
one	one	26	scheme	scheme	16	less	less	12
can	can	26	crisi	crisi	16	servic	servic	12
busi	busi	25	invest	invest	16	noth	noth	12
growth	growth	25	two	two	16	state	state	12
central	central	24	way	way	16	help	help	11
job	job	23	nation	nation	16	still	still	11
higher	higher	23	chancellor	chancellor	16	risk	risk	11
week	week	23	use	use	15	support	support	11
pandem	pandem	22	fiscal	fiscal	15	worker	worker	11
			real	real	15	end	end	11

Word cloud

Below is a word cloud base on the word frequency analysis carried out above.

