ABSTRACT

History textbooks provide a lens through which students view the nation's past. Due to their political significance, government - especially that of an authoritarian regime - has an incentive to use a more subjective narrative in history textbooks. This paper tests this claim by proposing four metrics to measure subjectivity of history textbooks used in Mainland China, Hong Kong and Taiwan. They are the adjective ratio, the ratio of positive to negative phrase, ratio of Chinese Communist Party (CCP) to Kuomintang (KMT) entities mentions and word embedding that measures distance between important political entities and surrounding adjectives. We find evidence that history textbooks tend to emphasize the government's point of view, but textbooks in Mainland China exhibit stronger degree of subjectivity by both the adjective and positive-to-negative phrase ratio. Furthermore, the ratio of CCP-to KMT entities is higher, and there is a stronger association between CCP entities and positive words in Mainland textbooks. Finally, by comparing textbooks before and after curriculum reform, we find that the positive-to-negative phrase ratio in post-1949 history registered a considerable increase over time in Mainland.

Keyword: History Textbooks, Authoritarian Regimes, Text Subjectivity, Part of Speech

Word Count: 7745

Correspondent Author:
Onyi Lam
onyilam@gmail.com
1-858-337-3573

1 Introduction

History textbooks provide a lens through which students view their nation's past. In the absence of political influences, history textbooks should not be biased toward or against any particular political actor, instead offering an evenhanded and relatively more objective account of the past. In reality, however, governments can use history textbooks "as ideological tools to promote a certain belief system and legitimize an established political and social order." (Apple and Christian-Smith 1991). Authoritarian regimes, in particular, have a strong incentive to homogenize political views of their citizens so that they do not wish to rebel against the regime. History textbooks in authoritarian regimes are arguably more vulnerable to government manipulation for this reason.

This paper measures the degree of subjectivity in history textbooks by using insights and tools developed in the field of natural language processing. Doing so also allows us to illustrate a less-studied linguistic channel through which persuasive communication manifests. Subjectivity in natural language processing refers to aspects of language used to express opinions and evaluations (Banfield 1982; Wiebe 1994). Subjective remarks can come in a variety of forms, including opinions, rants, allegations, accusations, suspicions, and speculation (Riloff et al. 2013). Though subjective remarks and slant can have similar meanings in some contexts, in this paper, we consider subjective remarks as text that emphasizes the writer's views and feelings and slant as factual but incomplete account of a given topic[1]. In other words, the main difference between slant and subjective remark is that subjective text contains more opinionated statements and less factual information such as numbers and details of the state. It is worth noting that by our definition, objective text, in absence of the author's own opinions and evaluations, can be slanted.

Compared to other information channels such as the mass media, history textbooks are arguably more formative in determining one's support to the regime and easier for the authority to manipulate. There are at least two reasons. First, unlike the mass media, history textbooks are mandatory in the curriculum for most students. Students are also highly incentivised to learn the material in order to do well in exams. Second, textbooks are typically read at a younger age than newspapers. Studies have shown that individuals' values and beliefs are more susceptible

[1]Yolanda Xue provided excellent research assistance.

[1]This definition of "slant" is consistent with that in Gentzkow and Shapiro 2010.

to outside influence at a younger age[2]. These important characteristics of history textbooks mean that understanding how history is being portrayed can help us understand the divergence in political views between citizens residing in different regions with shared history.

We study high school history textbooks used in mainland China, Hong Kong and Taiwan. Comparing textbooks used in these three regions with shared history allows us to contrast the attempt of governments of different political systems to influence the citizens' view on the Chinese Communist regime. To put the degree of subjectivity in perspective, we also include in our quantitative analysis a translated copy of "Search of Modern China" (SMC) - a book on modern Chinese history written by British-American scholar Jonathan Spence. SMC is frequently assigned as a reference book in Chinese History courses taught in universities in the United States. In addition to the modern history textbooks, we also analyze the textbooks used prior to major textbook reforms in the mainland and Taiwan to study the change in subjectivity due to change in government policy.

Our focus on history textbooks is not new. In particular, history textbooks in East Asia have been a source of international tension and domestic controversies. Previous studies have noted the divergence of historical account of World War II in textbooks used in East Asia (Sneider 2013). Despite Beijing and Seoul's criticism on Japanese nationalists' effort in whitewashing Japan's atrocities in World War II in some controversial textbooks, mainstream Japanese textbooks are found to be relatively devoid of overt attempts to promote patriotism. In contrary, official mainland Chinese textbooks are found to promote a nationalistic view of the past. China observers have also noted that Beijing uses "history education as an instrument for the glorification of the party, for the consolidation of the PRC's national identity, and for the justification of the political system of the CCP's one party rule." (Wang 2008) Controversy over history textbooks also took place in Taiwan where critics of a proposed textbook revision in 2015 said the changes were "China-centric" and denied the island its own perspective. Outside of East Asia, researchers have found that significant differences exist between the Israeli and Palestinian textbooks in which what was positive on one side could be negative on the other side

[2]Studies have suggested that political views are influenced most in early years. The most relevant theory in this respect, the impressionable years hypothesis, states that core attitudes, beliefs, and values crystallize during a period of great mental plasticity in early adulthood (the so-called impressionable years) and remain largely unaltered thereafter.

(Jala 2003) . These difference were said to breed hatred and contribute to ongoing conflicts[3].

While there exists several empirical studies that measure slant in mass media (Gentzkow and Shapiro 2010; Groseclose and Milyo 2005), none has conducted a quantitative analysis on education content. Previous studies on history textbooks (Cantoni et al. 2016; Sneider 2013; Ye 2016[4]) rely on human interpretation which inevitably involves a certain degree of subjective judgment. The quantitative approach used in this paper can mitigate the potential human bias in interpretation. By focusing on text subjectivity rather than slant, we also bypass the need to find a language benchmark that is required for measuring slant. For example, the current state-of-the-art method used to measure media slant in the U.S. requires the researcher to know the language used by Democrats and Republicans in order for the algorithm to "learn" the partisan rhetoric. These benchmark is often unavailable in authoritarian regimes or for text-of-interest that does not have readily comparable references such as history textbooks in this study and jurisdictional statement. In addition, some of these approaches require a labeled training set which is costly and time-intensive to prepare as they involve hiring many coders to judge the sentiment implied in the text[5]. Although this paper focuses on the cross-region difference in subjectivity, the approach is also fruitful in analyzing changes over time.

We offer a series of empirical tests for the hypothesis that restrictive government control is associated with a more subjective interpretation of history. First, we count how often adjectives and four-character idioms (Chengyu) are used relative to the number of total characters[6]. Four-character idioms are unique in East Asian languages and are often aphoristic. The motivation comes from the intuitive observation that provocative and inflammatory language is often aided by use of adjectives. Adjectives are usually difficult to verify and quantify (what constitutes a war "brutal"?), which also makes it difficult for the information receiver to dispute with the information sender. As an informal but extreme example of this tactic, propaganda from North Korea often involves use of superlatives and hyperbole. The Korean Central News Agency - the

[3]http://www.economist.com/blogs/pomegranate/2013/02/israeli-and-palestinian-textbooks

[4] The study examines if there are opposing historical accounts in mainland and Taiwan textbooks.

[5]For example, Qin at el 2017 manually code whether and how the posts cover a particular type of event. Theocharis et al. 2017 manually code whether the tweet contains uncivil languages.

[6]Existing statistical packages such as TextBlob and Usent also calculate subjectivity score. Both methods involve a manually prepared lexicon and is less transparent. In addition, TextBlob does not account for differences in text length between the two pieces.

state-controlled office in charge of communicating with the outside world- regularly promised 'strikes without mercy by the steadfast and ever-vigilant People's Army,' and referred the South Korean government as 'wicked running dogs' and 'loud-mouthed gangsters'. The sentiment analysis literature has also found that the presence of adjectives is strongly predictive on whether an article belongs to an opinion piece as opposed to a news piece (Hatzivassiloglou and Wiebe 1999; Bruce et al. 1999; Bruce and Wiebe 2000).

Second, we count the number of positive and negative phrases used and calculate their ratio. Using this measure allows us to test a more specific hypothesis that the ruling regime has an incentive to emphasize policy success and underplay the mistakes. We expect that the positive-to-negative ratio on specific historical episodes and time period to vary. This binary classification of words is the crudest form of sentiment analysis but in absence of reliable sentiment dictionary that captures more granular emotion classification, the frequencies of positive and negative phrases give a reasonable assessment of the writer's opinion on a given topic. Relatedly, several studies in Psychology have found that the use of pleasant and unpleasant adjectives could capture readers' attention, and are easier to remember than neutral adjectives (Herbert et al. 2007; Kissler et al. 2009).

Third, we count the number of times six entities of historical and partisan significance are mentioned. Intuitively, the more an entity is mentioned, the more details of the entity is provided. Previous research has found that relative number of tweets mentioning a party or candidate is a good predictor for the support of that party in an election (Tumasjan et al. 2010). The frequency of these entities could imply support for one partisan perspective over another. We consider three Chinese Communist Party (CCP) (Mao, Marx, and CCP) and Kuomintang (KMT) entities (Chiang, Sun yut-shan, and KMT) to assess the relative emphasis on partisan viewpoints. Mao Zedong - the central figure of the CCP from its founding to his death - is critical to the CCP's authority and presumably mentioned more in the mainland textbooks. Similarly, Chiang kai-shek - the Director-General of the KMT until his death in 1975 - is expected to be mentioned more in the Taiwan textbooks. To account for difference in book length, we calculate the ratio of mentions of CCP to KMT entities.

The first three methods simply count the number of times a specific phrase appears. To capture the ways important political entities are being described in more granularity, we use word embedding to measure distance between these political entities and surrounding adjectives.

Word embeddings use numeric vectors to represent words, and this representation allows us to measure the distance between the political entities and other adjectives. The method is originally introduced by Bengio et al. 2003, and the main idea is that words that tend to appear in similar contexts are likely to be related. To learn these word embedding, we implement Word2vec (Mikolov et al. 2012), a state-of-art method in the field of natural language processing, in Python.

We find that adjective ratio, positive-to-negative phrase ratio and the ratio of CCP to KMT entities mentioned are both higher in the mainland textbooks. Mainland textbooks also have a higher positive-to-negative phrase ratio in post-1949 history (the People's Republic of China was established in 1949) relative to the pre-1949 history. In addition, we find that the CCP entities exhibit a stronger association with positive words in mainland textbooks. Finally, by comparing textbooks before and after curriculum reform, we find that the positive-to-negative phrase ratio in post-1949 history registered a considerable increase over time in mainland. In contrary, textbooks in Taiwan have become less subjective over time by the same measure. Overall, the result is consistent with the hypothesis that the history textbooks in a more restrictive political environment exhibit stronger degree of subjectivity.

One way to interpret our findings is that authoritarian regimes use messages with emotional connotation to trigger desired cognitive association with the regime. Psychologists and marketeers have long noted that messages, despite devoid of useful information, can improve the audience's assessment of the persuader's issue or product by triggering associative thinking or associative feeling via connotation of the message (Tversky and Kahneman 1982). This association operates through a cognitive process that takes advantage of existing analogies that an audience may already have in mind, or by creating the relevant analogy by advertising attributes associated with that desired analogy (Brader 2005; Mullainathan et al. 2008). By using more positive words in association with the regime and in a specific time period, the mainland Chinese government can potentially promote goodwill towards the government among the students.

This paper contributes to the rapidly growing literature in social science that uses data-intensive text analysis techniques to elicit political content (for example: Gentzkow and Shapiro 2010, Tumasjan et al. 2010; Beauchamp 2016; Roberts et al. 2016; Theocharis et al. 2017). By pointing out the role of subjective sentiment and quantitatively assessing its extent in education

content, this paper also expands the literature on the role of emotions in political communication. Much of this literature aims to identify the persuasive effect of emotional cues in political campaigns on voters(Brader 2005; Lau et al. 2007), and this paper both expands the scope of content and contributes to the methodology. Finally, this paper also shows that authoritarian regime can use emotional words as a strategy to persuade students to think positively on the regime. Recent work in political economy has also analyzed authoritarian regimes' incentive to homogenize the population by shaping the views of their citizens (Alesina and Reich 2013; Cantoni et al. 2017), to which we provide indirect evidence.

2 History Textbooks in Mainland China, Hong Kong and Taiwan

We sought to select high school textbooks that are most widely used in the respective regions. Students exposed to these textbooks are generally between the age of 15 and 17. Our modern textbook (as of use in 2016) sample consists of 3 versions from the Mainland, 2 from Hong Kong and 3 from Taiwan. We purchased the physical copies through an online shopping platform *Buyippee*. An undergraduate research assistant then digitalized the textbooks into text files for computation by both voice typing and manual entry. We ignore the ancient Chinese history and use the text after the Opium War (A.D. 1842) only. In addition, we ignore all text in captions and appendices.

There are three approved versions in Mainland, and they are: Renmin, Renjiao and Yuelu. We acquired all three of them. The publishers are either state-owned or a subsidiary of state-owned enterprise. Provinces and cities can choose one of the three approved versions for their high schools [7]. Renjiao is the most popular of the three versions with the largest number of adopting regions. The organization of the content is very similar across the versions: there are three books in each version and each book has a different emphasis. The first book presents the historical events in a generally chronological order. The second book focuses on the economic development in the corresponding time period, whereas the third book focuses on cultural and technological progresses. In some sections of the books, information not directly related to

[7] http://godfreyxu.github.io/2013/01/30/high-school-history-textbook-version-of-provinces-and-cities.html

China is also presented. For example, the different political systems used in other parts of the world are discussed, and important western literature and scientific advancement of the concurrent time period are also mentioned. We exclude these discussions in our text analysis to ensure content is comparable across regions. In addition, the curriculum is separated into a mandatory and an elective portion, and we use the mandatory portion only.

In both Hong Kong and Taiwan, the respective Education Bureau issues general guidance for private publishers, who are responsible for writing the textbooks. The schools have the discretion to decide which publisher's version to use. In Taiwan, there are seven popular versions in the market, and our sample consists of three of the most widely used: Kangzi, Nane, and Lungtun. Kangzi is most popular textbook and used by 45% of the schools, followed by Nane (17%) and LungTun (16%)[8]. The curriculum is organized into four parts with a respective focus on Taiwan History, Chinese History, Ancient World History and Modern World History[9] and the events are presented in chronological order. To ensure comparability across regions, we use the part on Chinese History only. The Hong Kong textbooks sample includes two publisher versions: Manhattan Press, and the Modern Educational Research Society Limited. Both of which are among the most commonly used textbook publishers [10]. Manhattan Press separates the whole curriculum into six books while Modern Education has four. However, the content is very similar between the two.

To measure change in subjectivity due to change in government regulation, we analyzed two old textbooks that were used prior to major reforms in mainland China and Taiwan. The old Taiwan history textbook was used between 1983 and 1999 before the textbook market opened up for private publishers. Prior to the reform, the content in the history textbook was standardized by the Education Bureau. We also have a Mainland textbook version used in 2003, before a major reform took place in 2004 that aimed to increase students' patriotic spirits. In addition to the secondary school textbooks used in the three regions, we studied a translated copy of a commonly used Chinese History textbook in universities in the United States - "The

[8]The figure comes from a study that sampled 313 schools in 2009 (Mao 2013)

[9]https://www.sanmin.com.tw/learning/public/data/course

[10]There is no formal study that examines the fraction of schools using which textbook versions, but Manhattan Press, Modern Educational, Hong Kong Educational and Ling Kee are considered to be the four most widely used versions: http://www.com.cuhk.edu.hk/ubeat_past/051170/64.htm

Search for Modern China" (SMC)[11]. Since SMC is presumably free of political influences and a political objective, it serves as a reference comparison for our analysis. There are important differences between SMC and the textbooks, however. Namely, the scope of SMC is wider and the events are covered in greater depth; the length of SMC is also significantly longer than any of the high school history textbook. SMC tends to provide more cultural and social context by introducing and describing in details the lives of people who are lesser known in history whereas the narrative of the history textbooks is more consistent and unambiguous.

3 Descriptive Statistics

As part of preprocessing, we converted the text into simplified Chinese, and removed all numbers, punctuation and non-Chinese characters. Unlike English, there is no white space between words in Chinese so we need to segment the text into meaningful phrases. To do that, we used Jieba, a Python library, that is designed to segment Chinese sentences into individual phrases by finding the most probable combination based on the word frequency[12]. After segmenting the sentences, we removed the stop words from the text[13].

Table 1 provides the descriptive statistics of each version. The first column indicates the total character count in each version, which varies considerably across regions though not as much within. Textbooks in Hong Kong are the longest on average, followed by Mainland and Taiwan. Column (2) - (4) indicate the number of bigrams, trigrams and quadgrams that Jieba detected. In all versions, the number of trigrams and quadgrams is considerably less than the number of the bigrams. Importantly, The number of quadgrams exceeds the number of trigrams in all the Mainland versions while the reverse is true for Hong Kong and Taiwan versions. Quadgrams are often adjectives, and some of them are 4-character idioms, which contain a fable with a practical lesson. This provides the first suggestive evidence that Mainland textbooks contain more subjective elements[14]. The ratio of identified n-grams to book length

[11]The book was translated by a Taiwanese.

[12]We supplemented a list of individual names to the jieba dictionary to help Jieba distinguish names from other parts of the sentence.

[13]We use the list of Chinese stop words from an online resource: https://gist.github.com/dreampuf/5548203

[14]In the sentiment analysis literature, Pang et al. (2002) also find evidence that higher-order n-grams are useful features in predicting opinionated piece. They report that unigrams outperform bigrams when determining determine whether a movie review is positive or negative

is similar across all versions, which suggests that the dictionary is not prone to any potential language difference across the three regions. Column (5) and (6) show the character count of pre- and post- 1949 content. Character count of post-1949 history is smaller than that of pre-1949 in all textbooks. On average, the fraction of post-1949 character count is the lowest in the Taiwan textbooks and highest in Mainland textbooks, illustrating a stronger (weaker) emphasis on post-1949 history in Mainland (Taiwan) textbooks. The fraction of post-1949 content is similar between Hong Kong textbooks and the SMC.

3.1 Major Historical Events

We extract from the text several major historical events and classify them into pre- and post-1949 events to examine whether the treatment on individual events varies across versions [15]. Event such as the Cultural Revolution, which is widely acknowledged as result of a policy blunder by the Communist Party, is expected to receive larger differential emphasis across the region's textbooks whereas events that are not directly related to the perception of the regime such as the Foreign invasion during the Qing Dynasty should receive relatively similar emphasis.

The top panel in table 2 shows the number of characters and the length of the four major historical events prior to 1949 relative to the book length: Foreign Invasion (1842-1911), XinHai Revolution (1911), Second Sino-Japanese War (1937-45), and the Chinese Civil War (1945-1949). Foreign Invasion refers to all of the wars involving a foreign power since the Opium War but excluding the second Sino-Japanese War. Xinhai Revolution includes description on Sun yut-chan, and all immediate events leading up to the Revolution. The second Sino-Japanese War refers to the military conflict between China and Japan from 1937 to 1945. Finally, the Chinese Civil war includes all events surrounding around the conflict between the CCP and the KMT after the surrender of Japan in 1945, but not any previous conflicts. We do not include text from SMC in the event-level analysis because its writing style, which tends to intervene with multiple themes, makes it difficult to single out the portion that is relevant to the specific historical episodes. Foreign Invasion has the largest character count among the pre-1949 events in 5 of the 8 versions. The word count varies significantly even across versions within the same region. For example, Mainland's Yuelu only has 618 characters on the Sino-Japanese war

[15] Since Mainland versions have different emphases in each book, we exclude text in the books on economics and culture and technology and only use the text in the general history.

whereas Renmin has 3293 characters. By relative length, the Civil War receives the heaviest emphasis in the three Taiwan textbooks, whereas Foreign Invasion has the highest ratio in the two Hong Kong versions.

The bottom panel of table 2 prints the character count and the relative length of the three post-1949 historical episodes: Great Leap Forward (1958 - 61), Cultural Revolution (1966-76), and the Reform and Open Policy (1976 - current, encompassing the Tiananmen Square Protests in 1989). In all versions, the Reform and Open Policy receives the most characters. Both the Great Leap and the Cultural Revolution receive smaller coverage in the Mainland textbooks. On the other hand, the Cultural Revolution has the highest relative length in the two Hong Kong versions. Interestingly, Taiwan Lungtun has the highest ratio and Taiwan Nane the lowest on the coverage of Reform and Open policy, suggesting considerable variation in event emphasis within Taiwan versions. The Tiananmen Square Protests is noticeably absent in all Mainland versions but are mentioned in all Hong Kong and Taiwan versions.

4 Analysis

We construct four metrics to measure subjectivity. The first counts the number of times an adjective is used and calculate the ratio of adjectives to the total character count. Knowing the affective aspect of the opinions is also important in understanding government's influence so for the second measure, we calculate the positive-to-negative phrase ratio using a pre-existing sentiment dictionary. Third, to assess how often a figure of partisan significance is mentioned, we count the number of times important political entities including Mao Zedong and Chiang Kai-shek appear. Finally, we use word embedding to calculate distance between these political entities and other adjectives to illuminate how they are portrayed in more granularity.

4.1 Adjective Ratio

To determine whether a phrase is an adjective, we wrote a script in Python to automatically look up all the identified phrases in the pre-installed dictionary on a Mac notebook[16] . A

[16]jieba.posseg.cut in the Jieba module returns part of speech tagging but the result is not satisfying. For example, the noun 蛮夷 means savage or barbarian people. Using jieba.posseg.cut will not detect the negative connotation of this phrase because it will only return "noun" for the phrase. Ideally, we want the segmentation

phrase is considered as an adjective if the word definition contains the character 形 (English: adjective. See figure 1) and a four-character idiom if the character 成 is present. In addition to adjectives, we also examined the presence of adverbs. A phrase is considered as an adverb if the character 副 is present in the word definition (See figure 2). Table 3 lists the top three bigram and quadgram adjectives and their respective frequencies in each textbook version. 独立自主 (Act independently and of one's own initiative) is the most common quadgram for all three Mainland versions. It is worth noting that all popular quadgrams in Mainland versions have a positive connotation. 内忧外患 (Domestic trouble and foreign invasion) is one of the top five quadgrams in both Hong Kong versions. 革命 (Revolution/Revolutionary) and 经济 (Economy/Economic) are the two most popular bigrams in all textbook versions. Perhaps surprisingly, 民主 (Democracy/Democratic) is the third most commonly used bigram adjective in all Mainland textbook versions. [17].

Figure 4 and 5 plot the adjective and adverb ratio for each version. Consistent with our hypothesis, the three Mainland versions have the highest adjective ratio. On average, they are more than 1% higher than Hong Kong and Taiwan versions. In comparison, SMC has a similar adjective ratio as the Hong Kong and Taiwan versions. The adverb ratio is slightly higher in Taiwan versions relative to the Hong Kong and Mainland versions, but is otherwise similar with SMC. The variation in adverb ratio is smaller than that of the adjective ratio, with the difference between the versions with the largest and smallest ratio about 0.7%. To the extent that adverbs are more common in informal than formal writing, the smaller adverb in history textbook and a relatively lack of difference across regions could be explained by the difference in writing style that demands more formality in textbook texts. As a robustness check, we also plot the verb ratio in figure 14 in the appendix. Verb ratio in Mainland textbooks, while higher than SMC, is

algorithm to tell us that 蛮夷 is a subjective noun. While this is not currently available in jieba, using the "full" method allows us to separately look at each possible word combination to determine if any of the phrase has an adjective component. An example of jieba's full segmentation method follows. The sentence 我来到北京清华大学 can be segmented into 我/ 来到/ 北京/ 清华/ 清华大学/ 华大/ 大学 . The phrase 清华大学 is inevitably counted twice because of the different possible combination within the phrase. However, this allows us to separately examine the individual word within 清华大学 .

[17]The bigram adjectives are commonly regarded as nouns as well. This can make our result difficult to interpret. We verify that the overall pattern of our results remain unchanged after manually removing these noun/adjectives in the ratio calculations.

not very different from Taiwan versions and is lower than the Hong Kong versions.

Figure 6 and 7 plot the adjective ratio of the four pre-1949, and the three post-1949 events. There is no strong regional pattern in the pre-1949 events but among the post-1949 events, Reform and Open Policy has a noticeably higher adjective ratio in the Mainland versions, and likewise for the Great Leap Forward even though the pattern is not as strong[18].

Left hand side of figure 8 plots the adjective ratio of all text separated into pre- and post-1949. The pre-1949 adjective ratio of the Mainland textbooks are on average the highest, even though it is relatively similar to that of SMC and the Hong Kong textbooks. More strikingly, the average increase from pre- to post-1949 is much larger in Mainland textbook when compared with the other books. This provides direct evidence that the Mainland government emphasizes the positive aspects of the post-1949 history.

4.2 Text Polarity

Text polarity - whether it is positive, negative, or neutral - informs the readers on the author's affect or view. The polarity of text can also affect on the readers' cognitive response. Studies in Political Communication and Marketing have found that text with strong valence tend to be more memorable even though the effect of positive and negative messages is asymmetric[19]. To determine the phrase polarity, we assign a phrase as either positive or negative using a Chinese sentiment dictionary that is made available by researchers in Chinese natural language processing [20]. The dictionary consists of 4570 positive and 4374 negative phrases. While the list of reference phrases is not exhaustive, the phrases are not restricted to adjectives. Some of these phrases are subjective nouns such as 爱戴 (love and support) and 褒扬 (praise). Phrases not found in the dictionaries are excluded from our ratio calculation.

Table 5 reports the number of positive and negative phrases, and their ratio in each version. Mainland versions have the highest positive-to-negative phrase ratio, followed by Hong Kong

[18]Due to the low total character count in Great Leap in Mainland versions, the ratio is not the best measure of subjectivity.

[19]Lau et al. 2006 find that negative political advertising can stimulate knowledge about the campaign. In Lee et al. 2007 and Chevalier and Mayzlin 2006, the researchers found that negative product reviews can change initial attitudes and influence purchase decision

[20]The dictionaries are available at: http://www.keenage.com/html/c_bulletin_2007.htm. The dictionary has the drawback of not having many idioms.

and Taiwan. In particular, Renmin's ratio is approximately 2 times higher than SMC, which has the lowest ratio of all. On average, the positive-to-negative phrase ratio is slightly higher in Hong Kong than in Taiwan[21].

To show in which period more positive or negative phrases are used, the right hand side of figure 8 plots the pre- and post-1949 positive-to-negative phrase ratio. In the pre-1949 period, the value is similar across all textbooks with approximately 2 to 3 positive to 1 negative phrase. However, Mainland versions have the largest increase from pre- to post- 1949 with approximately 5 to 7 positive to 1 negative phrase in the post-1949 period. In contrast, Hong Kong versions have a much smaller increase, while Taiwan versions have a small decrease on average. SMC has a roughly similar ratio in pre- and post-1949. The result strongly supports the notion that the Mainland textbooks portray a more positive image of the Chinese Communist regime compared with the textbooks of the other two regions.

Figure 9 and 10 plot the positive-to-negative phrase ratio of the major historical episodes. No region-specific pattern emerges and variance is high across versions for the pre-1949 events. For the post-1949 events, the ratio is noticeably higher in all three Mainland versions in Reform and Open Policy, with a value approximately 2 times higher than the Hong Kong and three times higher than the Taiwan versions. However, the ratio is similar across the board in Cultural Revolution and Great Leap Forward.

4.3 Political Entities

The number of times a political entity is mentioned indicates the historical importance of the political entity. Furthermore, the frequency can suggest the political entity's significance specifically to the party. Specifically, "Mao Zedong", as well as "Chinese Communist Party" are expected to be more frequently cited in Mainland textbooks because of their significance to

[21]Table 8 and 9 in the appendix also show the top 5 most used positive and negative phrases. 严重 (serious) is the most common negative phrase, and is the top 5 most frequent phrases in all versions. 新 (new) is the most common positive phrase. There is significant overlap among the most frequently used positive and negative phrases in textbooks belonging to the same region. For example, the top 5 most frequently used positive phrases in the Mainland Renjiao and Yuelu versions are identical. In addition, 专制 (dictatorial) and 严重 (serious) are present in all three Mainland versions as top five most frequently used negative phrases. Three positive phrases, and four negative phrases shared the top five between the two Hong Kong versions. In Taiwan, Kangxi and Lungtun share more commonly used phrases, but less so with Nanye.

Mainland politics. In contrary, "Kuomintang" is expected to be mentioned relatively more in Taiwan textbooks after adjusting for the number of times CCP entities are mentioned.

We count the number of times six important CCP and KMT entities are mentioned in each version, and then calculate the ratio of CCP to KMT entities mention frequencies. The six entities are Mao Zedong, the Chinese Communist Party, Karl Marx, Chiang Kai-Shek[22] , Sun Yat-Sen, the Kuomintang. The first three are considered as CCP and the last three KMT entities.

Table 4 presents the number of mentions for each of the 6 individuals and entities. There is a significant difference in count for Chiang, Marx, and Mao between Mainland and Taiwan versions. All Mainland versions mention Chiang less than 15 times but more than 45 times for Mao. The opposite pattern is observed in the Taiwan versions. Likewise, Marx is mentioned less than 5 times in all Taiwan versions but at least 19 times in the Mainland versions. Hong Kong versions mention Marx 9 times on average. The other 3 entities receive a more even-handed treatment. We find that the number of mentions of CCP (average 113 times) is not very different between Mainland and Taiwan versions (average 103 times). Sun also receives relatively similar emphasis between Mainland and Taiwan versions, with about 30 mentions on average in both. Interestingly, KMT is mentioned more in the Mainland versions that in the Taiwan versions. On average, Mainland versions mention KMT 96 times while Taiwan versions 47 times only.

The ratio is presented in figure 3. Consistent with our hypothesis, Mainland versions have the three highest CCP to KMT ratio. In particular, Renmin mentions CCP and KMT entities in a 3.5 to 1 ratio, while SMC has a 2 to 1 ratio. Hong Kong and Taiwan textbooks have a ratio between 1 to 2 with Taiwan versions the lowest among the three regions on average.

This analysis provides evidence that not only Mainland textbooks, but Taiwan textbooks also have a tendency to emphasize their respective party's perspective. While both the adjective ratio and the positive-to-negative ratio provide an absolute measure of subjectivity for which the larger the value the more subjective the text scores, the interpretation of the ratio of CCP-to-KMT entities is not as straightforward because a low CCP-to-KMT entities ratio also signals sided coverage of the KMT entities[23].

[22]The Taiwan versions address Chiang Kai-Shek using his adopted name "Zhongzheng", and in many cases in the old Taiwan textbook, the honorific "Chairman Chiang" is used.

[23]It is also worthwhile to note that the CCP to KMT ratios can be interpreted as either a subjectivity or slant measure based on our definition. Opinion representing the regime or political entity can be treated as subjective content. In contrary, factual information in support of the political entity can be considered as slant Similar

4.4 Word Embedding

Counting the number of times a political entity is mentioned and calculating the ratio can mask true subjectivity because of one's tendency to mention and attack their opponents. It is possible that a low CCP-to-KMT entities ratio is a result from both heavy mentions of CCP and KMT entities but that CCP entities are praised whereas KMT entities are attacked. To understand how a political entity is being portrayed in more granularity, we study the word embedding of the political entity. In Linguistics, word embeddings aim at quantifying semantic similarities between words based on their distributional properties. The basic idea is that words with similar distributions of surrounding words have similar meanings. Mathematically, a word embedding $W : words \rightarrow \mathbb{R}^N$ is a parameterized function mapping words in some language to high-dimensional vectors. Using this framework allows us to capture the sentiment surrounding the political entities of interest by computing the distance with their surrounding adjectives in the high-dimensional embedding space trained by the history textbooks from each region.

We use Word2vec (Mikolov et al. 2013), a collection of models that are used to produce word embeddings through training of the neural network, to consider the associations between entities of interest and words with emotional connotation. One type of model within Word2vec is called Continuous Bag of Words (CBOW). It can be used to predict co-occurrence relationships using the conditional probability of observing the target word given the input context words. Context words are represented by multiple words for a given target word. For example, CBOW treats "The", "cat", "over", "the", "puddle" as context and and predicts the target word "jumped". The training goal of the CBOW model is to arrive at vector representations of words that best predict the target word. Formally the objective function is given by:

$$J_\theta = \frac{1}{T}\sum_{t=1}^{T} log\ p(w_t|w_C) \tag{1}$$

where θ represents all the variables we optimize. w_t denotes the target word, and w_C denotes the context words. t denotes the training step. Denote V as the vocabulary in the text. The conditional probability of the target word can be represented by a softmax function, which uses a neural network structure to learn the parameters:

$$p(w_t|w_C) = \frac{exp(u_t^T v_{w_c})}{\sum_{w_i \in V} exp(u_w^T v_{w_c})} \tag{2}$$

approach is used in study of media bias in the U.S.. By counting the number of times a think tank is cited in the newspaper, Groseclose and Milyo 2005 aims to capture liberal bias in the media. We made no distinction here.

u_w and and v_w are two representations of the word w. u_w comes from rows of the input to hidden weight matrix in the neural network, and v_w comes from columns of hidden to output matrix. The inner product $u_t^T v_{w_c}$ computes the log-probability of word w_c, which we normalize by the sum of the log-probabilities of all words. The goal of the algorithm is to learn the weights in the input to hidden layer, and hidden layer to the output matrix[24]. (A more detailed explanation of the CBOW model and neural network learning can be found in Rong (2014)).

To measure the similarity between words, we calculate the cosine distance between the word vectors in the high-dimensional embedding space. The cosine distance between vector A and B is defined as:

$$\frac{A \cdot B}{\|A\| \, \|B\|} \tag{3}$$

The resulting similarity ranges from -1 meaning exactly opposite, to 1 meaning exactly the same, with 0 indicating orthogonality. We can then examine the distance between important political entities and other phrases.

Rather than treating each textbook version separately, we combine individual version into a corpus for each region and train our model based on region-specific text. We do that because we want to maximize statistical power in explaining cross-region difference in language use. We train each model with an embedding dimension of 500 and with context size of 6 using a python library gensim, which has a built-in CBOW model. We then search for the closest adjectives around the important political entities in each embedding space. The result for each region is presented in the appendix. Positive adjectives such as 拨乱反正 (bring order out of chaos) and 解放思想 (liberate thoughts) are found to be closest with Mao Zedong in the Mainland textbooks, whereas more neutral phrases such as 整肃 (enforce) are among the closest adjectives in Hong Kong and Taiwan textbooks. With respect to Chiang Kai-Shek, more positive phrases such as 取得胜利 (to get victory) and 固守 (defend tenaciously) are in the list of the closest adjectives in Taiwan textbooks, and noticeably more negative phrase such as 全军覆没 (the whole army is wiped out) in Mainland textbooks.

To facilitate comparison across regions, we calculated the ratio of positive-to -negative phrase using the top 20 surrounding adjectives of the 6 political entities. Table 6 shows that Mainland versions have the highest positive-to-negative phrase ratio on all CCP entities and Sun. On the other hand, Taiwan versions have the highest positive-to-negative ratio on Chiang and

[24]This can be achieved by the gradient descent method using a random initialization.

KMT. Hong Kong versions' positive-to-negative ratio tend to fall in the middle between that of Mainland and Taiwan. With a value between 0.6 to 1.5, SMC's ratio does not fluctuate a lot across the 6 entities and tend to be more balanced. In summary, while the result is consistent with the hypothesis that content of Mainland textbooks are more likely to describe positively on the CCP entities, the analysis also reveals that Taiwan textbooks have tendency to use do the same on the KMT entities.

We can visualize each embedding space trained by the model by using the t-Distributed Stochastic Neighbor Embedding method (t-SNE) (Maaten et al. 2008). This is a dimensions reduction technique that is particularly well suited for the visualization of high-dimensional datasets such as ours. The idea is to embed high-dimensional points in low dimensions in a way that respects similarities between points[25]. By visualizing them in a 2-dimensional plane, we can illustrate the semantic distance between the political entities themselves.

Figure 13 shows the visualization of the embedding space. While the relationship between Marx, CCP and Mao is similar to that of Sun and Chiang and KMT in Mainland textbooks, the two clusters are far from each other, which illustrates that the semantic distance between the CCP and KMT entities is far. In Hong Kong, CCP and Marx are close to each other while Mao appears to be further apart. Mao is closer to the KMT, Chiang and Sun cluster. Finally in Taiwan, Mao, CPP, KMT, Chiang and Sun are clustered and Marx alone is far away. This echoes with the previous finding that Marx is scarcely mentioned in Taiwan textbooks.

5 Prior and After Reform

History textbooks in both Mainland China and Taiwan have undergone significant revision in the last two decades. Before 1996, Taiwan's history textbooks were written and published by the

[25] t-SNE starts by converting the high-dimensional Euclidean distances between data points into conditional probabilities that represent similarities. Suppose there are only two high-dimensional objects, x_1 and x_2.The conditional probability $p_{2|1}$. The goal is to learn a 2-dimensional projection y_2 and y_1 that reflects the similarities between x_1 and x_2 as well as possible. As with their high-dimensional counterparts, the similarity of the projection points can be represented by conditional probability. The observation is that if the map points y_1 and y_2 correctly model the similarity between the high-dimensional data points x_1 and x_2, the conditional probabilities $p_{j|i}$ and $q_{j|i}$ will be equal. The algorithm that minimizes the sum of Kullback-Leibler divergences over all data points using a gradient descent method with respect to y.

National Institution for Compilation and Translation under the Ministry of Education (MOE). All schools were required to use the nationally standardized history textbooks. Since 1996, the MOE has opened up the textbook market for private publishing companies. Schools have also been given the discretion to choose their textbooks. This has allowed for more varied narratives and a deviation from the official interpretation. In contrary in Mainland China, a nationwide education reform took place in 2001 with an explicit goal to instill patriotism and love for social socialism in students [26]. The reforms in these two regions represented opposite direction in changes in the intensity of government influence and therefore, we expect the respective reforms to have opposite effect on the textbook subjectivity. Namely, text subjectivity is expected to increase in Mainland textbooks and decrease for the Taiwan textbooks after the reform. We test these implications next[27].

The bottom 2 rows of table 1 print the summary statistics of the pre- and post-reform textbooks. Both the Mainland and Taiwan pre-reform textbooks have larger word count than their post-reform textbooks. The post-1949 ratio of the old Taiwan textbook is significantly smaller than that of the post-reform textbook with only one paragraph allocated to post-1949 events. The old Mainland textbook as well allocates a larger fraction of its total word count to pre-1949 history compared to the post-reform textbooks: the ratio is over 2 in the old textbook and less than 1.5 in the modern versions on average. This indicates a stronger emphasis on post-1949 history in both Mainland and Taiwan after the reform[28].

The aggregate adjective ratio increases from 0.52 to 0.55 in the Mainland textbook as shown in figure 4 while that of the pre-reform Taiwan textbook is similar to the average of the modern versions. Figure 11 plots the adjective ratio of the old textbook and the average of the

[26] Quoting from Appendix C of Cantoni et al, 2016: in the Ministry of Education's "Framework for Basic Education Reform" (2001), a new curriculum should meet the following objectives (in the order of appearance in the original document): it should reflect the times, and make students patriotic, communitarian, [and] love socialism. Students should inherit and carry forward the great traditions of the Chinese nation and its revolution; and be equipped with an awareness of the legal system under a socialist democracy. The new curriculum should promote compliance with national laws and with societal ethics, and gradually form in students a correct world view, a correct view of life, and a correct value system.

[27] We do not conduct the word embedding exercise in this section because the corpus of the pre-reform textbook is much smaller and not enough to train a neural network.

[28] With more years of recent history available, it is possible that the stronger stronger emphasis on post-1949 history in post-reform textbooks is due to more information at hand when the textbooks were written.

modern textbook versions by historical event[29]. The adjective ratio remains relatively unchanged in the four pre-1949 episodes in both Mainland and Taiwan. Among the post-1949 episodes, there is a noticeable increase for the Great Leap Forward in the Mainland textbook. Contrary to our hypothesis, we also observe a small decrease for Cultural Revolution in Mainland textbooks. The inconclusive evidence can be attributed to the smaller word count of the event in the modern Mainland textbook - Cultural Revolution's word count in the modern versions is less than half of that in the old version.

The left hand side of figure 12 plots the overall adjective ratio of the text used in pre- and post-1949 history. The increase from pre- to post-1949 in the old Taiwan textbook is larger than that of the post-reform ones, which signals a more subjective writing style in the post-1949 content in the pre-reform version. An increase of slightly larger magnitude from pre- to post-1949 is seen in the pre-reform Mainland textbook, but the slope is almost identical to that of the post-reform Mainland textbooks. The right hand side of figure 12 shows that the post-reform Mainland textbooks have a much larger increase in positive to negative phrase ratio from pre- to post-1949 period than the old Mainland textbook. This is consistent with our hypothesis that there is more emphasis on the positive aspects of post-1949 history in the Mainland after the reform. Compared to the pre-reform Taiwan textbook, the post-modern textbooks has a smaller decrease in positive-to-negative phrase ratio from pre- to post-1949 history, which suggests a more evenhanded coverage.

Turning to the number of mentions of the 6 important political entities, table 4 shows that there is a large difference between the Mainland and Taiwan pre-reform versions among the CCP entities. CCP (Mao, Marx) is mentioned 152(62, 14) times in the old Mainland textbook and 74(2, 1) times in the old Taiwan textbook. However, the difference is less striking for KMT entities. Among them, both Chiang (41, 42) and Sun (37, 39) are mentioned about equal number of times. Moreover, KMT is mentioned more in Mainland (116 times) than in Taiwan (32 times). The result signals that CCP entities are more polarizing than the KMT entities before the respective reforms. After the reform in Taiwan, all CCP entities are mentioned more. On the other hand, all KMT entities are mentioned less after the reform in Mainland.

[29]Because the old Taiwan version contains very little material on post-1949 history, we do not include post-1949 events on the Taiwan textbooks in the analysis

6 Writing Style

The difference in style across the three regions in speech and informal writing style is nontrivial especially between Hong Kong and the other two regions[30]. Can the hitherto result be explained by difference in writing styles across regions? That is, do Mainland Chinese tend to use more adjectives, and emphasize the positives in their writing? First, we note that the style difference in formal writing such as that in textbooks is usually smaller. Nevertheless, to evaluate the possibility that writing style differs, we select representative newspapers from the three regions and quantitatively examine their adjective ratio and positive-to-negative phrase ratio. Examining the writing style of newspapers serves a good proof of concept because newspapers are regularly consumed by the public which reflects the common writing style in that region.

Instead of using all the text in the newspaper, this analysis focuses on the news section on the economy and stock market only because the writing on the economy is usually more formal and therefore more similar to that of history textbooks, and unlike political news, is less subject to political influences. In creating the sample, we avoid choosing a newspaper that operates under direct government influences such as the state-owned newspaper. We select one newspaper from each region: Shanghai Morning Post (新闻晨报) circulated mainly in Shanghai, Oriental Daily (東方日報) in Hong Kong and Liberty Times (自由時報) in Taiwan. All three of them are among most popular newspapers in their respective regions. The news content is available from Wisers, a Hong Kong private company. We use all news articles in 2016 as the input and exclude editorials.

As shown in the top panel of table 7, we find that Taiwan newspaper has the highest adjective ratio, followed by that of Mainland and Hong Kong. The positive-to-negative phrase ratio of the three newspapers is presented at the bottom panel. With a range of 3.5-4.3, the ratio of the newspaper is not too different from that of the history textbooks. Liberty Times has the largest value, followed closely by Shanghai Morning Post, and then Oriental Daily. The result says that adjective ratio as well as the positive-to-negative phrase ratio in Mainland writings do not tend to be higher. This supports our assumption that the adjectives and phrases with emotional connotation are used to express a more subjective narrative. However, it should be noted that the adjective ratio in the newspapers is higher than history textbooks in general. This

[30]The primary spoken language in Hong Kong is Cantonese while the other two regions primarily speak Mandarin.

indicates that the writing style of newspapers tend to be different from that of history textbooks. To the extent that difference in the formality and rigidity of the language requirement - i.e. factors that are orthogonal to cultural difference in writing - could explain this difference in writing style between newspapers and textbooks, our conclusion is still valid.

7 Conclusion

This paper provides the first quantitative analysis of education content. We have shown that Mainland history textbooks exhibit a stronger degree of subjectivity by both the adjective ratio and the positive-to-negative phrase ratio. Moreover, Mainland textbook tend to mention the CCP entities more, and the CCP entities are closer to positive adjectives by semantic distance. The result broadly confirms the notion that authoritarian government has a stronger incentive to use a subjective narrative to align the political views of the students with that of the regime. By studying the textbook before and after reform in Mainland China and Taiwan, we have also shown that looser government control of education content led to a less subjective interpretation of historical events.

This paper highlights an aspect of persuasive communication that has received relatively less attention in the social science literature by clarifying the distinction between subjectivity and slant and focus on the former. The methods demonstrated in this paper can also be used to analyze the subjectivity of other politically-relevant content such as politicians' speeches and jurisdictional statements. There are important caveats, however. First, our approach is comprised of several individual measure, so it could be difficult to assess the overall subjectivity when different measures yield contradicting results. Second, the adjective ratio does not capture other subjective elements such as subjective nouns[31]. Previous study in Natural Language Processing has found that certain extraction pattern can be leveraged to identify subjective words (for example, the pattern ,"expressed *object*" often extracts subjective nouns, such as "concern", "hope", and "support"), and some bootstrapping algorithms can automatically generate these extraction patterns (Riloff et al. 2003). We leave this possible extension for future work.

Whether attitudes and opinions can be changed by subjectivity is unanswered in this paper. Complicating this question further in this context is that while the government's intent

[31] Jieba allows us to partially address this issue by separating out the adjective component of subjective noun.

to persuade should be explicit, students' political views are still influenced by the curriculum (Cantoni et al. 2016). This implies that persuasion can operate through both the cognitive and rational processing channel. While this paper has illustrated the possibility of persuasion via cognitive processing, future research can explicitly examine how subjectivity interacts with receivers' information on senders' intention to influence opinion.

References

[1] S. Adwan and D. Bar-On. *Learning Each Other's Historical Narrative: Israelis and Palestinians*. Peace Research Institute in the Middle East.

[2] A. Alesina and B. Reich. Nation building. 2013.

[3] M. W. Apple and L. K. Christian-Smith. *The politics of the textbook*. 1991.

[4] A. Banfield. *Unspeakable Sentences*. Boston: Routledge and Kegan Paul.

[5] N. Beauchamp. Predicting and interpolating state-level polls using twitter textual data. *American Journal of Political Science*.

[6] T. Brader. Striking a responsive chord: How political ads motivate and persuade voters by appealing to emotions. *American Journal of Political Science*, 49(2):388–405, 2005.

[7] D. Cantoni, Y. Chen, D. Yang, N. Yuchtman, and Y. J. Zhang. Curriculum and ideology. *Journal of Political Economy*, 125(2):338–392, 2017.

[8] M. Gentzkow and J. M. Shapiro. What drives media slant? evidence from us daily newspapers. *Econometrica*, 78(1):35–71, 2010.

[9] T. Groseclose and J. Milyo. A measure of media bias. *The Quarterly Journal of Economics*, pages 1191–1237, 2005.

[10] V. Hatzivassiloglou and J. Wiebe. Effects of adjective orientation and gradability on sentence subjectivity. *Proceedings of the 18th conference on Computational linguistics*, 1:299–305, 2000.

[11] C. Herbert, T. Ethofe, S. Anders, M. Junghofer, D. Wildgruber, W. Grodd, and J. Kissler. Amygdala activation during reading of emotional adjectives — an advantage for pleasant content. *Social Cognitive and Affective Neuroscience*, 4(1):35–49, 2009.

[12] D. Kahneman and A. Tversky. Judgment under uncertainty: Heuristics and biases. 1982.

[13] J. Kissler, C. Herbert, P. Peyk, and M. Junghofer. Early cortical responses to emotional words during reading. *Psychological Science*, 2007.

[14] R. Lau, L. Sigelman, and I. Brown Rovner. The effects of negative political campaigns: A meta-analytic reassessment. *The Journal of Politics*, 69(4):1176–1209, 2007.

[15] J. Lee, D.-H. Park, and I. Han. The effect of negative online consumer reviews on product attitude: An information processing view, 2008.

[16] L. v. d. Maaten and G. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008.

[17] C.-Y. Mao. 選什麼？如何選？為何而選？— 高中教師選擇歷史教科書之研究。教育科學研究. 教育科學研究, pages 123–144, 2013.

[18] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[19] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

[20] S. Mullainathan, J. Schwartzstein, and A. Shleifer. Coarse thinking and persuasion. *The Quarterly Journal of Economics*, 123(2):577–619, 2008.

[21] T. O'Hara, J. Wiebe, and R. Bruce. Selecting decomposable models for word-sense disambiguation: Thegrling-sdm system. *Computers and the Humanities*, 34(1-2):159–164, 2000.

[22] B. Qin, D. Stromberg, and Y. Wu. The political economy of social media in china. *Manuscript*, 2017.

[23] E. Riloff, J. Wiebe, and T. Wilson. Learning subjective nouns using extraction pattern bootstrapping. *Proceedings of the Seventh Conference on Natural Language Learning*, 2003.

[24] M. Roberts, B. Stewart, and E. Airoldi. A model of text for experimentation in the social sciences. *Journal of the American Statistical Association*, 111:988–1003, 2016.

[25] X. Rong. word2vec parameter learning explained. *Manuscript*, 2014.

[26] F. R. Rosselli, J. J. Skelly, and D. M. Mackie. Processing rational and emotional messages: The cognitive and affective mediation of persuasion. *Journal of Experimental Social Psychology*, 31:163–190, 1995.

[27] D. C. Sneider. History textbooks and the war in asia.

[28] J. D. Spence. *The search for modern China*. WW Norton & Company, 1991.

[29] Y. Theocharis, B. Pablo, Z. Fazekas, S. A. Popa, and O. Parnet. A bad workman blames his tweets the consequences of citizens' uncivil twitter use when interacting with party candidates. *Journal of Communication*, 66:1007–1031, 2016.

[30] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe. Predicting elections with twitter: What 140 characters reveal about political sentiment. *ICWSM*, 10:178–185, 2010.

[31] Z. Wang. National humiliation, history education, and the politics of historical memory: Patriotic education campaign in china. *International Studies Quarterly*, 52(4):783–806, 2008.

[32] J. Wiebe, R. Bruce, and T. O'Hara. Development and use of a gold standard data set for subjectivity classifications. *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, page 246–253, 1999.

[33] J. M. Wiebe. Tracking point of view in narrative. *Computational Linguistics*, 20(2):233–287, 1994.

[34] J. M. Wiebe, R. F. Bruce, and T. P. O'Hara. Development and use of a gold-standard data set for subjectivity classifications. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pages 246–253. Association for Computational Linguistics, 1999.

[35] A. Ye. Remembering different pasts: An analysis of history textbooks in mainland china and taiwan. 2016.

8 Tables

Table 1: Summary Statistics of Each Textbook Version

Region	Book	Book length	Bigram	Trigram	Quadgram	Pre-1949	Post
Hong Kong	Manhattan	70627	26206	1317	1070	42466	28
			0.37	0.02	0.02	0.60	0.
Hong Kong	Modern	91085	33866	1502	1356	58114	32
			0.37	0.02	0.01	0.64	0.
Mainland	Renjiao	41544	16778	867	986	19651	21
			0.40	0.02	0.02	0.47	0.
Mainland	Renmin	55479	22433	1226	1469	29325	26
			0.40	0.02	0.03	0.53	0.
Mainland	Yuelu	38908	15688	857	941	20029	18
			0.40	0.02	0.02	0.51	0.
Taiwan	Nanyi	30792	11112	444	345	23136	76
			0.36	0.01	0.01	0.75	0.
Taiwan	Kangxi	33640	12679	506	419	24180	94
			0.38	0.02	0.01	0.72	0.
Taiwan	Lungtun	40457	15527	610	540	28606	11
			0.38	0.02	0.01	0.71	0.
United States	Search of Modern China	410198	140591	7345	3962	235946	174
			0.34	0.02	0.01	0.58	0.
Mainland	old	56925	22746	1357	1416	40397	16
			0.40	0.02	0.02	0.71	0.
Taiwan	old	46939	16025	552	347	46665	2
			0.34	0.01	0.01	0.99	0.

The count is the number of phrases that are recognized in the Dictionary. The second row of each version represents the normalized by book length.

Table 2: Character Count and Ratio of Major Historical Episodes

	Hong Kong		Mainland			Taiwan			Ma
	Manhattan	Modern	Renjiao	Renmin	Yuelu	Nanyi	Kangxi	Lungtun	
Foreign Invasion	12326	16865	2608	3526	2232	2414	2063	2095	7
	0.175	0.185	0.063	0.064	0.057	0.078	0.085	0.052	0
Xinhai Revolution	4448	4937	1496	1472	1184	1778	2846	2598	1
	0.063	0.054	0.036	0.027	0.030	0.058	0.061	0.064	0
Civil War	3460	4112	2822	890	1000	1907	2166	2631	3
	0.049	0.045	0.068	0.016	0.026	0.062	0.064	0.065	0
Sino-Japanese War	3989	5991	1498	3293	618	1641	2016	1332	7
	0.056	0.066	0.036	0.059	0.016	0.053	0.060	0.033	0
Great Leap Forward	1484	1905	382	230	457	528	345	449	
	0.021	0.021	0.009	0.004	0.012	0.017	0.010	0.011	0
Cultural Revolution	5834	4406	940	876	429	523	1249	625	2
	0.083	0.048	0.023	0.016	0.011	0.017	0.037	0.015	0
Reform and Open	6303	7514	2308	2848	2774	1475	1644	4808	3
	0.089	0.083	0.056	0.051	0.071	0.048	0.049	0.119	0

The second row indicates the ratio of the event length to the book length. Since the old Taiwan textbook contains a very post-1949 history, we do not separate them into individual historical episodes.

Table 3: Most popular adjectives in each textbook

Region	Book	Bigram	Count	Quadgram
Hong Kong	Manhattan	革命 (Revolution)	214	内忧外患 (domestic trouble and foreign invasion)
		经济 (Economy)	178	实事求是 (be true to facts)
		开放 (open up)	69	前所未有 (unprecedented)
Hong Kong	Modern	革命 (Revolution)	250	独立自主 (act independently and of one's own initiative)
		经济 (Economy)	224	内忧外患 (domestic trouble and foreign invasion)
		开放 (open up)	98	有识之士 (man of insight)
Mainland	Renjiao	革命 (Revolution)	158	独立自主 (act independently and of one's own initiative)
		经济 (Economy)	140	百家争鸣 (contention and flourishing of numerous schools of thought)
		民主 (Democracy)	110	百花齐放 (flourishing art and literature)
Mainland	Renmin	革命 (Revolution)	266	独立自主 (act independently and of one's own initiative)
		民主 (Democracy)	154	拨乱反正 (bring order out of chaos)
		经济 (Economy)	156	实事求是 (be true to facts)
Mainland	Yuelu	经济 (Economy)	139	独立自主 (act independently and of one's own initiative)
		革命 (Revolution)	128	百花齐放 (flourishing art and literature)
		民主 (Democracy)	124	深入人心 (strike a deep chord in the hearts of the people)
Taiwan	Nanyi	革命 (Revolution)	89	有识之士 (man of insight)
		经济 (Economy)	62	痛定思痛 (draw a lesson from a bitter experience)
		国际 (International)	28	救亡图存 (save the nation from doom and strive for its survival)
Taiwan	Kangxi	革命 (Revolution)	111	有识之士 (man of insight)
		经济 (Economy)	62	百家争鸣 (contention and flourishing of numerous schools of thought)

Table 4: Number of Mentions of Important Figures and Entities

Region	Version	Mao	CCP	Marx	Chiang	KMT	Sun
Hong Kong	Manhattan	104	226	10	30	91	63
Hong Kong	Modern	82	331	8	41	151	68
Mainland	Renjiao	63	91	33	10	94	35
Mainland	Renmin	73	147	47	8	110	36
Mainland	Yuelu	47	100	19	12	85	21
United States	Search of Modern China	403	1109	91	290	439	201
Taiwan	Nane	25	94	0	21	38	35
Taiwan	Kangxi	25	96	4	21	29	21
Taiwan	Lungtun	44	121	4	17	74	33
Mainland	old	62	152	14	41	116	37
Taiwan	old	2	74	1	42	32	39

Table 5: Number of Positive and Negative Words in Each Version

Region	Version	Positive	Negative	Positive to Negative Ratio
Hong Kong	Manhattan	3376	1003	3.366
Hong Kong	Modern	4088	1149	3.558
Mainland	Renjiao	2415	531	4.548
Mainland	Renmin	3721	740	5.028
Mainland	Yuelu	2404	495	4.857
United States	Search of Modern China	21171	8367	2.530
Taiwan	Nane	1605	581	2.762
Taiwan	Kangxi	1433	465	3.082
Taiwan	Lungtun	1973	564	3.498
Mainland	old	3217	833	3.862
Taiwan	old	2002	636	3.148

Table 6: positive to negative ratio of top 20 closest adjectives

	Chiang	Mao	KMT	CCP	Sun	Marx
Mainland	1.0	5.0	0.5	3.5	4.5	6.0
Hong Kong	2.5	3.0	3.5	3.0	1.25	0.6
Taiwan	6.0	1.0	3.0	2.25	1.5	2.0
SMC	1.5	0.6	0.7	1.0	1.0	0.75

Table 7: Adjective and positive-to-negative ratio in major newspapers of the 3 regions

	Oriental Daily	Shanghai Morning Post	Liberty Times
	Hong Kong	Mainland	Taiwan
Adjective	168272	98368	50709
Word Count	1403410	777956	389745
Adjective Ratio	0.120	0.126	0.130
Positive	59298	36255	17595
Negative	16988	8663	4096
Positive-Negative Ratio	3.491	4.185	4.296

9 Figures

Figure 1: Screenshot of the dictionary application on mac that shows the word 'Important'

Figure 2: Screenshot of the dictionary application on mac that shows the word 'Very'

Figure 3: CCP to KMT entities

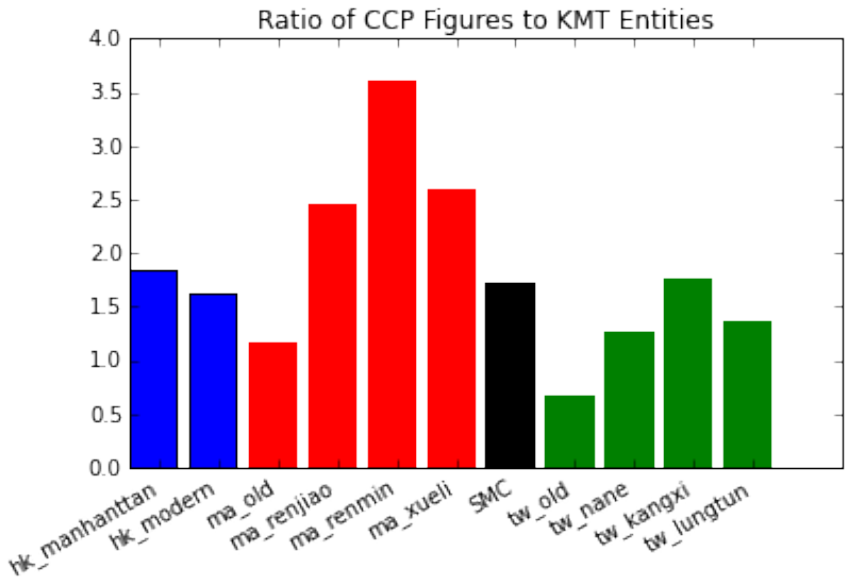


Figure 4: Adjective Ratio

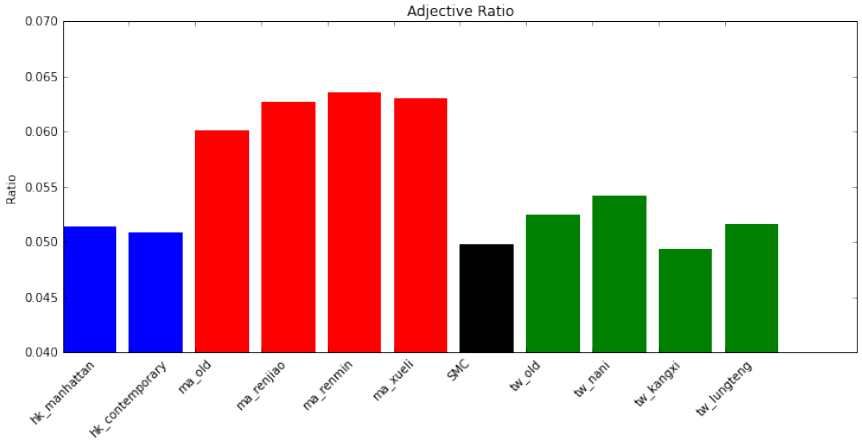


Figure 5: Adverb Ratio

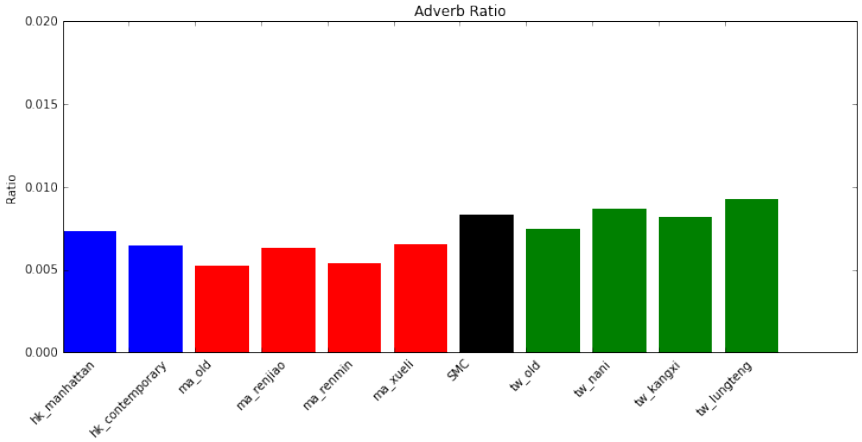


Figure 6: Adjective Ratio of pre-1949 events

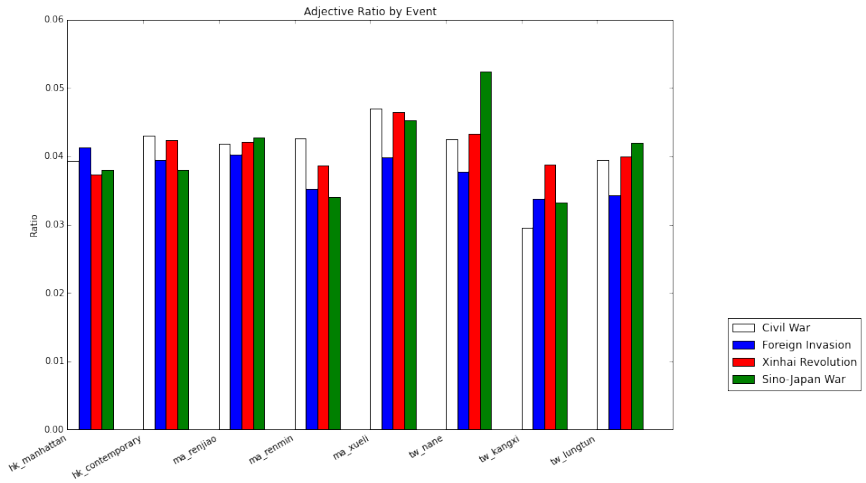


Figure 7: Adjective Ratio of post-1949 events

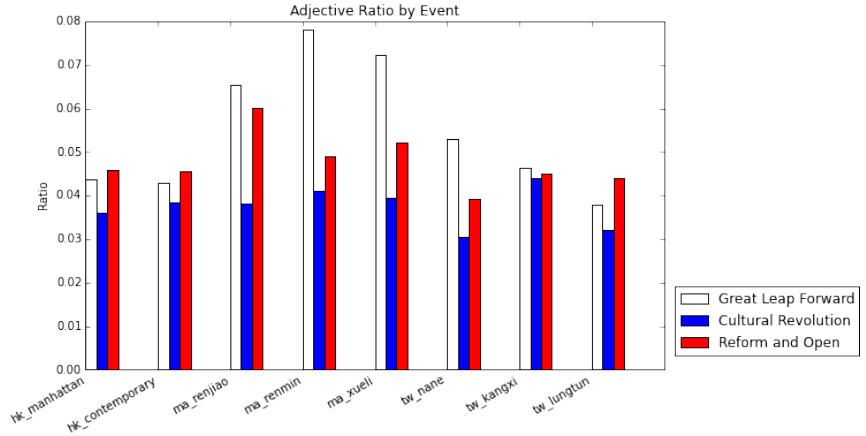
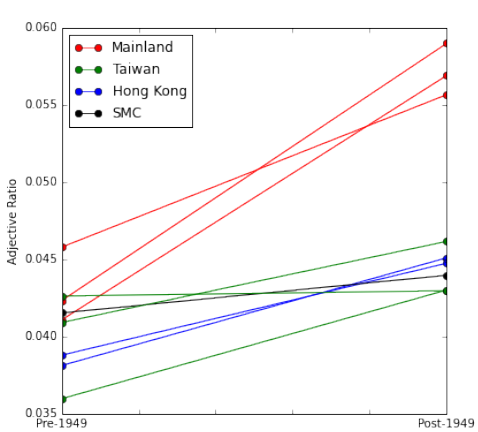


Figure 8: Adjective and Positive-to-Negative Ratio in Pre- and Post-1949



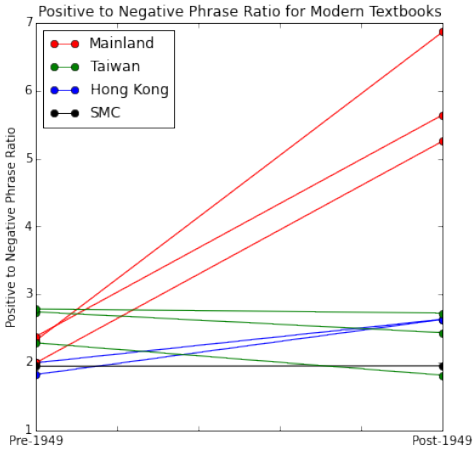


Figure 9: Positive to Negative Phrase Ratio in 4 major pre-1949 Historical Episodes

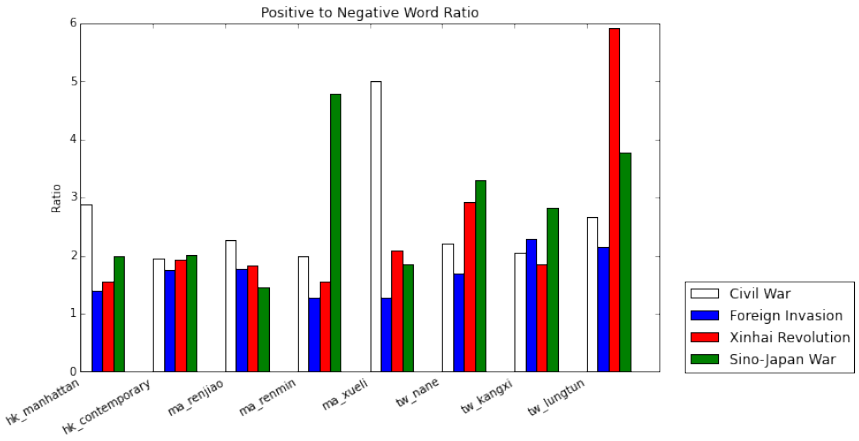


Figure 10: Positive to Negative Phrase Ratio pre- and post- 1949

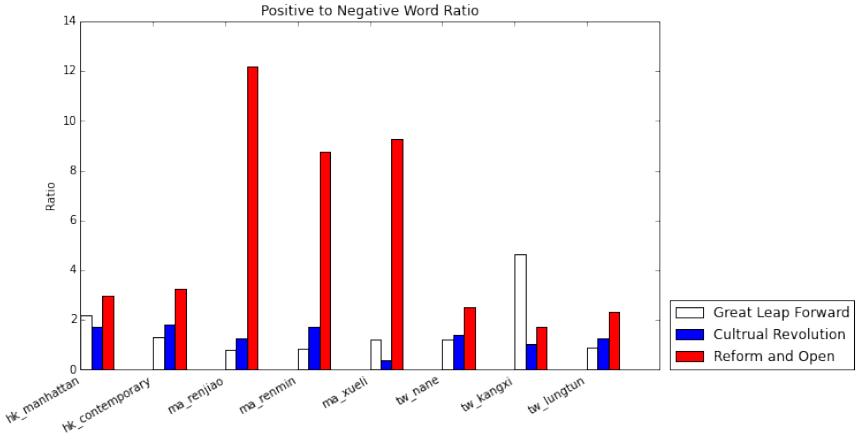


Figure 11: Adjective Ratio in Pre- and Post-Reform Versions on Individual Events

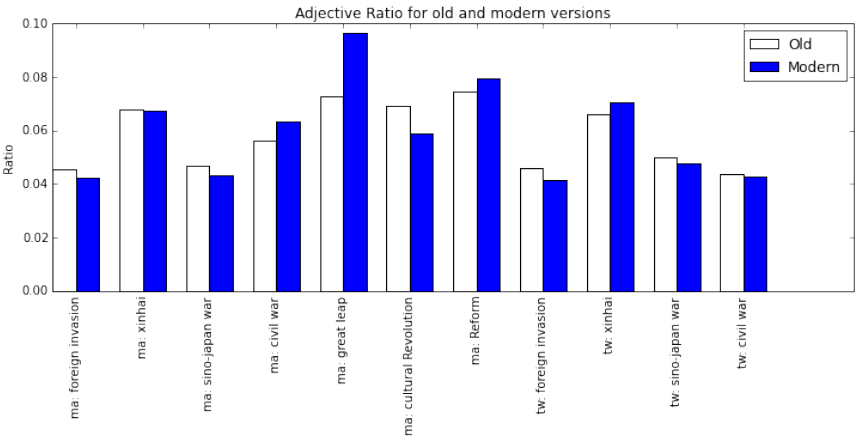


Figure 12: Adjective and Positive-to-Negative Ratio in Pre- and Post-Reform Versions in Pre- and Post-1949

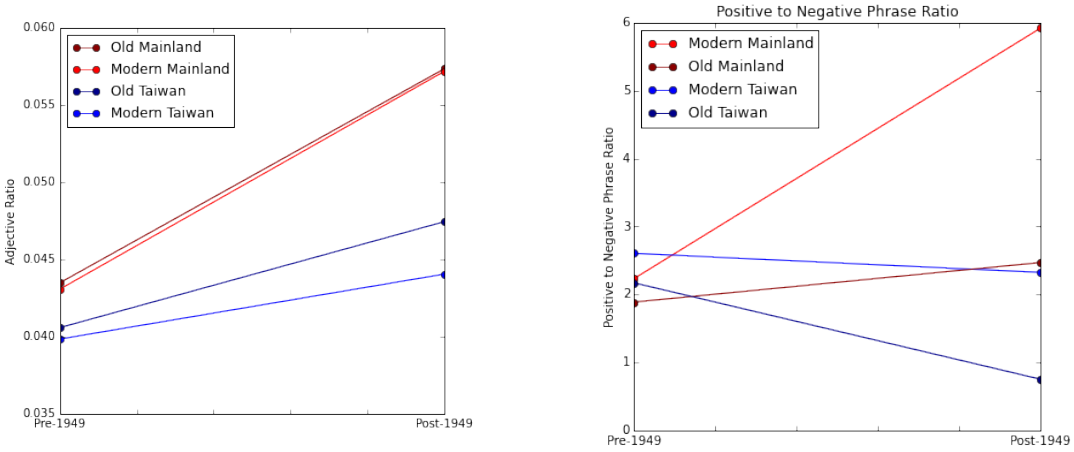
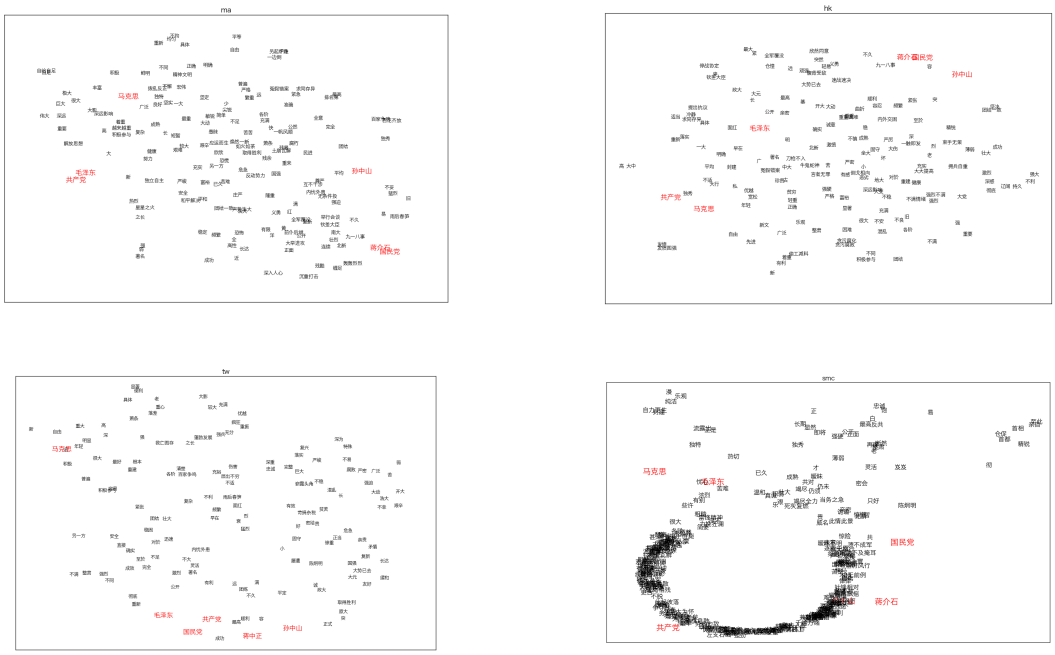


Figure 13: Word embedding visualization of Mainland textbooks by t-SNE



Note: The phrases highlighted in red are the 6 political entities of interest: 蒋介石 (Chiang Kai-Shek), 毛泽东 (Mao Ze-dong), 孙中山 (Sun Yat-sen), 国民党 (Kuomintang) , 共产党 (Chinese Communist Party), 马克思 (Marx)

10 Online Appendix

10.1 Additional Tables and Figures

Region	Version	1	2	3	4	5
Hong Kong	Manhattan	支持 (support) 57	新 (new) 47	积极 (active) 38	和平 (peace) 35	平等 (fair) 34
Hong Kong	Modern	和平 (peace) 76	支持 (support) 62	积极 (active) 46	统一 (unified) 43	重要 (important) 43
Mainland	Renjiao	新 (new) 93	和平 (peace) 40	科学 (science) 30	重要 (important) 27	统一 (unified) 26
Mainland	Renmin	新 (new) 133	科学 (science) 72	和平 (peace) 56	基础 (foundation) 51	主要 (primary) 47
Mainland	Yuelu	新 (new) 75	科学 (science) 47	和平 (peace) 43	重要 (important) 43	统一 (unified) 34
United States	Search of Modern China	新 (new) 347	重要 (important) 179	支持 (support) 164	主要 (primary) 142	接受 (accept) 137
Taiwan	Nane	新 (new) 28	重要 (important) 28	中正 (center) 21	积极 (enthusiastic) 21	主要 (primary) 19
Taiwan	Kangxi	传统 (tradition) 31	重要 (important) 30	支持 (support) 26	主要 (primary) 23	中正 (center) 21
Taiwan	Lungtun	主要 (primary) 42	新 (new) 41	重要 (important) 38	支持 (support) 35	传统 (tradition) 27
Mainland	old	新 (new) 83	主要 (primary) 65	和平 (peace) 57	统一 (unified) 50	基础 (foundation) 32
Taiwan	old	统一 (unified) 32	美 (beautiful) 23	和平 (peace) 22	允 (allow) 20	平等 (fair) 20

Table 8: Top 5 Positive Phrases

The number next to the phrase representing the number of times that phrase shows up. The column number represents the rank of the phrase. The phrase 'Revolution' and 'Economy' are excluded from the list.

Region	Version	1	2	3	4	5
Hong Kong	Manhattan	严重 (serious) 48	反 (oppose) 23	不平 (unfair) 21	专制 (dictatorial) 16	难以 (difficult to) 16
Hong Kong	Modern	严重 (serious) 52	反 (oppose) 40	难以 (difficult to) 25	官僚 (bureaucracy) 18	不平 (unfair) 15
Mainland	Renjiao	封建 (feudal) 46	斗争 (struggle) 20	严重 (serious) 16	官僚 (bureaucracy) 10	错误 (mistake) 10
Mainland	Renmin	封建 (feudal) 55	斗争 (struggle) 51	严重 (serious) 28	官僚 (bureaucracy) 20	专制 (dictatorial) 11
Mainland	Yuelu	斗争 (struggle) 27	严重 (serious) 19	专制 (dictatorial) 16	旧 (old) 15	错误 (mistake) 13
United States	Search of Modern China	似乎 (as if) 141	官僚 (bureaucracy) 122	攻击 (attack) 118	严重 (serious) 99	抗议 (protest) 97
Taiwan	Nane	严重 (serious) 17	反 (oppose) 8	讨 (beg for) 8	名义 (name) 6	所谓 (so-called) 6
Taiwan	Kangxi	斗争 (struggle) 17	策略 (strate 14	严重 (serious) 13	所谓 (so-called) 10	专制 (dictatorial) 9
Taiwan	Lungtun	反 (oppose) 19	斗争 (struggle) 17	严重 (serious) 16	动荡 (unstable) 11	官僚 (bureaucracy) 11
Mainland	old	封建 (feudal) 87	斗争 (struggle) 72	严重 (serious) 42	错误 (mistake) 37	反 (oppose) 23
Taiwan	old	复 (turn over) 27	所谓 (so-called) 16	严重 (serious) 13	不平 (unfair) 11	伪 (fake) 9

Table 9: Top 5 Negative Phrases

The number next to the phrase representing the number of times that phrase shows up. The column number represents the rank of the phrase.

Table 10: Mainland 10 closest adjectives

Chiang (蒋介石)		Mao (毛泽东)		KMT (国民党)		CCP(共产党)		Sun(孙中山)		Marx(马克思)	
Word	Distance	Word	Distance	Word	Distance	Word	Distance	Word	Distance	Word	Distance
复新 (again)	0.7202	独秀	0.6510	正面 (positive)	0.6635	重新 (again)	0.5732	腐朽 (rotten)	0.5724	广泛 (extensive)	0.6825
(again)		(duk-sau) (name)		(positive)		(again)		(rotten)		(extensive)	
北新	0.7111	拨乱反正	0.5554	轰轰烈烈	0.5740	正面	0.5494	平均	0.4795	具体	0.6699
(North New)		(bring order out of chaos)		(vigorous)		(positive)		(average)		(concrete)	
全军覆没	0.70277	解放思想	0.5518	大举进攻	0.5341	各阶	0.5469	尊严	0.4627	苦苦	0.6586
(the whole army is wiped out)		(liberate thoughts)		(launch a large-scale attack)		各阶		(dignity)		(strenuously)	
大举进攻	0.7011	正确	0.5435	危急	0.4671	充满	0.5443	另起炉灶	0.4616	大	0.5647
(launch a large-scale attack)		(correct)		(critical)		(be brimming with)		(make a fresh start)	0.4616	(big)	0.5647
正面	0.6918	准确	0.4973	不均	0.4570	广泛	0.5207	准确	0.4564	坚定	0.5517
(positive)		(Accurate)		(unequal)		(extensive)		(Accurate)		(firm)	
南大	0.6819	明确	0.4922	稳定	0.4506	拨乱反正	0.5142	红 (red)	0.4522	焕然一新	0.5111
(South Big)		(clear-cut)		(stable)		(bring order out of chaos)	0.5142	(red)	0.4522	(take on an entirely new look)	0.5111
壮烈	0.6714	最重	0.4600	恐慌	0.4490	团结	0.4984	满	0.4484	深远影响	0.5019
(heroic)		(heaviest)		(frightened)		(united)		(full)		(have far-reaching influence)	
钦差大臣	0.6678	连续	0.4565	取得胜利	0.4404	团结一致	0.4801	深入人心	0.4066	良好	0.4785
(imperial envoy)		(consecutive)		(get victory)		(united)		(strike a deep chord in the hearts of the people)		(fine)	
公开	0.5995	热烈	0.4218	均匀	0.42925	独特	0.4474	旧	0.3868	独特	0.4718
(make public)		(ardent)		(even)		(unique)		(old)		(unique)	
恐怖	0.5886	星星之火	0.4160	残暴	0.3887	焕然一新	0.4465	易	0.3815	精神文明	0.4696
(scary)		(a single spark can start a prairie fire)		(brutal)		(take on an entirely new look)		(simple)		(spiritual civilization)	

Notes: We use Jieba's part-of-speech tagging to define adjectives in this analysis.

Table 11: Hong Kong 10 closest adjectives

Chiang (蒋介石)		Mao (毛泽东)		KMT (国民党)		CCP(共产党)		Sun(孙中山)		Marx(马克思)	
Word	Distance	Word	Distance	Word	Distance	Word	Distance	Word	Distance	Word	Distance
束手无策	0.7419	老	0.5909	束手无策	0.6012	薄弱	0.5646	薄弱	0.5406	独秀	0.7022
(be at a loss at what to do)		(old)		(be at a loss at what to do)		(weak)	0.5646	(weak)		(duk-sau) (Name)	0.7022
不久	0.7303	乐观	0.5643	重建	0.4973	独秀	0.5627	成功	0.5179	广泛	0.6730
(soon)		(optimistic)		(rebuild)		(duk-sau) (Name)		(success)		(extensive)	
精锐	0.6846	亲密	0.5441	薄弱	0.4798	坚决	0.5195	落实	0.4874	充满	0.6220
(selected)		(close)		(weak)		(firm)		(relaxed)		(be brimming with)	
公开	0.6456	整肃	0.5313	壮大	0.4756	团结	0.5041	著名	0.4414	不良	0.6043
(open)		(enforce)		(strong)		(unite)		(famous)		(bad)	
大势已去	0.6018	最高	0.5296	健康	0.4707	团结一致	0.4662	激愤	0.4253	着重	0.5947
(the game is as good as lost)		(the highest)		(healthy)		(united)		(indignant)		(emphasize)	
烈	0.5929	独秀	0.5183	精锐	0.4685	最高	0.4557	不久	0.4237	封建	0.5920
(intense)		(duk-sau) (Name)		(selected)		(highest)		(soon)		(feudal)	
义勇	0.5599	大中	0.5106	坚决	0.4602	大势已去	0.4441	独秀	0.4128	不适	0.5905
(righteous and courageous)		(central)		(firm)		(the game is as good as lost)		(duk-sau) (Name)		(unsuitable)	
激愤	0.5376	言者无罪	0.5016	强烈不满	0.4560	大党	0.4350	封建	0.3999	优越	0.5870
(indignant)		(blame not the critics)		(strong resent)		(big party)		(feudal)		(superior)	0.5870
速战速决	0.5305	面红	0.4904	各阶	0.4261	精锐	0.4316	长	0.3966	著名	0.5755
(fight a quick battle)		(be red in the face)		every aspect		(selected)		(long)		(famous)	
突然	0.5202	公开	0.4657	持久	0.4247	落实	0.4186	确实	0.3678	偷工减料	0.5656
(suddenly)		(public)		(lasting)		(implement)		(real)		(do shoddy work and use inferior materials)	

Notes: We use Jiena's part-of-speech tagging to define adjectives in this analysis.

Table 12: Taiwan 10 closest adjectives

Chiang (蒋介石)		Mao (毛泽东)		KMT (国民党)		CCP(共产党)		Sun(孙中山)		Marx(马克思)	
Word	Distance	Word	Distance	Word	Distance	Word	Distance	Word	Distance	Word	
顺利	0.7834	最高	0.7222	最高	0.7448	最高	0.8217	屡遭	0.6208	旧	0.8652
(smooth)		(the highest)		(the highest)	0.7448	(the highest)		(to suffer repeatedly)		(old)	
最高	0.7704	重新	0.7029	重新	0.6985	屡遭	0.7415	友好	0.5914	深为	0.8635
(the highest)		(again)		(again)		(to suffer repeatedly)		(friendly)		(be deeply)	
取得胜利	0.7652	严峻	0.6317	严峻	0.6489	小	0.6811	顺利	0.5526	年轻	0.8239
(to get victory)		(severe)		(severe)		(small)		(smooth)		(young)	
正式	0.6539	整肃	0.5313	成功	0.6337	成功	0.6698	成功	0.5436	优越	0.8165
(formal)		(enforce)		(success)		(success)		(success)		(superior)	
成功	0.6431	崭露头角	0.5907	屡遭	0.6238	稳固	0.6666	有利	0.5375	蓬勃发展	0.8015
(success)		(come to prominence)		(to suffer repeatedly)	0.6238	(stable)		(advantageous)		(develop vigorously)	
诚	0.6141	成功	0.5888	顺利	0.6069	重新	0.6633	大势已去	0.5273	各阶	0.7983
(sincere)		(success)		(smooth)		(again)		(the game is as good as lost)		(all sectors)	
小	0.6070	腐败	0.5518	腐败	0.5700	严峻	0.6408	远	0.5199	自由	0.7967
(small)		(rotten)		(rotten)		(severe)		(far away)		(free)	
友好	0.5991	强烈	0.5417	稳固	0.5522	团结	0.6385	不幸	0.4984	百家争鸣	0.7915
(friendly)		(strong)		(stable)		(united)		(unfortunate)		(contention and flourishing of numerous schools of thought)	
屡遭	0.5891	广泛	0.5402	小	0.5453	顺利	0.6369	弱	0.4935	救亡图存	0.7886
(to suffer repeatedly)		(widespread)		(small)		(smooth)		(weak)		(save the nation from doom and strive for its survival)	
固守	0.5793	激烈	0.5376	灵活	0.5425	有利	0.6286	矛盾	0.4657	伤害	0.7880
(defend tenaciously)		(violent)		(flexible)		(advantageous)		(contradiction)		(harm)	

Notes: We use Jiena's part-of-speech tagging to define adjectives in this analysis.

Table 13: SMC 10 closest adjectives

Chiang (蒋介石)		Mao (毛泽东)		KMT (国民党)		CCP(共产党)		Sun(孙中山)		Marx(马克思)	
Word	Distance	Word	Distance	Word	Distance	Word	Distance	Word	Distance	Word	
两败俱伤	0.7250	犀利	0.7452	桀傲不驯	0.8895	仅仅只是	0.8367	悻悻	0.8157	宽大为怀	0.8433
(neither side gains)		(sharp)		(arrogant and unyielding)		(only)		(angry)		(be magnanimous)	
日经	0.6794	才 0.7360	乱事	0.8884	胆怯	0.8317	悻悻然	0.7838	适宜	0.8295	
(already)		(actually)		(troubled event)		(cowardly)		(angry)		(appropriate)	
骨肉相残	0.6391	独秀	0.7073	争执不下	0.8873	为所欲为	0.8165	偏僻	0.7781	豪言壮语	0.7558
(fratricidal fighting)		(duk-sau) (Name)		(neither could convince the other)		(do as one pleases)		(far away)		(heroic utterance)	
不惜一切	0.5726	未公开	0.7022	亲自	0.8780	尖	0.8160	娇弱	0.7380	马首是瞻	0.7384
(at all cost)		(undisclosed)		(personally)		(sharp)		(weak)		(follow somebody's lead)	
宽宏	0.5724	从先	0.6996	少壮	0.8744	最早	0.8129	悻然	0.7280	宽大	0.7180
(magnanimous)		(in the past)		(young and strong)		(earliest)		(angry)		(spacious)	
左支右绌	0.5179	深谋远虑	0.6940	亲密	0.8724	一己之私	0.8098	活泼	0.7130	冰消瓦解	0.6898
(be in straitened circumstance)		(think deeply and plan carefully)		(close)		(pursue one's own ends)		(live)		(disintegrate)	
争执不下	0.5079	经直	0.6782	暂	0.8719	纯美	0.7736	险些	0.7077	正巧	0.6793
(neither could convince the other)		(straight)		(temporarily)		(pure beauty)		(almost)		(chance to)	
反共	0.4962	整肃	0.6771	溃不成军	0.8707	浴火重生	0.7609	最强	0.6976	欢愉	0.6598
(anti-communism)		(enforce)		(be defeated and flee in great disorder)		(rebirth)		(strongest)		(happy)	
正前	0.4917	心知肚明	0.6689	毋须	0.8695	正是	0.7570	无足轻重	0.6969	浅薄	0.6339
(in front)		(be well aware)		(unnecessary)		(as is)		(be of little significance)		(shallow)	
桀傲不驯	0.4607	拙劣	0.6667	逐级	0.8685	浓烈	0.7550	大不相同	0.6847	耀眼	0.6211
(arrogant and unyielding)		(clumsy and inferior)		(step by step)		(strong (flavor))		(very different)		(dazzle)	

Notes: We use Jiena's part-of-speech tagging to define adjectives in this analysis.

Figure 14: Verb Ratio