School of Social Sciences and Philosophy Assignment Submission Form

Student Names:	**Zhang Tianxin** **Darragh Kane O'Toole** **Brendan Walsh**
Student ID Number:	21352232 18431135 22334595
Programme Title:	P.Grad.Dip Applied Social Data Science
Module Title:	POP77022: Quantitative Text Analysis for Social Scientists
Assessment Title:	Republic of China Decision-Making During WWII: Which Military Issues Concerned the Wartime Chinese Government the Most?
Topic:	Republic China's military decision-making between 1937 and 1945
Lecturer (s):	Dr. Constantine Boussalis
Date Submitted	2023.03.12

We have read and we understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at: http://www.tcd.ie/calendar

I have also completed the Online Tutorial on avoiding plagiarism 'Ready, Steady, Write', located at: http://tcd-ie.libguides.com/plagiarism/ready-steady-write

Signed:

张天欣

DKOᵀ

Brand Walsh

Date: 12ᵗʰ of March, 2023

Republic of China Decision-Making During WWII: Which Military Issues Concerned the Wartime Chinese Government the Most?

Zhang Tianxin, Darragh Kane O'Toole, & Brendan Walsh

March 12, 2023

Abstract

We study which issues most occupied and concerned the military leadership of Republican China (中华民国), 1912-1949, between the Second Sino-Japanese War of 1937-1945 by exploring archival documents. Our research is primarily exploratory and descriptive. . Moreover, in our intensive reading, we've yet to encounter quantitative textual analysis covering the topic of wartime China nor the military brass in that period, thus making our work here original research.

Because we don't assume which topics and terms are most or least prevalent, we employ Structured Topic Modelling (STM), an unsupervised learning method to decipher the most common terms and topics concerning military leadership. Because our research is exploratory, we didn't predict we'd find definitive answers nor shed large amounts of light, which ultimately was the case. However, we're opening a black box. We hope our research can spark interest and further study on this critical period in not just Chinese politics, society, and history, but globally considering the Japanese imperial ambitions at that time, the existential crisis that was World War II, and China re-emerging as a global player in the past three decades.

Introduction

Upwards of 24 million people died in China during World War II. The images of Nanjing evoked by victim interviews; the Christian Bale film, The Flowers of War (金陵十三钗); and Iris Chang's magnum opus, *The Rape of Nanking*[1] – while horrific – don't come close to the whole picture of wartime China. With this research, we seek to examine decision processes of senior Republican Chinese military officials during World War II, 1937-1945.

The Chinese Civil War (国共内战)[2] effectively went on hiatus once the Japanese imperial threat emerged in China following the Mukden Incident in September 1937 from which Japan's forces in China, the China Expeditionary Army (支那派遣) launched a full-scale invasion of China. Following the Northern Expedition between 1926-1928 (北伐战争) led by Generalissimo Chiang Kai-shek (蒋介石), the Guomindang (国民党) formed the Nationalist government (中华民国国民政府) in China, making it effectively a one-party state. Because of this and because we can't access Chinese communist military archives from this period, we use the following research question to understand how the Chinese military brass led by the nationalists behaved during this period:

Which Military Issues Concerned the Wartime Chinese Government the Most?

Our exploratory and descriptive research can tell us much about how this eight-year period fits into Chinese and global history, specifically how the unequal Treaty of Shimonoseki[3] (Japanese: 下条, Mandarin: 马关条约)[4] interfluenced Chinese military behaviour towards invading Japanese forces. Precisely, what happened amongst the republican Chinese hierarchy during the Mukden (modern-day Shenyang(沈阳)) Incident, the Rape of Nanjing(南京大屠杀), the Battle of Shanghai (淞沪会战), the Battle of Changde (常德会战). In the spirit of Chang's question *"What keeps certain events in history and assigns the rest to oblivion?"*,[5] our research provides a different angle on Chinese Studies and history. We're unaware of studies similar to ours, i.e., quantitative text analysis of Chinese military leadership during the Second Sino-Japanese War. Put otherwise, to the best of our knowledge, our research is original.

---

[1]Chang, Iris. *The Rape of Nanking: The Forgotten Holocaust of World War II.* New York, NY: Basic Books. 2004

[2]The final phase of this (1945-1949) is also called the War of Liberation (解放战争) amongst Chinese communists

[3]signed the 17th of April, 1895

[4]This treaty ended the first Sino-Japanese War, 1894-1895 in which Japan annexed the island of Formosa/Taiwan and the Pescadores/Penghu Islands (Article 2), forced China to pay harsh reparations akin to the later Treaty of Versailles(Article 4), and provided Japan access to several large Chinese cities and ports (Article 6)

[5]Chang, 2004, Page 200

A note on Chinese placenames and terminology

We use the Hanyu Pinyin system (汉语拼音) of romanisation rather than the Postal Romanisation system (邮政式拼音) when writing Chinese placenames and terms. In many Chinese history texts (and strangely some modern writing on China), the Postal system is used – this means Beijing is Peking, Chongqing is Chungking, Tianjin is Tientsin, and Nanjing is Nanking. Hence, where applicable, we write these terms in Hanyu Pinyin with footnotes delineating the antiquated names. Furthermore – as you may have noticed by now – we translate most Chinese terms and placenames we use into simplified Mandarin in brackets beside the English terms.

Reviewing the Literature

We're unaware of studies similar to ours, i.e., quantitative text analysis of Chinese military leadership during the Second Sino-Japanese War. While quantitative text analysis has grown exponentially in popularity and power, it's odd the Chinese military during this period hasn't been studied more, considering in the five years afterwards, the Guomindang (国民党) was roundly defeated by the CCP (中国共产党), propelling Communism to major advances in the Third World, according to Westad.[6]

As we've said throughout this paper, we won't find 'correct' answers – after all, according to Grimmer & Stewart, all language models are wrong, but some are useful [7] – but we can garner insight into this critical period in Chinese history by examining Chinese military texts from this period. Because we've not encountered any research applying similar methodology to our to this topic, we can't write a literature review directly, if you will. We drew inspiration from our interest in the period, reading broader East Asian history, and three articles using similar methodology to ours.

The first of these articles by Fridlund, Oiva, & Paju examines the zeitgeist surrounding humanism in Germany between 1829 and 1850.[8] To do this, they employed a topic modelling method on 100 texts from the German language press.[9] A key element we noted in this study was the authors inferred from the topic model because of how the topics manifested. This advised us to be cautious in our analysis, tentative in our conclusions, and – above all – to hone our domain-knowledge to parse and decipher our results more intelligently.

---

[6] Westad, Odd Arne. The Global Cold War. 11th printing. Cambridge: Cambridge University Press. 2015 [Originally published in 2005], Page 65.

[7] Grimmer, Justin & Brandon M. Stewart. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts". *Political Analysis* 21, No.3 (2013): Page 269.

[8] Fridlund, Mats, Mila Oiva, & Petri Paju. *Digital Histories: Emergent Approaches within the New Digital History.* 1st edition. Helsinki: Helsinki University Press. 2020. Pages 259-278.

[9] Remember, Germany wasn't a unified political entity at this time – hence, German language, not German or Germany.

The second of these articles by Spirling examines how settlers treated Native Americans.[10] Spirling does this by quantitatively analysing almost 600 treaties and related documents between settlers and Native Americans. Like Fridlund, Oiya, & Paju, Spirling must infer from his analysis using his eminent domain knowledge. His results reveal much about settler-native interaction and treatment. For our purposes, Spirling's research offered us a template and proffered the benefits of using historical documents – Chinese military documents, in our case – with emergent technology to provide previously arduous or even impossible insights.

The third and final of these article comes closer to our research. Mitter and Moore examine documents describing the experience and memory of China in the anti-Japanese war (抗日战争), 1937-1945.[11] Mitter & Moore identified several relative themes related to power transferring from the Guomindang to the CCP, mentioning a wide variety of topics such as American influence in the civil war, and CCP propaganda. This article illustrates how important this type of research – namely analysing older documents – is for historical research, such as ours.

Describing our Corpus of the Republic of China's Military Leadership, 1937-1945

To compile our corpus, we selected four volumes from the (PR China) state-sponsored Second Historical Archive of China (中华第二历史档案馆编)[12]. These archival documents contain multivarious document types, including orders, instructions, reports, and correspondence within the Republic of China's military leadership during the Second Sino-Japanese War, 1937-1945. This archive collection is at present the most comprehensive collection of Republic of China military archives opened so far. We determine that because these documents originated from or were presented to Chinese military brass, military leaders were cognisant of and cared about these issues.

We chose this corpus so we could illuminate the usually opaque wartime decision-making processes. As we mentioned above, we're unaware of other quantitative textual research of the Republic of China's military leadership during this existential moment of Japanese imperialism.

---

[10]Spirling, Arthur. "U.S. Treaty Making with American Indians: Institutional Change and Relative Power, 1784-1911". *American Journal of Political Science* 56, No.1 (2012): Pages 84-97.

[11]Mitter, Rana & Aaron William Moore. "China in World War II, 1937-1945: Experience, Memory, and Legacy". *Modern Asian Studies* 45, No.2 (2011): Pages 225-240.

[12]Second Historical Archive of China (1993), Compilation of Historical Archival Data of the Republic of China, Edit. 5, Vols. 1, 2, 3, & 4: Military, Jiangsu: Jiangsu Ancient Books Publishing House.

Methodology

The war-time decision-making process is usually opaque. Nonetheless, archives in China opened in the past two decades allow us to open lift the lid on black box that is Chinese military decision-making. With quantitative methods, we can now process voluminous archives, and conduct a larger-scale analysis than traditional archive studies.

Because we don't assume which topics and terms are most or least prevalent, we employ Structured Topic Modelling (STM) – an unsupervised learning method – in R. If a topic is prevalent in our corpus, it may indicate the Chinese military leadership was more concerned or focussing more on a specific issue. We initially used 35 topic clusters, but found too many overlapping topics. Therefore, we optimised and reduced the number of clusters to 15, providing more sensible analysis and results.

Limitations

Ultimately, we were limited by the R's Mandarin textual analysis tools, which is unfortunately out of our control. For example, useful collocations weren't emerging the way they would if the Mandarin tools were as good as other languages. Mandarin is highly complex, making accurately capturing these nuances and meaning difficult for automated text analysis tools. As we found in the study, the Mandarin quantitative text analysis tools are still underdeveloped. Specifically, the pre-processing tools are suboptimal. For instance, the Mandarin collocation function didn't identify some common collocations like 敌军(Enemy Force), 日军(Japanese Army), 我军(Our Army), 军部(Army HQ), 师部(Division HQ), 我部(Our Unit), potentially biasing our results. Resultantly, the Mandarin language's complext may limit the accuracy and reliability of the quantitative text analysis; we're yet unable to overcome these algorithmic limitations.

While our Mandarin text analysis project here can potentially provide valuable insights into Chinese history and culture, it's incumbent on us as researchers to recognise these limitations, and cautiously interpret our results. Further research and validating using larger and more diverse datasets is necessary to establish reliable and robust models. Moreover, we can't ignore potential biases in sources documents (archives, in our case) when interpreting our results. Perhaps, more quantitative research of Mandarin-language corpora may improve Mandarin quantitative tools.

Summarising our Results

After running the STM, we ascertained the 15 topic clusters. After dropping Topic 4, Topic 5, Topic 11, Topic 13, we got 11 useful topics. We ascertained 15 topic clusters from our STM model. We determined only 11 of these topics were informative, excluding Topic 4, Topic 5, Topic 11, and Topic 13. These four topics were incoherent because they didn't cover

times, places, people, or the battle units. Our most prevalent topics in descending order are:

- **Topic 1: Doctrine and Organisation**. Top words: Should, Army, Unit, (dropped words: four, three).
- **Topic 2: Southern China War Zone (Zhejiang Province, Hubei Province, Jiangxi Province)**. Top words: enemy, army, division, mountain.
- **Topic 3: Northern China War Zone, (Beiping (Today's Beijing), Shandong Province)**. Top words: Army, Enemy, division, unit).
- **Topic 6: Battlefield of Hunan, Yunnan**. Top words: Enemy, Army, Division, Main Force (Dropped words: 88r, three).
- **Topic 7: Doctrine and Organisation**. Top words: Should, Army, Unit (Dropped words: four, three).
- **Topic 8: The Yunnan-Burma War Zone**. Top words: Sandong (located in Yunan Province), Enemy, Army, battle position.
- **Topic 9: The Yunnan-Burma Battlefield**. Top words: Retreating to South-Eastern Region, main force, South, West.
- **Topic 10: Yunnan-Myanmar War Zone**. Top words: Dongjie (A village in Yunnan), Enemy, Army, Battle Station.
- **Topic 12: Yunnan-Myanmar War Zone**. Top Words: Yichuan (A county in Yunnan Province), battle station, South, city.
- **Topic 14: Yunnan-Myanmar War Zone**. Top words: Hekou (A Town in Yunnan Province), enemy, army, battle position.
- **Topic 15: Yunnan-Myanmar War Zone and Southern China War Zone**. Top Words: Enemy, Army, Division, main force (Dropped words: Three, Four, nearby).

Based on our findings, we can conclude the most prevalent issues concerning the military leadership of the Republic of China were mainly battlefield scenarios, and the doctrine and organisation of the army. Among these battlefields, the most prevalent were in the Southeast China region (Zhejiang, Fujian, and Guangdong), the country's wealthiest region. The second most prevalent battlefield in our corpus was the Burma-Yunnan battlefield; this was the only land channel since 1940 through which Allied aid was sent to the Chinese.

Before war between China and Japan broke out, the Chinese military was highly decentralised; it was divided between the Central Army (in Nanjing) and the local armies. Chiang Kai-shek and his generals directly commanded the rear, whereas the warlords – who obeyed Chiang – commanded the local armies. The organisation and doctrine varied across different forces. Unifying them into one commandership was a crucial task for the central military leadership after the Sino-Japanese War broke out in September 1937.

Notably missing in our findings are issues like supplies, logistics, mobilisation, training, discipline, morale, and soldiers' welfare (such as health, wages, and living conditions), suggesting they weren't priorities for the Chinese military command during this period. From these findings, we can theorise the Chinese military leadership was indifferent to these issues, potentially explaining the Chinese military's poor performance during the war.

Discussion

Unless you're a keen reader of modern Chinese history, you might have missed a glaring omission in our results: no mention of Chongqing. (重庆)[13] This seems strange considering it was the provisional capital city of the Republican government once Nanjing was sacked, it was also, according to Bix, always the *"main target of* (Japanese) *air attack.*[14] We speculate this is because Chongqing wasn't the frontline – i.e., no ground battle happened in Chongqing. However, Nanjing emerges often because it was the capital.

Worth noting here is Grimmer & Stewart's argument that language is complex, and automated content analysis (quantitative text analysis) can never replace careful and close reading of text, but rather amplify and augment careful reading and thoughtful analysis.[15] However, we identify an oft undiscussed benefit to quantitative text analysis: sparking interest. We certainly don't paint a comprehensive picture of wartime decision-making in the Republic of China, but we illustrated some fascinating and instructive themes; our findings have whetted our interest, and can do likewise for other researcher by identifying consequential topics – i.e., signposting.

Our findings are prove consequential in that we discover themes of indifference to fundamental military issues, such as living conditions, wages, and even morale amongst Chinese military leadership. Contrarily, our findings demonstrate the Chinese military brass was more concerned with organisational and doctrinal issues – i.e., military theory over practice. Needless to say, the one serving soldier amongst us tore his hair out at these findings; after all, a military marches on its stomach. We can tentatively say this at least contributed to poor Chinese military performance against the Japanese, a much better trained and armed military. Why the Chinese ultimately won is beyond our scope, however.

Conclusion

Following the Northern Expedition between 1926-1928 (北伐战争) led by Generalissimo Chiang Kai-shek (蒋介石), China was effectively a one-party state, led by the Guomindang (国民党). Deciphering its military decision-making during an eight-year long existential moment is a worthwhile and belated process. With this research, we've lifted the lid on the black box of Republican military decision-making. Our here discoveries are very enlightening. We can tentatively conclude some reasons for the Chinese military performing so poorly against the Japanese. Our findings demonstrate the Chinese military leadership were mostly concerned with doctrinal and theoretical issues, rather than more substantial, contemporary, and pressing issues, such as morale, rather amazingly.

---

[13]Chongqing used to be labelled as Chungking

[14]Bix, Herbert P. Hirohito and the Making of Modern Japan. 2nd edition. London: Harper Perennial. 2016 [Originally published in 2000]. Page 347.

[15]Grimmer & Stewart. "Text as Data". Page 268.

While unfortunate that we can't access Chinese communist military archives from this period to complete the picture, or compare with the Republic military, accessing these archives has been worthwhile. Future research here will include quantitively analysing Japanese military documents from this time. If communist archives open, researchers should include them too. What a tremendous piece of untapped research. With our pioneering exploratory and descriptive research, we have some of the first clues from the horse's mouth as to what the warring parties were thinking during Second Sino-Japanese War, 1937-1945.

Group Participation

We met twice in February. At our first meeting, we planned and structured our project, and provisionally divided the work. We agreed Darragh would do the literature review and historical background; Tianxin would create and describe our corpus and methodology; and Brendan would do the abstract, most of the introduction, summary, and compile the final document. At our second meeting two weeks later, we updated each other on each of our progress, and delineated a time frame.

We met three times in March to share our progress and work to date. Then, we began creating a machine readable text corpus (Tianxin did most of this). Once we'd completed this, we met on the 8th of March to discuss and interpret our results. Two days later, we finalised our finding and constructively critiqued each others' work. We met for the last time on the 12th of March to review and subsequently submit the completed document.

Breakdown of work: - Abstract: Brendan

– Introduction: Brendan and Darragh

- Literature review: Darragh

– Corpus: Tianxin

– Methodology: Tianxin and Brendan

- Summary and discussion: All members had contributions from All members

– Limitations: Darragh and Tianxin

– Programming: Mostly Tianxin, with some by Darragh and Brendan

– Compiling the final document: Brendan