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

Winnie Huiheng Zeng[1], Menghan Jiang[2], Jieyu Chen[3], Chu-Ren Huang[4], & Kathleen Ahrens[5]

The Hong Kong Polytechnic University[1,4,5]
Shenzhen MSU-BIT University[2]
South China Agricultural University[3]

JWLLP31, 16 Dec 2022, Kyung Hee University

Introduction

- 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**)

Corpus Creation

- 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		℃
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		［①Ｄ］
上面		一来		﹤．］

- Topic number selection: N=2, 4
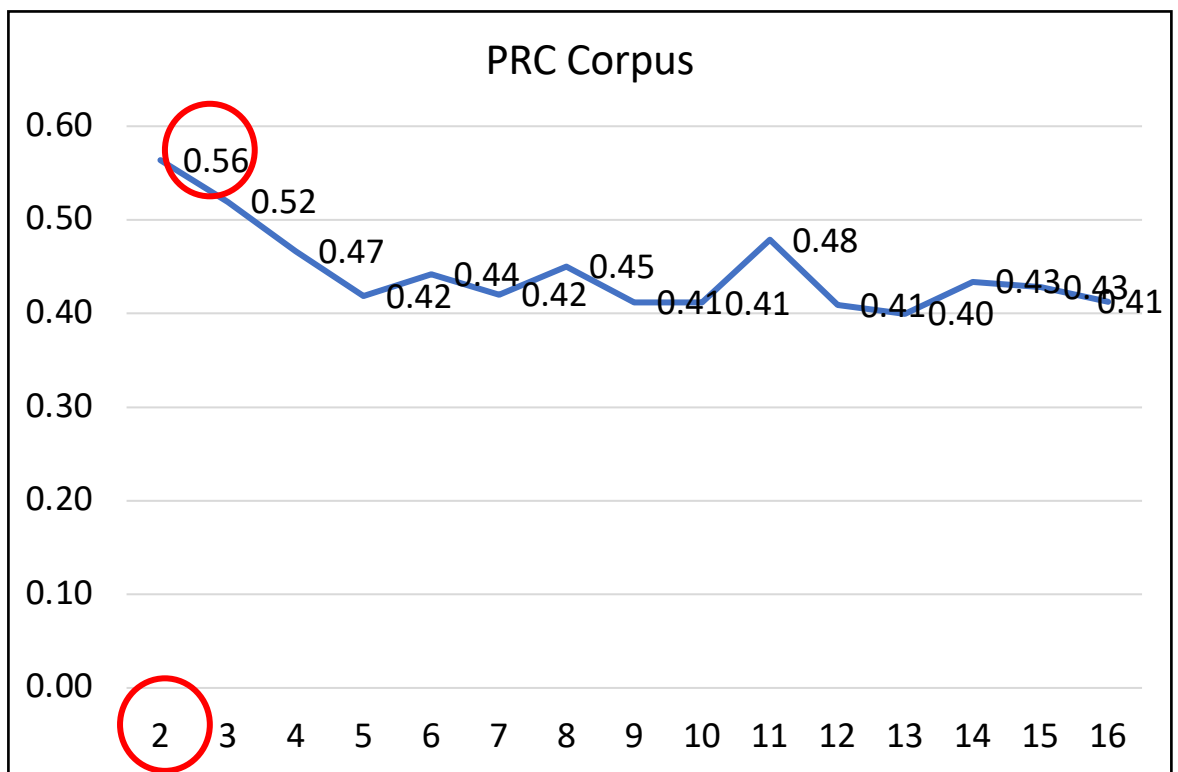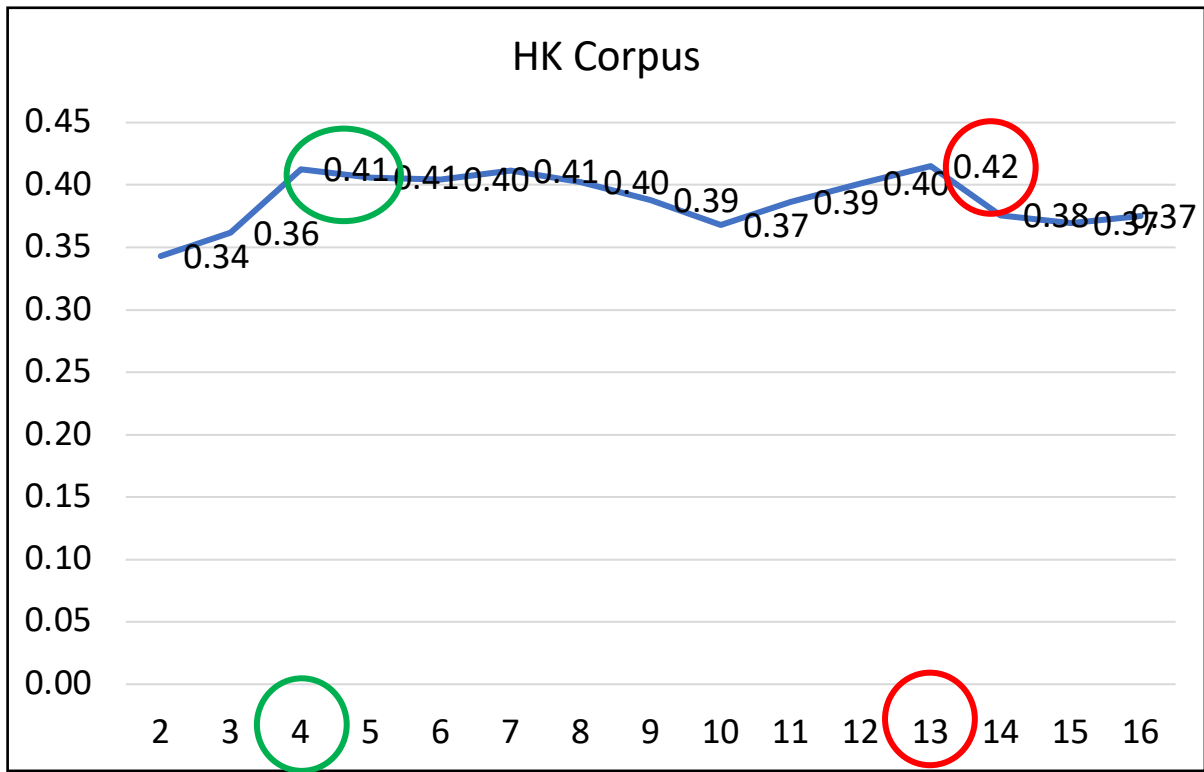
Figure. Coherence scores for the topic numbers

HK Corpus (Top 500 words)

ML Corpus (Top 500 words)

Figure. Word cloud of the most common words in the HK and ML corpora of government news on Covid-19

HK Corpus

ML Corpus

HK Corpus

ML Corpus

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)
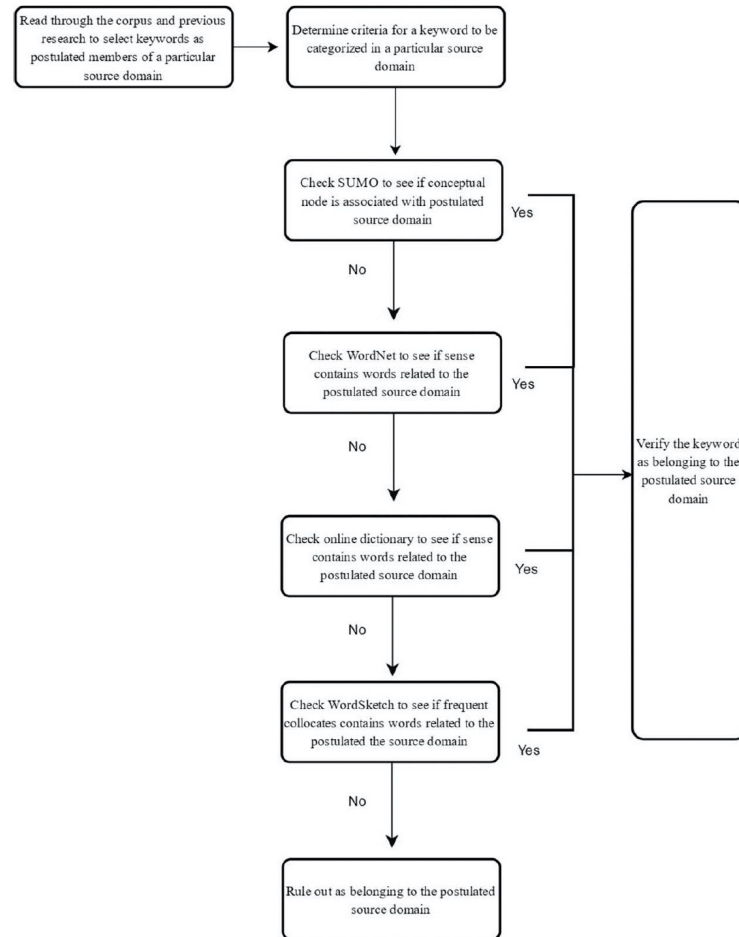
No.	WAR Keywords	Frequency	No.	WAR Keywords	Frequency
1	防控	11219	26	战场	31
2	抗疫	1135	27	暴力	25
3	部署	959	28	战疫	20
4	一线	705	29	敌人	19
5	阻击战	364	30	抵制	17
6	战胜	328	31	战争	16
7	斗争	324	32	堡垒	16
8	打赢	315	33	战役	14
9	战略	206	34	阵地	13
10	攻坚	191	35	战果	13
11	战	176	36	主战场	13
12	防	137	37	保卫	11
13	胜利	135	38	歼灭战	11
14	和平	118	39	抗争	9
15	防线	118	40	对垒	9
16	攻关	95	41	攻击	7
17	奋战	80	42	战士	7
18	打击	79	43	打仗	6
19	保卫战	67	44	战术	4
20	抗	66	45	打败	3
21	战斗	66	46	防守	3
22	军队	66	47	武装	2
23	第一线	64	48	伤亡	2
24	威胁	56	49	击败	1
25	前线	40	50	守卫	1
				Total	**17382**

- Sentiment analysis of WAR metaphorical framing

- ## 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?

References

- 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 ~