

Topic Modelling and Sentiment analysis of Government News on Covid-19 in Hong Kong and Mainland China

Winnie Huiheng Zeng¹, Menghan Jiang², Jieyu Chen³, Chu-Ren Huang⁴, & Kathleen Ahrens⁵

The Hong Kong Polytechnic University^{1,4,5}
Shenzhen MSU-BIT University²
South China Agricultural University³

JWLLP31, 16 Dec 2022, Kyung Hee University



- Effective government communication is vital for delivering information to interact with and influence the public's knowledge and behaviors during the pandemic (Capano et al., 2020; Hale et al., 2021).
- This study aims to investigate the **communication strategies of the governments** in Hong Kong and Mainland China in response to the Covid-19 pandemic.
- Data: Corpus of government news on Covid-19 in **Hong Kong (HK)** and **Mainland China (ML)**



RQ1: Which topics are discussed in the government news on Covid-19 in Hong Kong and Mainland China?

RQ2: What are the rhetorical strategies (e.g., metaphorical framing) used in the government news on Covid-19 in Hong Kong and Mainland China?

RQ3: What is the sentiment polarity of the government news on Covid-19 in Hong Kong and Mainland China?



- Corpus creation
- Data Analysis
 - Part 1: Topic modeling (**RQ1**)
 - Part 2: Rhetorical strategy analysis (**RQ2**)
 - 战争 *zhan4zheng1* WAR metaphorical frame
 - Part 3: Sentiment analysis (**RQ2**)



- Two parallel corpora (Timeframe: December 2019 to July 2021)
 - **HK corpus:** 5,662 Chinese government news on Covid-19 in [Hong Kong \(HK\)](#) (<https://www.news.gov.hk/chi/index.html>)
 - **ML corpus:** 2,304 Chinese government news on Covid-19 in [Mainland China \(ML\)](#) (<http://www.gov.cn/xinwen/index.htm>)
- Data cleaning and pre-processing
 - Chinese text segmentation and POS tagging
 - Jieba 0.42.1, <https://github.com/fxsjy/jieba> /Python 3.6.13



Latent Dirichlet Allocation-LDA (Blei et al., 2003)

- 1) Removal of stop words
- 2) Removal of “疫情” 相关词 (i.e., “疫情”, “新冠”, “肺炎”, “冠状”, “冠状病毒”)
- 3) Turn the tokens into a bag-of-words, i.e., a list of tuples with the token and its number of occurrences in the corpus
- 4) Determine the topic number
 - Topic coherence measure
 - Manual interpretation of the LDA results
- 5) LDA training (Topic numbers: 2, 3.....16)
- 6) Visualization



Stop word list (1,117)

- HK corpus: traditional Chinese stop word list
- ML corpus: simplified Chinese stop word list

百度停用词		中文停用词		哈工大停用词
一		\$		—
一下		0)) ,
一些		1) ÷ (1 -
一切		2		" ,
一则		3)、
一天		4		= (
一定		5		:
一方面		6		→
一旦		7		°C
一时		8		&
一来		9		*
一样		?		—
一次		=		~~~~
一片		"		'
一直		"		.
一致		、		『
一般		。		. —
一起		《		./
一边		》		--
一面		—		』
万一		一些		="
上下		一何		【
上升		一切		[*]
上去		一则		} >
上来		一方面		[⑤]]
上述		一旦		[① D]
上面		一上		~ 1

Topic Modeling - Coherence Score



- Topic number selection: N=2, 4

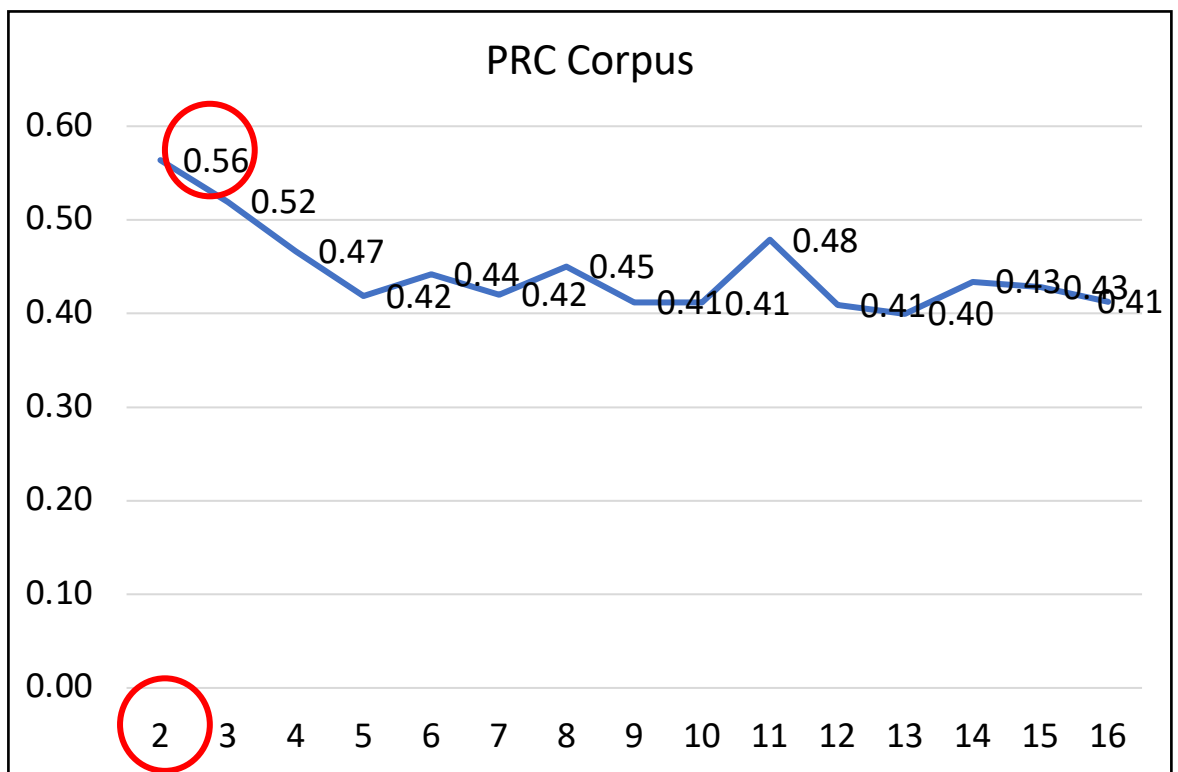
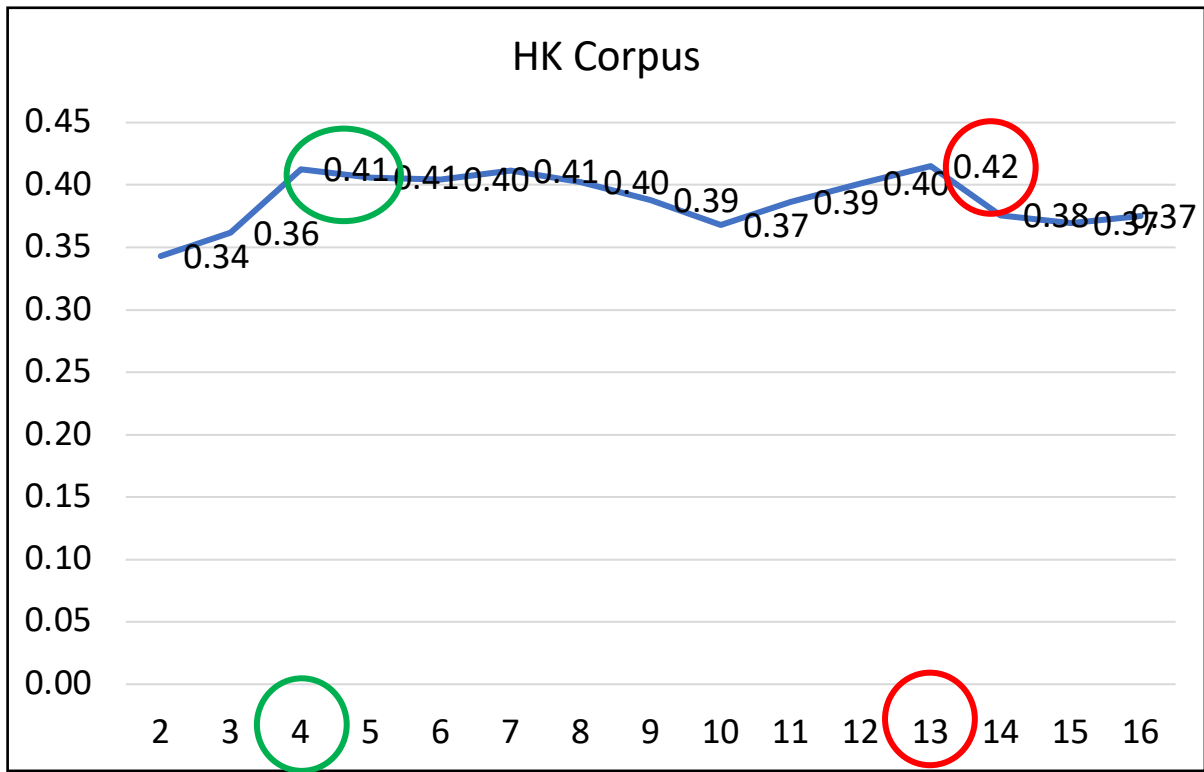


Figure. Coherence scores for the topic numbers



HK corpus

- Theme #1: 疫苗接种和政府 “vaccination and government”
- Theme #2: 检测和个案 “detection and cases”

ML corpus

- Theme #1: 中国和防控工作 “China and the prevention and control”
- Theme #2: 确诊病例和境外输入 “confirmed cases and imported cases from abroad”



Insights into the different government responses to the Covid-19 pandemic between Hong Kong and Mainland China:

The results of the LDA modeling indicate variations between the two regions:

- Mainland China: A proactive strategy with an emphasis on the topic of 防控 *fang2kong4* ‘prevention and control’ initiatives
- Hong Kong: A reactive strategy by focusing on the 检测 *jian3ce4* ‘detection’ measures.
- The Mainland China government responded to the Covid-19 pandemic from the perspective of the government, and the Hong Kong government responded more from the perspective of the public citizens.



- Metaphor has been found to be an effective rhetorical device to shape the human experience of epidemic diseases, such as the conceptualizations of AIDS (Sontag, 1989), SARS (Wallis & Nerlich, 2005), and flu (Dobric & Weder, 2016) in media discourse.
- WAR metaphor is the predominant type to talk about the Covid-19 pandemic in public discourse (Craig, 2020; Wicke & Bolognesi, 2020, 2021).

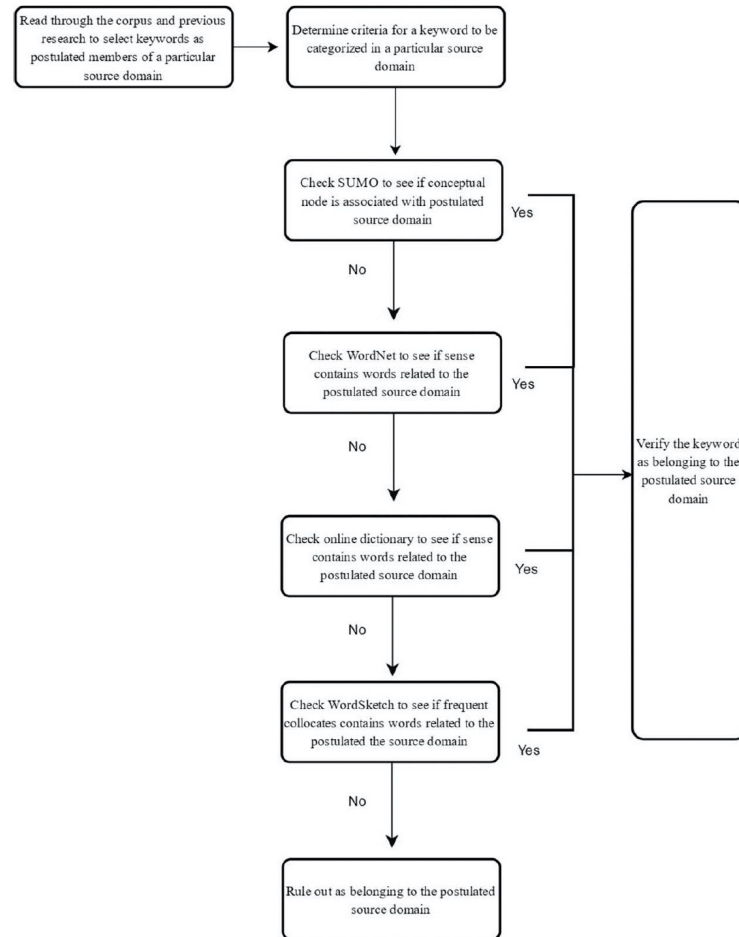


- To see the lexical units (i.e., metaphorical keywords) within WAR metaphorical frame are used to talk about Covid-19 in government news in Hong Kong and Mainland China?
- How does the use of the pervasive figurative framing of WAR change over time?



Determine lexical units associated with metaphors of WAR

- Analysis of sample data
- Source domain verification procedures (Ahrens & Jiang, 2020)





Frequencies of 50 Common Lexical Units of WAR Metaphors in ML Corpus

No.	WAR Keywords	Frequency
1	防控	11219
2	抗疫	1135
3	部署	959
4	一线	705
5	阻击战	364
6	战胜	328
7	斗争	324
8	打赢	315
9	战略	206
10	攻坚	191
11	战	176
12	防	137
13	胜利	135
14	和平	118
15	防线	118
16	攻关	95
17	奋战	80
18	打击	79
19	保卫战	67
20	抗	66
21	战斗	66
22	军队	66
23	第一线	64
24	威胁	56
25	前线	40

No.	WAR Keywords	Frequency
26	战场	31
27	暴力	25
28	战疫	20
29	敌人	19
30	抵制	17
31	战争	16
32	堡垒	16
33	战役	14
34	阵地	13
35	战果	13
36	主战场	13
37	保卫	11
38	歼灭战	11
39	抗争	9
40	对垒	9
41	攻击	7
42	战士	7
43	打仗	6
44	战术	4
45	打败	3
46	防守	3
47	武装	2
48	伤亡	2
49	击败	1
50	守卫	1
	Total	17382



• Sentiment analysis of WAR metaphorical framing

	A	B	C	D	E	F
1	Concordance	Concordance	WAR Lexical Unit	Target Domain	Sentiment	File names
2	疫情/n 快速/d 响应/v 处置/v , /x 强化/v 重点/n 人员/n 疫苗/n 接种/v , /x 从严/d 落实/a 外防/n 输入/v 措施/n 等/u 方面/n 对/p 疫情/n 防控/vn 工作/vn 作出/v 了/ul 具体/a	部署/n . /x 《/x 通知/v 》 /x 强调/v , /x 按照/p “/x 管/vn 行业/n 必须/d 管/vn 安全/an、 /x 管/vn 业务/n 必须/d 管/vn 安全/an、 /x 管/vn 生产/vn 经营/vn 必须/d 管/vn 安全/an”/x 的/	部署	疫情防控	Positive	中国政府网_政策_20210430_2107.txt
3	/v 做好/v 防控/vn 应急/vn 物资/n 储备/vn 等/u 工作/vn . /x 市场监管/n 总局/n、 /x 农业/n 农村部/n、 /x 国家/n 林业/n 和/c 草原/n 局/n 等/u 多个/m 部门/n 也/d 作出/v 具体/a	部署/n , /x 加强/v 重点/n 环节/n 重点/n 场所/n 监管/vn , /x 突出/v 饲养/v、 /x 繁育/v、 /x 运输/vn、 /x 出售/vn、 /x 购买/v 等/u 环节/n , /x 加强/v 检验/vn 检疫/vn 力度/n , /x 对/p 竹鼠/	部署	疫情防控	Positive	中国政府网_新闻_20200123_11.txt
4	/v . /x 国资委/nt 党委/j 26/m 日/m 召开/v 专题会议/n , /x 对/p 国资委/nt 和/c 中央/n 企业/n 扎实/a 做好/v 新型/b 冠状病毒/l 感染/v 的/uj 肺炎/n 疫情/n 防控/vn 工作/vn 做出/v 具体/a	部署/n . /x 中央/n 企业/n 快速/d 响应/v , /x 在/p 投入/v 大量/n 人力物力/n 主动/b 参与/v 疫情/n 防控/vn 的/uj 同时/c , /x 积极/ad 捐款捐物/n . /x 招商局/nt 集团/n、 /x 东风公司/n、 /x 国家/n 能源/	部署	疫情防控	Positive	中国政府网_新闻_20200129_72.txt
5	坚决/ad 打赢/v 疫情/n 防控/vn 阻击战/n 系列/q 述评/v 之一/r 新华社/nt 记者/n 这/r 是/v 在/p 举国上下/l 万众一心/i 抗击/vn 新型/b 冠状病毒/l 感染/v 肺炎/n 疫情/n 关键时刻/n 的/uj 重要/a	部署/n . /x 3/m 日/m , /x 中央政治局/nt 常务委员会/n 召开/v 会议/n , /x 研究/vn 下/f 一步/m 疫情/n 防控/vn 工作/vn . /x 习近平/nrfq 总书记/n 强调指出/l , /x 同/p 时间/n 赛跑/v、 /x 与/p 病魔/n 较量/d , /	部署	疫情防控	Positive	中国政府网_新闻_20200205_229.txt
6	/d 命令/n , /x 防控/vn 就是/d 责任/n . /x 以/p 习近平/nrfq 同志/n 为/p 核心/n 的/uj 党中央/nt 以/p 对/p 人民/n 高度负责/n 的/uj 使命/v 担当/v , /x 在/p 关键时刻/n 作出/v 重要/a	部署/n , /x 吹响/vn 疫情/n 防控/vn 阻击战/n 的/uj 嘹亮/a 号角/n . /x 一系列/m 务实/d 有效/a 的/uj 举措/v , /x 彰显/v 与/p 病魔/n 较量/d 的/uj 决心/v -/x -/x 在/p 党中央/nt 坚强/v	部署	疫情防控	Positive	中国政府网_新闻_20200205_255.txt
7	/n 防控/vn 一线/m 的/uj 医务/n 工作者/n 和/c 广大干部/ns 职工/n 送/v 去/v 党中央/nt 的/uj 关怀/nr 和/c 慰问/v , /x 对/p 进一步/d 做好/v 疫情/n 防控/vn 工作/vn 作出/v 重要/a	部署/n , /x 为/p 我们/r 打赢/v 疫情/n 防控/vn 阻击战/n 注入/v 强大/a 信心/n 和/c 力量/n . /x 疫情/n 就是/d 命令/n , /x 生命/vn 重于泰山/i . /x 在/p 以/f 习近平/nrfq 同志/n 为/p 核心/	部署	疫情防控	Positive	中国政府网_新闻_20200211_366.txt
8	/v 疫情/n 防控/vn 的/uj 原则/n、 /x 思路/n、 /x 目标/n、 /x 任务/n、 /x 重点/n 等/u 提出/v 新/a 要求/v , /x 对/p 统筹/v 疫情/n 防控/vn 与/p 经济社会/n 发展/vn 作出/v 新/a	部署/n , /x 是/v 指导/n 做好/v 疫情/n 防控/vn 和/c 当前/t 各项/r 工作/vn 的/uj 重要/a 遵循/v . /x 国务院/nt 常务会议/n 和/c 中央/n 应对/v 新冠/n 肺炎/n 疫情/n 工作/vn 领导/n 小组会议/	部署	疫情防控	Positive	中国政府网_新闻_20200214_455.txt
9	发生/v 以来/f , /x 湖北/ns 多/m 地/n 确诊/v 人数/n 也/d 不断/d 增加/v、 /x 持续/vd 高位/n , /x 医疗/n 资源/n 缺乏/v , /x 救援/vn 物资/n 告急/v ! /x 关键时刻/n , /x 党中央/nt 紧急/a	部署/n , /x 充分考虑/n 疫情/n 现状/n、 /x 人力资源/n 储备/vn 及/c 受援/v 地/uv 市/n 医疗/n 资源/n 缺口/n 等/u 情况/n , /x 统筹安排/n 19/m 个/m 省份/n 对口/vn 支援/v 湖北省/ns 除/p 武汉市/	部署	疫情防控	Positive	中国政府网_新闻_20200218_507.txt



- Part3: Sentiment analysis

- Baidu's platform (Cheng, Wang, & Huang, 2021)

https://ai.baidu.com/tech/nlp_apply/sentiment_classify

- WAR metaphorical frames in the HK corpus

- Diachronic patterns

- RQ1: Which topics are discussed in government news on Covid-19 in Hong Kong and Mainland China, and **how do they change over time with the development of the pandemic?**
- RQ2: How does the use of the pervasive figurative framing of WAR **change over time?**
- RQ3: What is the sentiment polarity of the government news on Covid-19 in Hong Kong and Mainland China, and **how does it change over time?**



- Ahrens, K., & Jiang, M. (2020). Source domain verification using corpus-based tools. *Metaphor and Symbol*, 35(1), 43-55.
- Capano, G., Howlett, M., Jarvis, D. S., Ramesh, M., & Goyal, N. (2020). Mobilizing policy (in) capacity to fight COVID-19: Understanding variations in state responses. *Policy and Society*, 39(3), 285-308.
- Craig, D. (2020). Pandemic and its metaphors: Sontag revisited in the COVID-19 era. *European Journal of Cultural Studies*.
- Dobric, N., & Weder, F. (2016). Media conceptualizing illnesses-the case of the flu. *Continuum*, 30(1), 126–142.
- Hale, Thomas, et al. "A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker)." *Nature human behaviour* 5.4 (2021): 529-538.
- Sontag, S. (1989). *Illness as metaphor and AIDS and its metaphors*. New York: Doubleday.
- Wallis, P., & Nerlich, B. (2005). Disease metaphors in new epidemics: The UK media framing of the 2003 SARS epidemic. *Social Science and Medicine*, 60(11), 2629–2639.
- Wicke, P., & Bolognesi, M. M. (2020). Framing COVID-19: How we conceptualize and discuss the pandemic on Twitter. *PloS one*, 15(9), e0240010.
- Wicke, P., & Bolognesi, M. M. (2021). Covid-19 discourse on twitter: How the topics, sentiments, subjectivity, and figurative frames changed over time. *Frontiers in Communication*, 6, 651997.



Thank you for your attention ~

