Through the Lens of History: Methods for Analyzing Temporal Variation in Content and Framing of State-run Chinese Newspapers

Shijia Liu
Northeastern University
liu.shij@northeastern.edu

David A. Smith
Northeastern University
dasmith@ccs.neu.edu

Abstract

State-run Chinese newspapers are believed to strategically select and frame news articles to align with the shifting political tides of the country. This paper describes methods to quantify these changes in content and framing over time. Looking at more than 50 years of articles from the *People's Daily* and *Reference News*, we analyze differences in name mentions and sentiment in news articles for politicians before and after their deaths, as well as during and not during certain political events. We find significant estimates of difference, reflecting the changes in various aspects of the political environment in China during different time periods. We also apply change point detection methods to identify turning points in time series data of name mentions and sentiment. The identified turning points show a high co-occurrence with crucial political events and deaths of politicians. Furthermore, we utilize topic modeling to analyze the framing choices for articles written in different decades. The changes in frequent topic words are more significant in *People's Daily* than in *Reference News*, which is consistent with the focus shifts of the Chinese central government in history. Finally, by using pre-trained language models to predict masked names in news articles, we analyze the distinctiveness of the language used to report individuals.

(Published on Jan 1953) ...一九五三年的政治工作任务，应围绕着打好仗这一总任务，遵照彭德怀司令员一九五二年十月二十五日向全军发布的命令，更深入地加强对部队的爱国主义与国际主义教育...

...the political work tasks of 1953 should revolve around the overall task of winning the war, following the order issued by Commander Peng Dehuai to the entire army on October 25, 1952, to further strengthen the patriotic and internationalist education of the troops...

(Published on April 1968) ...高举毛泽东思想伟大红旗，同破坏活学活用毛泽东思想群众运动的中国赫鲁晓夫以及反革命修正主义分子彭德怀、罗瑞卿作坚决的斗争，勇敢地捍卫了伟大的毛泽东思想......

...Holding high the great banner of Mao Zedong Thought, we resolutely fought against the Chinese Khrushchev who sabotaged the mass movement of studying and applying Mao Zedong Thought, as well as the counter-revolutionary revisionists Peng Dehuai and Luo Ruiqing, bravely defending the great Mao Zedong Thought...

Figure 1: Excerpts of articles mentioning Peng Dehuai from *People's Daily*, published in two different eras. English translations provided by ChatGPT.

1 Introduction

Throughout the history of the People's Republic of China, the rhetoric of state-run newspapers has reflected the ideologies and power dynamics of party and government leaders. Who appears on the front pages of the newspapers and how political figures are described and framed with regard to particular political events reflect shifts in the agenda of influential leaders (Jaros and Pan, 2018). Furthermore, it is widely believed that the Chinese government has utilized various methods to control the spread of information and influence public opinion in order to stabilize society (Cook, 2022). Among these strategic controls, the existence of radical changes in reporting on politicians due to the occurrence of certain political events is an interesting phenomenon worthy of close examination. When and how such changes occur provide a unique perspective for studying the interaction of government action and societal changes in modern China.

In Chinese state-run newspapers, the language used to report political figures is often carefully

Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), pages 8143–8172
April 29 - May 4, 2025 ©2025 Association for Computational Linguistics

chosen to show where the politicians stand in various power struggles. Figure 1 shows two excerpts from articles in the *People's Daily* that mention Peng Dehuai, a well-known and controversial political and military figure of the PRC, published in two different time periods. The first was published during the Korean War, when Peng was the commander of the People's Volunteer Army and was highly regarded by Mao Zedong as a war hero. The second was published during the Cultural Revolution, when Peng was completely removed from power and labeled as a "counter-revolutionary revisionist." The word choices and the overall sentiments of these two excerpts reflect the changing political status of Peng during two political periods.

Given the example above, we ask: Is there a systematic strategy utilized by state-run Chinese newspapers that gives quantifiable evidence of difference in reporting for articles written during certain political periods? Moreover, does such evidence also manifest itself in the difference in reporting when certain political figures are alive or dead? If such evidence exists, it would strongly support the claim that Chinese newspapers select the ways in which news is reported according to changing political tides and the impact of political events in the central government. Furthermore, we would also like to understand the shifts in framing choices of articles from different decades in order to have a holistic picture of changes of focus in Chinese society that could be influenced by the government in different time periods of the regime. Lastly, we wonder whether the context used to report certain politicians is unique enough so that it can be easily identified by pre-trained language models when the actual names of the politicians are not given. In other words, can language models encode such context so that the separation of the embeddings for different politicians is significant? If so, it would imply that reporters in China take politicians' identities and power status into consideration when constructing language and rhetoric in their writings.

The contributions of this paper are as follows: (1) As far as we know, this is the first work to analyze the complete text of two widely-circulated, state-run PRC publications, *People's Daily* and *Reference News*, for the effects of political events and the deaths of certain politicians on changes in news reporting. (2) We measure features of news coverage and employ change point detection methods to quantify such effects and discover their social meaning. (3) We use topic modeling to identify the framing choices of news articles from different decades, examining the implied change of focus of the Chinese government. (4) We use a pre-trained BERT model to predict masked names in news articles and quantify how unique the language used to report different politicians is.

2 Related Work

Previous work has focused on strategies Chinese state media use to shape the perspectives and opinions of a global audience (Xiaoling, 2010; Curtin, 2012). Specifically, Sun (2009) discusses the vast difference in the coverage of current events between Chinese and international media outlets. Lee (2003) outlines the interplay between nationalism and globalism in Chinese media. In addition, Fan et al. (2024) examines the strategies used by Chinese state media on Twitter in order to reach foreign publics. Hou et al. (2021) explores the changing rhetoric of Chinese political discourse in different periods in the PRC era.

Although previous research has examined the corpus of the *People's Daily*, most of this work focuses on studying particular political issues, such as international relations (Guan, 2018; Guan and Liu, 2019), gay portrayal (Huang, 2018) and public health policy (Yang, 2021). Liu and Chang (2020) study the changes in different aspects of citizenship rights through news articles in different eras.

Looking at changes in US news coverage, Card et al. (2016) introduce a model to analyze the framing of news articles related to immigration. Card et al. (2022) use an approach based on contextual embedding of text to analyze 140 years of US political speeches that reveal changing framing of immigration by different political parties.

3 Methods

In this section, we detail the main methods used to perform our experiments, including the extraction of descriptive statistics, change point detection, topic modeling and masked token prediction by a pre-trained language model.

3.1 Descriptive Statistics and Change Point Detection

Firstly, we compute the descriptive statistics of the differences of the normalized count of articles mentioning a politician's name and the average sentiment of sentences mentioning the politician before and after their death, as well as during and not during certain political events.

We define the following variables in our model:

- Y denotes the response variable where we estimate the differences on. We experiment with two possible settings for Y: the (normalized) count of articles mentioning a politician's name and the average sentiment of sentences mentioning the politician in articles.
- L is an independent variable that indicates whether the politician is alive or dead ($L = 0$ for alive and $L = 1$ for dead).
- E is another independent variable that indicates whether articles are written in the middle of particular political events ($E = 1$ for in-event and $E = 0$ otherwise).

We then proceed to compute the average difference statistics (ADS). The two ADS quantities with respect to two independent variables L and E are defined as:

$$ADS_L = \mathbb{E}[Y \mid L = 1] - \mathbb{E}[Y \mid L = 0] \quad (1)$$

$$ADS_E = \mathbb{E}[Y \mid E = 1] - \mathbb{E}[Y \mid E = 0] \quad (2)$$

Secondly, we obtain yearly time series data on normalized name counts and average sentiment for different politicians. Then we perform change point detection (CPD) on the data. Change point detection aims to discover abrupt and significant changes in the behavior of a time series. In this case, we attempt to identify such changes in normalized name counts and average sentiment to see if they coincide with important political events and deaths of politicians.

Many algorithms have been proposed for change point detection (Van den Burg and Williams, 2020). In our analysis, we choose to use kernel change point detection (Arlot et al., 2019), which maps data to a higher-dimensional space and finds change points that minimize a certain cost function. We evaluate two kernel functions in our experiments: linear and radial basis functions.

3.2 Framing Analysis and Topic Modeling

We perform our main framing analysis on the agenda-setting mechanism of the *People's Daily* and *Reference News* using statistical topic modeling (Blei et al., 2003a). Topic modeling has been widely used in automated framing analysis (Roberts et al., 2013; Boydstun and Gross, 2013; Nguyen et al., 2013; Tsur et al., 2015; Field et al., 2018). In our case, we fit Latent Dirichlet Allocation (LDA) topic models (Blei et al., 2003b) to different decades of newspaper corpora and identify which topic words carry the highest cumulative weights among all topics extracted by the models. Those top topic words are considered to show aspects of news reporters' framing choices.

3.3 Predicting Masked Names

We also run experiments on predicting the masked names of the politicians listed in Table 1 in order to measure how similar the discourse around different politicians is. A masked language model, such as BERT (Devlin et al., 2019), is trained to predict masked tokens in sentences given the information of the context surrounding them. In order to answer the research question of how unique the context used to report a politician is, we will mask the full names of the politician and replace the names with the masking token, in addition to a general-purpose title. Then we will make a pre-trained Chinese BERT model (Cui et al., 2021) predict the masked names based on the context in the given text.

4 Datasets and Results

In this section, we first report the details of the datasets used in our experiments. We then proceed to analyze our experimental results. Table 1 and Table 2 list the politicians and political events of interest in our analysis. These politicians and political events are subjectively selected based on our judgment of the importance and representativeness of people and events in the history of the PRC.

4.1 Datasets

We perform our experiments on news articles from the *People's Daily* and *Reference News*, two widely circulated state-run newspapers in China. *People's Daily* is the official newspaper of the Central Committee of the Chinese Communist Party. *Reference News* is published by the Xinhua News Agency and translates and republishes articles by foreign news agencies. The news articles are scraped from `https://www.laoziliao.net/rmrb` and `https://www.laoziliao.net/ckxx`. The articles from *People's Daily* date from May 1946 to December 2003 (1,315,525 articles in total) and the articles from *Reference News* date from March 1957 to December 2002 (569,490 articles in to-

Politician	**Highest Government Position Held**	**Birth and Death**
Mao Zedong (毛泽东)	Chairman of the Chinese Communist Party	1893.12–1976.9
Chiang Kai-shek (蒋介石)	Leader of the Republic of China	1887.10–1975.4
Zhou Enlai (周恩来)	Premier of People's Republic of China	1898.3–1976.1
Peng Dehuai (彭德怀)	Minister of National Defense	1898.10–1974.11
Liu Shaoqi (刘少奇)	Chairman of People's Republic of China	1898.11–1969.11
Zhu De (朱德)	Chairman of NPCSC	1886.12–1976.7
Hua Guofeng (华国锋)	Chairman of the Chinese Communist Party	1921.2–2008.8
Deng Xiaoping (邓小平)	Chairman of Central Military Commission	1904.8–1997.2
Hu Yaobang (胡耀邦)	General Secretary of the Chinese Communist Party	1915.11–1989.4
Zhao Ziyang (赵紫阳)	General Secretary of the Chinese Communist Party	1919.10–2005.1
Jiang Zemin (江泽民)	General Secretary of the Chinese Communist Party	1926.8–2022.11

Table 1: Chinese politicians of interest for our snalysis

Political Events	**Time of Occurrence**
Establishment of PRC (Est. PRC)	to 1949.10
Korean War (KR War)	1950.6–1953.7
Anti-rightist Campaign (ARC)	1957.6–1957.12
The Great Leap Forward (GLF)	1958.1–1962.12
Cultural Revolution (CR)	1966.5–1976.10
Reform and Opening-up (RO)	1978.12–
Tiananmen Square Protests (TSP)	1989.4–1989.6

Table 2: Political events of interest for our analysis

tal).[1] During the scraping process, we extract the URLs, the publication dates, the titles and the text contents of news articles. The scraped data are available at `https://github.com/sliu126/chinese_newspapers.git`.

4.2 Results of Descriptive Statistics and Change Point Detection

In this section, we report the ADSs (eqs. 1–2) for our descriptive statistics analysis. We compute the ADSs on two quantities: (normalized) name count and sentiment scores, as stated in §3.1.

We normalize name counts by dividing them by the total number of articles mentioning the particular name in the corpus; hence, the normalized name count is between 0 and 1, and the ADSs for name count would be between -1 and 1. Sentiment scores of sentences are computed using the Google Cloud API[2] and are normalized between -1 (negative) and 1 (positive). We exclude politicians who lived beyond the latest articles in our dataset.

Figure 2 and Figure 3 show ADS_L for normalized name count and sentiment scores for different politicians, computed on the corpus of *People's Daily*. As we can see in Figure 2, the majority of the politicians receive more reports while alive, except for *Zhou Enlai* and *Liu Shaoqi*. We can also see in Figure 3 that most people have articles about them written with more positive sentiment after their death, suggesting that Chinese news reporters tend to show more respect to deceased politicians. The fact that *Zhou Enlai* was mentioned both more in count and more positively in sentiment after his death corresponds to the common agreement that he was posthumously very revered by the Chinese public and by the higher leadership. Interestingly, *Mao Zedong* was described with more positive sentiment while alive, possibly due to the fact that he was held up as the leader and savior of China while he was in power. After his death, articles about him were written in a more objective manner.

Figure 4 and Figure 5 illustrate the ADS_E for *Mao* in different political events. Note that the in-event and out-of-event name counts are computed as the average name counts per month in order to mitigate the fact that out-of-event time is much longer than in-event time for some events. Unsur-

[1] *Reference News* was circulated only among the highest government officials before 1957, when it became available to the public under the order of Mao Zedong.

[2] https://cloud.google.com/natural-language/docs/analyzing-sentiment

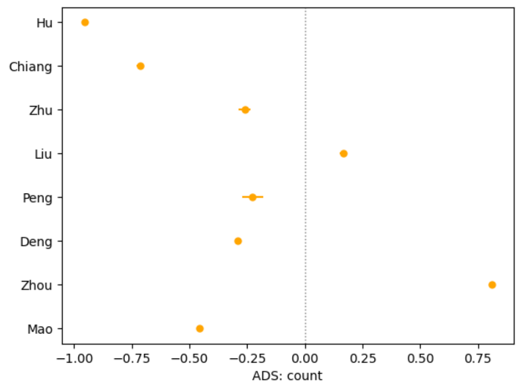

Figure 2: ADS (95% CI included) with difference in (normalized) name count when people are alive and dead (corpus: *People's Daily*)

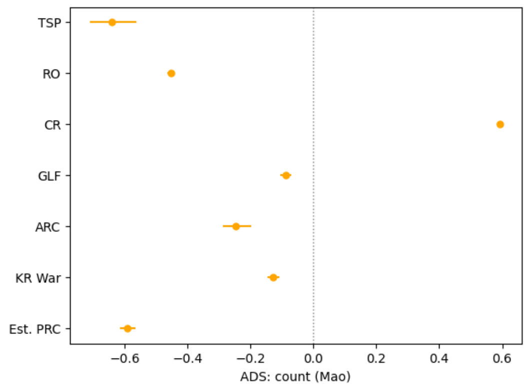

Figure 4: ADS (95% CI included) with difference in (normalized) name count for different events (politician: *Mao Zedong*, corpus: *People's Daily*)

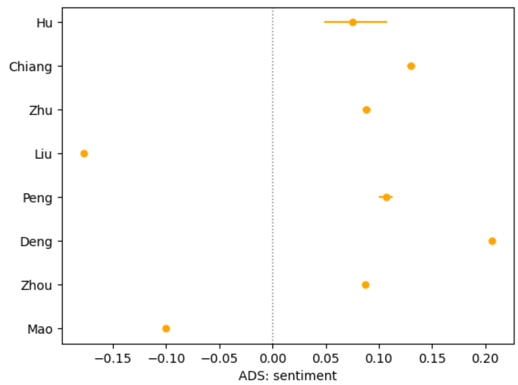

Figure 3: ADS (95% CI included) with difference in sentiment when people are alive and dead (corpus: *People's Daily*)

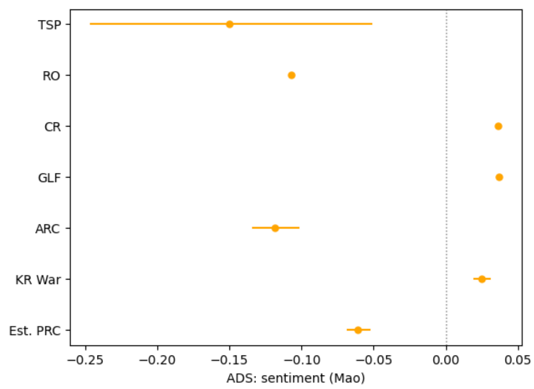

Figure 5: ADS (95% CI included) with difference in sentiment for different events (politician: *Mao Zedong*, corpus: *People's Daily*)

prisingly, *Mao* was mentioned much more during the Cultural Revolution, and articles about him were more positive during it. After the Reform and Opening-up, reporters tended to write less about him and in a more neutral manner.

Other results on different politicians and different corpus are included in Appendix A.

Table 3 and Table 4 show results for change point detection on *People's Daily* data. We use Python's Ruptures package (Truong et al., 2020) to detect change points in time series data. The minimum size between two change points is set to 1 year and the number of change points is set to 3. We want the algorithm to detect two political events (Establishment of PRC and Cultural Revolution) and the deaths of politicians. A change point identified within 1 year of the true occurrence time of an event is considered a successful detection. We choose politicians from Table 1 who have experienced the two events while in power and died in the domain of our time series data. We also provide the time series plots of normalized name count and sentiment of *Mao Zedong* in *People's Daily* in Figure 6 and Figure 7 in order to present a general picture of what the data for our analysis look like.

As we can see from the results, the algorithm is able to identify at least one expected change point for all politicians. For name count data, all three expected change points are identified with similar frequency. For average sentiment data, death (DEA) is the most commonly identified change point across all politicians, while the Establishment of the PRC (EPRC) and Cultural Revolution (CR) are almost tied in the second place. In general, we can conclude that the two major political events and death constitute major turning points in the way Chinese

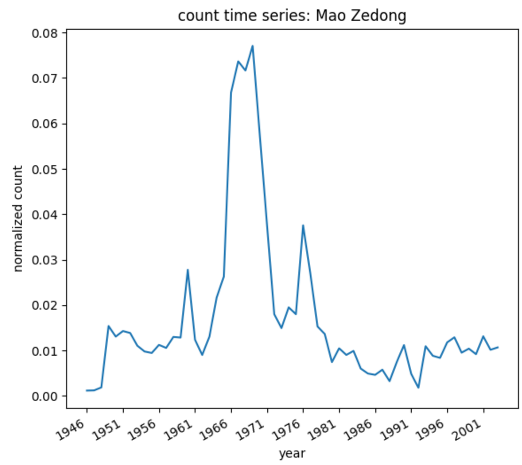

Figure 6: Time series plot for normalized name count for *Mao Zedong*, corpus: *People's Daily*)

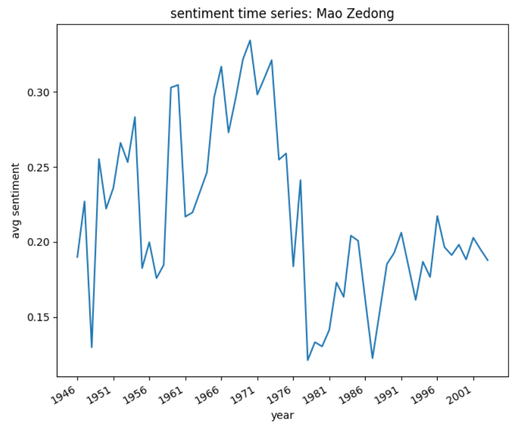

Figure 7: Time series plot for sentiment for *Mao Zedong*, corpus: *People's Daily*)

politicians are reported in state-run newspapers. Note that the change points detected that do not represent the two events and the death are turning points relevant to each individual politician. For example, the sentiment change points detected for *Deng Xiaoping* correspond to the Establishment of PRC, the "Criticize Deng" campaign in 1975, Mao's death in 1976 and Deng's Southern Tour in early 1992.

Change point detection results on *Reference News* are included in Appendix B.

4.3 Results of Topic Modeling

We use spaCy's (Honnibal and Montani, 2017) `zh_core_web_sm` model to segment and parse news articles. Afterwards, we fit an LDA topic model using the Mallet package (McCallum, 2002) (version 2.0.8) with the number of topics set to 10.

Politician	CP: linear	CP: rbf
Mao Zedong	ALL	ALL
Zhou Enlai	CR	EPRC, DEA
Deng Xiaoping	CR, DEA	CR, DEA
Peng Dehuai	EPRC, CR	EPRC, CR
Liu Shaoqi	DEA	CR
Zhu De	ALL	ALL
Chiang Kai-shek	EPRC	EPRC

Table 3: Change points detected on normalized name count over time on *People's Daily*, ALL means {EPRC, CR, DEA}

Politician	CP: linear	CP: rbf
Mao Zedong	CR, DEA	CR, DEA
Zhou Enlai	EPRC, DEA	DEA
Deng Xiaoping	EPRC	EPRC
Peng Dehuai	DEA	DEA
Liu Shaoqi	CR, DEA	CR, DEA
Zhu De	CR, DEA	CR, DEA
Chiang Kai-shek	EPRC, DEA	EPRC, DEA

Table 4: Change points detected on average sentiment over time on *People's Daily*

For each decade of our newspaper data, we extract the top five topic words with the most cumulative weights across all topics. We remove stopwords[3] during the process.

Table 5 and Table 6 show the top 5 topic words identified in retrieved topics of different decade of news articles from *People's Daily* and *Reference News*. As we can see, the distributional shifts in topics are more obvious in *People's Daily* than in *Reference News*, with "Production" and "Revolution" in earlier decades and "Economy" and "Development" in later decades, which corresponds to a shift of focus from political (class) struggle to economy development of the Chinese government. However, the top topic words from *Reference News* are always about different countries, reflecting the fact that the newspaper strategically selects articles from foreign news agencies that mainly focus on China's diplomacy.

Other results on selected politicians are shown in Appendix C. Note that we only show politicians who have significant coverage, with at least 100 articles mentioning them in all decades in the newspapers.

[3]We use the Chinese stopwords list from `https://github.com/stopwords-iso/stopwords-zh`

Decade	Top 5 Topic Words with Weights				
1940	The Masses 群众 0.0628	Work 工作 0.0504	China 中国 0.0380	Production 生产 0.0334	Government 政府 0.0295
1950	Work 工作 0.0787	Production 生产 0.0527	China 中国 0.0510	Government 政府 0.0375	Soviet Union 苏联 0.0374
1960	Revolution 革命 0.0744	China 中国 0.0652	Production 生产 0.0540	The U.S. 美国 0.0538	Struggle 斗争 0.0443
1970	Revolution 革命 0.0783	China 中国 0.0583	Struggle 斗争 0.0489	Chairman Mao 毛主席 0.0386	Work 工作 0.0358
1980	China 中国 0.0600	Work 工作 0.0446	Development 发展 0.0420	Country 国家 0.0354	Economy 经济 0.0323
1990	China 中国 0.0786	Development 发展 0.0531	Economy 经济 0.0457	Work 工作 0.0445	Enterprise 企业 0.0373
2000	China 中国 0.0694	Development 发展 0.0605	Work 工作 0.0486	Enterprise 企业 0.0330	Construction 建设 0.0305

Table 5: Top topic words identified in different decades of articles in *People's Daily*

4.4 Results of Predicting Masked Names

In this section, we report the results of masking different politicians' names in articles from *People's Daily* and having the Chinese BERT model (Cui et al., 2021) predicting the masked politicians.[4] From each decade of articles mentioning the politicians listed in Table 1, we randomly sample 100 articles and remove the full names of the politicians. Then we insert "[MASK]先生" at the places where the politicians' names are removed. We append "先生" after the masking token to encourage the language model to predict a surname. We then map the predicted surnames to actual Chinese politicians based on our knowledge of Chinese history.

Table 7 and Table 8 show the top five politicians who could replace the masked surnames of *Mao* and *Zhou* with their average token probabilities per sentence, predicted by the Chinese BERT model. As we can see, the model is able to identify those who have worked closely with Mao and Zhou. It is unsurprising that Zhou appears in Mao's top replacements and Mao appears in Zhou's, since these two politicians are constantly paired up to perform government duties. Deng Xiaoping is identified as a top replacement for Mao, probably because they have both been the highest leader of the PRC. Li Xiannian and Chen Yun, who appear in both tables, have been in the central power since the establishment of the PRC and have assumed crucial roles in the government (Li was the Chairman of PRC and Chen was the Vice Premier).

We also calculate the top five replacements for different politicians during different political events listed in Table 2. The results for selected politicians are in Appendix E. In general, the predicted probability of the true names is higher when the masked politicians were either in power or in the center of a political turmoil. For example, *Peng Dehuai's* replacements include his own name in top five names during the Korean War, the Anti-rightist Campaign and the Great Leap Forward. However, his name disappears during the Cultural Revolution when he was ostracized by the central government (see Appendix E.1). Similarly, *Liu Shaoqi's* own name ranks higher as a replacement for himself during the Anti-rightist Campaign and the Great Leap Forward, but lower before the establishment of PRC and during the cultural revolution when he was not acting as the chairman of PRC (see Appendix E.2).

Finally, we are also interested in the gaps be-

[4] We use Huggingface's `hfl/chinese-bert-wwm` model in our experiments.

Decade	Top 5 Topic Words with Weights				
1950	The U.S. 美国 0.1182	China 中国 0.0977	Government 政府 0.0739	Country 国家 0.0596	Soviet Union 苏联 0.0560
1960	The U.S. 美国 0.1163	China 中国 0.0920	Soviet Union 苏联 0.0768	Government 政府 0.0620	Country 国家 0.0519
1970	Soviet Union 苏联 0.1156	The U.S. 美国 0.1155	China 中国 0.1029	Country 国家 0.0585	Government 政府 0.0537
1980	The U.S. 美国 0.0893	China 中国 0.0857	Soviet Union 苏联 0.0645	Government 政府 0.0407	Country 国家 0.0389
1990	The U.S. 美国 0.0776	China 中国 0.0653	Country 国家 0.0404	Japan 日本 0.0333	Economy 经济 0.0318
2000	The U.S. 美国 0.1027	China 中国 0.0559	Company 公司 0.0393	Government 政府 0.0337	Taiwan 台湾 0.0331

Table 6: Top topic words identified in different decades of articles in *Reference News*

Politician	Avg Probability
Mao Zedong (毛泽东)	0.3112
De Xiaoping (邓小平)	0.0385
Zhou Enlai (周恩来)	0.0348
Li Xiannian (李先念)	0.0318
Chen Yun (陈云)	0.0252

Table 7: Top 5 politicians' names generated by BERT to replace the masked *Mao Zedong's* name in articles from *People's Daily*

Politician	Avg Probability
Zhou Enlai (周恩来)	0.1815
Li Xiannian (李先念)	0.0759
Chen Yun (陈云)	0.0342
Mao Zedong (毛泽东)	0.0300
Yang Shangkun (杨尚昆)	0.0267

Table 8: Top 5 politicians' names generated by BERT to replace the masked *Zhou Enlai's* name in articles from *People's Daily*

tween the predicted probabilities of the true name and of other replacement names for different politicians, since they could give us information on how unique the language used to report those politicians is. Therefore, we define the uniqueness ratio to be the ratio between the predicted probability of the true name and the highest predicted probability of replacement names. Furthermore, we would like

Politician	Uniq. Ratio	Entropy
Mao Zedong	8.0831	1.4606
Zhou Enlai	2.3913	1.8865
Chiang Kai-shek	1.2846	2.2631
Deng Xiaoping	1.2615	2.2132
Zhu De	1.1411	2.3053
Hu Yaobang	1.1140	2.2607
Jiang Zemin	0.9863	2.2735
Liu Shaoqi	0.8970	2.2807
Peng Dehuai	0.8158	2.3136
Hua Guofeng	0.7427	2.2456
Zhao Ziyang	0.4763	2.2221

Table 9: Uniquenes ratio and entropy calculated from predictions made by Chinese BERT for different politicians

to know how certain the language model is when making predictions, which can be represented by the entropy of the predictions. Table 9 shows the uniqueness ratio and entropy (in bits) of different politicians. Interestingly, those with higher uniqueness ratios are the exact people who have been in power for a long period of time, with the highest ratio belonging to Mao himself. Those with lower uniqueness ratios are the ones who have been in power and then removed from power, reflecting the tragic fact that the language used to report politicians in the PRC could change drastically for the same people with different political status. The language model tends to be more certain when predicting the replacements for Mao and Zhou, corresponding to the fact that the language used to report these two is more individualized.

5 Conclusion

We extract statistics on coverage of selected politicians and employ change point detection methods to news articles from *People's Daily* and *Reference News* to discover the difference in the ways of reporting during different political events and when deaths of political figures occur. Results show that such differences exist and can be traced to changes in political tides and public opinion in different eras of the PRC. Furthermore, through topic modeling we discover the changes in framing choices on news articles in different decades, which reflect the evolving agenda and shifts of focus of the Chinese government. Finally, by predicting masked names of politicians using a BERT model, we identify the difference in the uniqueness of the context used to report different politicians. We discover that those who have held power for a longer period of time tend to have more uniqueness in the language used to report them in articles. We hope that the methods used in this work can be applied to news articles from other media outlets and political systems.

Future work may involve inferring power networks among politicians by examining the closeness of language used by reporters when discussing them during political events. It would also be interesting to compare the results with existing discoveries of "inner circles" among Chinese political elites. In addition, it would be informative to study the framing choices adopted by newspapers published under different regimes on certain Chinese political events, since they might reflect the links between rhetoric in news reporting and ideologies of news agencies. We believe that these are exciting directions that could add valuable knowledge to our understanding of how news reporting is done in the PRC.

6 Limitations

We believe that we have conducted a comprehensive study on the corpus of two Chinese state-run newspapers that reveals multiple aspects of Chinese politics in different eras of the PRC. However, we suggest the following limitations that might open up avenues for future work:

- The implications of the top replacement names given by BERT can be further examined. How these names correlate to the political careers of the masked people is an interesting question to explore. For example, we can extract personal networks of certain politicians from their Wikipedia pages or history textbooks and compare the information to the experimental results we have.

- Regarding framing, it would be interesting to compare the results of topic modeling on *People's Daily* with corpora of newspapers published under different regimes, such as Taiwanese newspapers. We believe that newspapers that carry different ideologies would frame the same political events differently.

- It is also interesting to make large language models identify the top topic words in sets of articles and compare the results with the LDA analysis. This would inspect LLMs' abilities of framing analysis on news articles, which may provide useful information on how powerful LLMs currently are.

References

Sylvain Arlot, Alain Celisse, and Zaid Harchaoui. 2019. A kernel multiple change-point algorithm via model selection. *Preprint*, arXiv:1202.3878.

David M. Blei, Michael I. Jordan, Thomas L. Griffiths, and Joshua B. Tenenbaum. 2003a. Hierarchical topic models and the nested chinese restaurant process. In *Proceedings of the 16th International Conference on Neural Information Processing Systems*, NIPS'03, page 17–24, Cambridge, MA, USA. MIT Press.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003b. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3(null):993–1022.

Amber E. Boydstun and Justin H. Gross. 2013. Identifying media frames and frame dynamics within and across policy issues.

Dallas Card, Serina Chang, Chris Becker, Julia Mendelsohn, Rob Voigt, Leah Boustan, Ran Abramitzky, and Dan Jurafsky. 2022. Computational analysis of 140 years of us political speeches reveals more positive but increasingly polarized framing of immigration. *Proceedings of the National Academy of Sciences*, 119(31):e2120510119.

Dallas Card, Justin Gross, Amber Boydstun, and Noah A. Smith. 2016. Analyzing framing through the casts of characters in the news. In *Proceedings of* *the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1410–1420, Austin, Texas. Association for Computational Linguistics.

Sarah Cook. 2022. Countering beijing's media manipulation. *Journal of Democracy*, 33(1):116–130.

Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, and Ziqing Yang. 2021. Pre-training with whole word masking for chinese bert. *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, 29:3504–3514.

Michael Curtin. 2012. Chinese media and globalization. *Chinese Journal of Communication*, 5(1):1–9.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Yingjie Fan, Jennifer Pan, and Jaymee Sheng. 2024. Strategies of chinese state media on twitter. *Political Communication*, 41(1):4–25.

Anjalie Field, Doron Kliger, Shuly Wintner, Jennifer Pan, Dan Jurafsky, and Yulia Tsvetkov. 2018. Framing and agenda-setting in Russian news: a computational analysis of intricate political strategies. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3570–3580, Brussels, Belgium. Association for Computational Linguistics.

Tianru Guan. 2018. Framing the boundary of sino-japanese conflicts in china's communication sphere: A content analysis of the news coverage of japan and sino-japanese controversies by the people's daily between 2001 and 2015. *Journal of Chinese Political Science*, 23(4):603–618.

Tianru Guan and Tianyang Liu. 2019. Fears, hopes and the politics of time-space: The media frames of japan in the chinese people's daily. *International Communication Gazette*, 81(6-8):664–685.

Matthew Honnibal and Ines Montani. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear.

Renkui Hou, Chu-Ren Huang, and Kathleen Ahrens. 2021. Language change in chinese political discourse based on the relationship between sentence and clause. In *Proceedings of the 35th Pacific Asia conference on language, information and computation*, pages 244–250.

Yixiong Huang. 2018. Media representation of tongxinglian in china: A case study of the people's daily. *Journal of homosexuality*, 65(3):338–360.

Kyle Jaros and Jennifer Pan. 2018. China's newsmakers: Official media coverage and political shifts in the xi jinping era. *The China Quarterly*, 233:111–136.

Chin-Chuan Lee. 2003. The global and the national of the chinese media. *Chinese media, global contexts*, pages 1–31.

Stephanie Na Liu and Tsan-Kuo Chang. 2020. Continuities and changes of media construction of citizenship rights in china: the case of the people's daily, 1978–2012. *Asian Journal of Communication*, 30(5):343–362.

Andrew Kachites McCallum. 2002. Mallet: A machine learning for language toolkit. Http://mallet.cs.umass.edu.

Viet-An Nguyen, Jordan Boyd-Graber, and Philip Resnik. 2013. Lexical and hierarchical topic regression. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'13, page 1106–1114, Red Hook, NY, USA. Curran Associates Inc.

Margaret E. Roberts, Brandon M Stewart, Dustin Tingley, and Edoardo M. Airoldi. 2013. The structural topic model and applied social science. In *International Conference on Neural Information Processing*.

Wanning Sun. 2009. Mission impossible? soft power, communication capacity, and the globalization of chinese media. *International journal of communication*, 4:19.

Charles Truong, Laurent Oudre, and Nicolas Vayatis. 2020. Selective review of offline change point detection methods. *Signal Processing*, 167:107299.

Oren Tsur, Dan Calacci, and David Lazer. 2015. A frame of mind: Using statistical models for detection of framing and agenda setting campaigns. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1629–1638, Beijing, China. Association for Computational Linguistics.

Gerrit JJ Van den Burg and Christopher KI Williams. 2020. An evaluation of change point detection algorithms. *arXiv preprint arXiv:2003.06222*.

Zhang Xiaoling. 2010. Chinese state media going global. *East Asia Policy*, 2(1):42–50.

Zheng Yang. 2021. Military metaphors in contemporary chinese disease coverage: a case study of the people's daily, 1946–2019. *Chinese Journal of Communication*, 14(3):259–277.

A Additional Descriptive Statistics Results

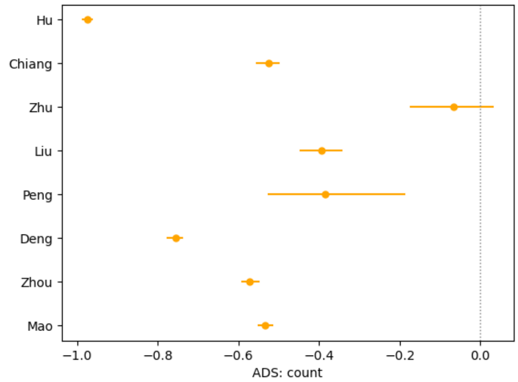

Figure 8: ADS (95% CI included) with difference in (normalized) name count when people are alive and dead (corpus: *Reference News*)

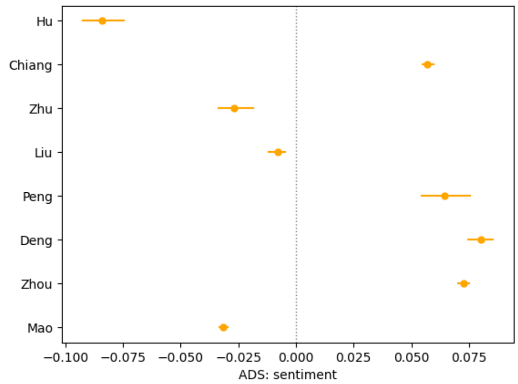

Figure 9: ADS (95% CI included) with difference in sentiment when people are alive and dead (corpus: *Reference News*)

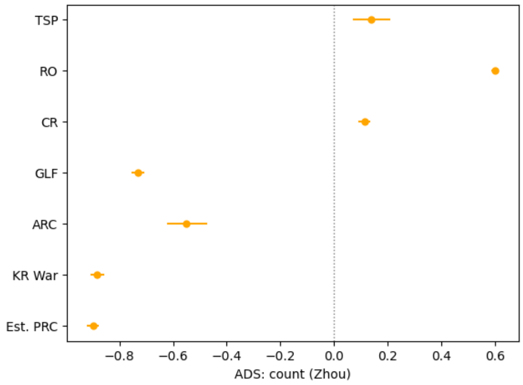

Figure 10: ADS (95% CI included) with difference in (normalized) name count for different events (politician: *Zhou Enlai*, corpus: *People's Daily*)

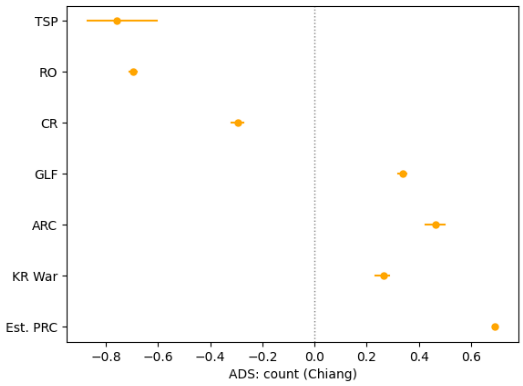

Figure 11: ADS (95% CI included) with difference in (normalized) name count for different events (politician: *Chiang Kai-shek*, corpus: *People's Daily*)

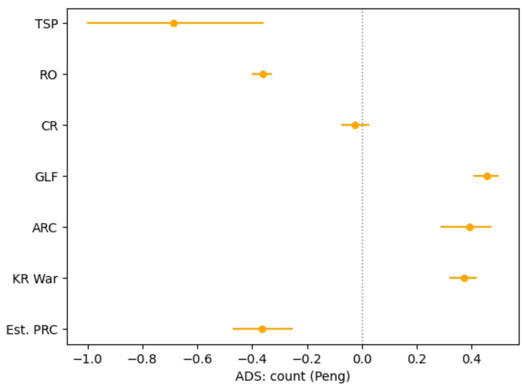

Figure 12: ADS (95% CI included) with difference in (normalized) name count for different events (politician: *Peng Dehuai*, corpus: *People's Daily*)

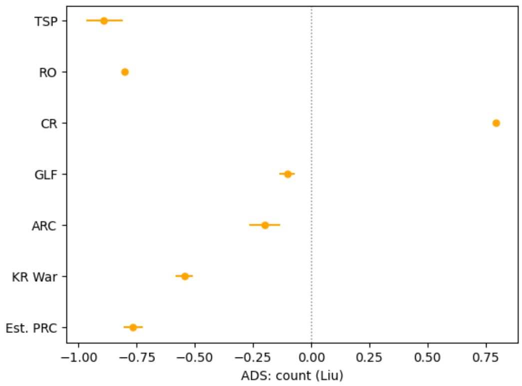

Figure 13: ADS (95% CI included) with difference in (normalized) name count for different events (politician: *Liu Shaoqi*, corpus: *People's Daily*)

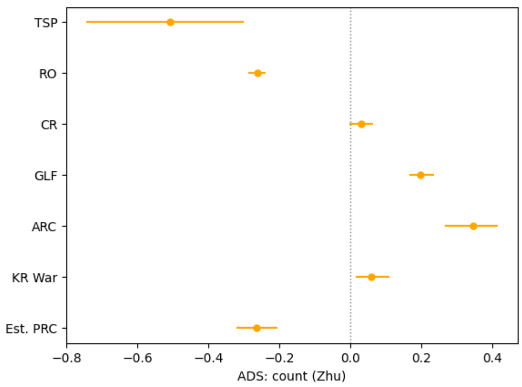

Figure 14: ADS (95% CI included) with difference in (normalized) name count for different events (politician: *Zhu De*, corpus: *People's Daily*)

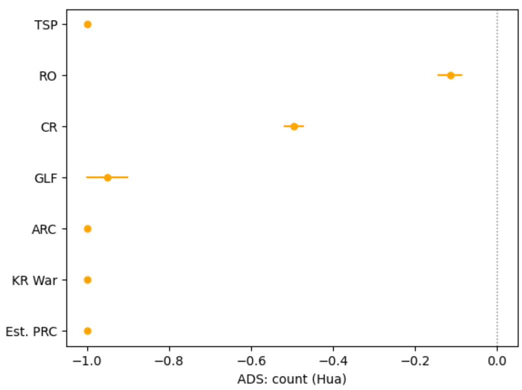

Figure 15: ADS (95% CI included) with difference in (normalized) name count for different events (politician: *Hua Guofeng*, corpus: *People's Daily*)

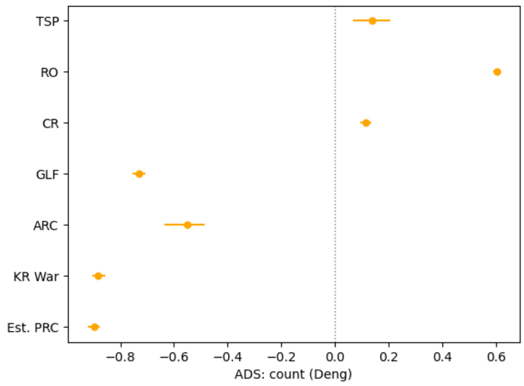

Figure 16: ADS (95% CI included) with difference in (normalized) name count for different events (politician: *Deng Xiaoping*, corpus: *People's Daily*)

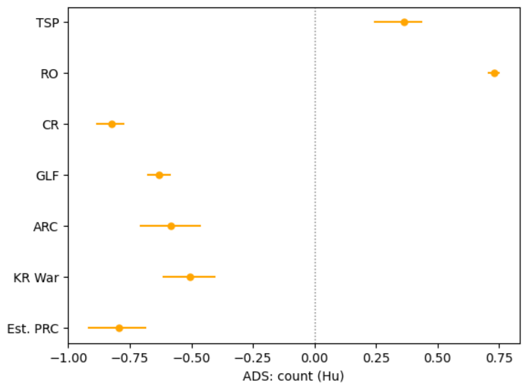

Figure 17: ADS (95% CI included) with difference in (normalized) name count for different events (politician: *Hu Yaobang*, corpus: *People's Daily*)

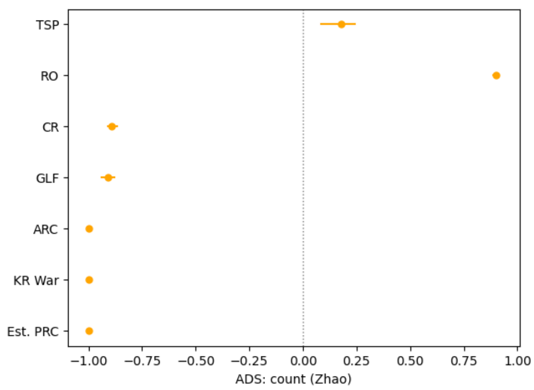

Figure 18: ADS (95% CI included) with difference in (normalized) name count for different events (politician: *Zhao Ziyang*, corpus: *People's Daily*)

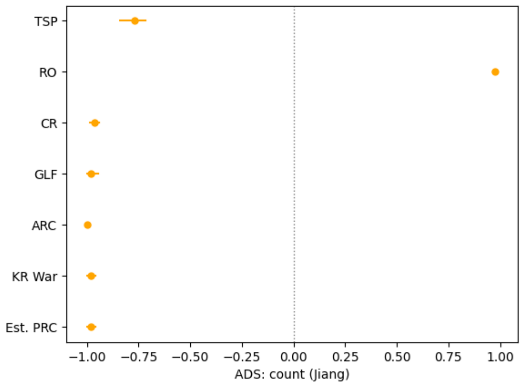

Figure 19: ADS (95% CI included) with difference in (normalized) name count for different events (politician: *Jiang Zemin*, corpus: *People's Daily*)

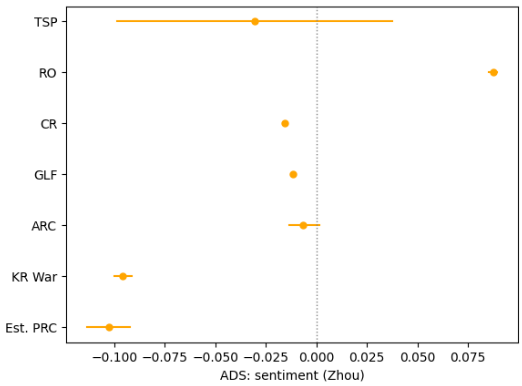

Figure 20: ADS (95% CI included) with difference in sentiment for different events (politician: *Zhou Enlai*, corpus: *People's Daily*)

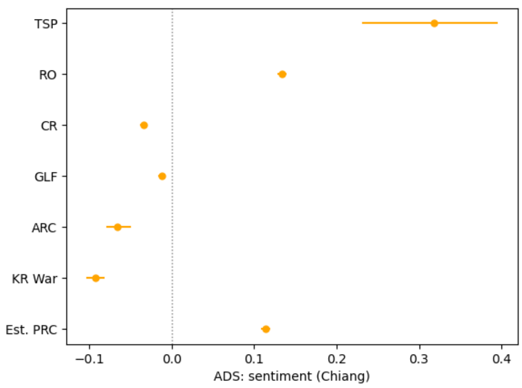

Figure 21: ADS (95% CI included) with difference in sentiment for different events (politician: *Chiang Kai-shek*, corpus: *People's Daily*)

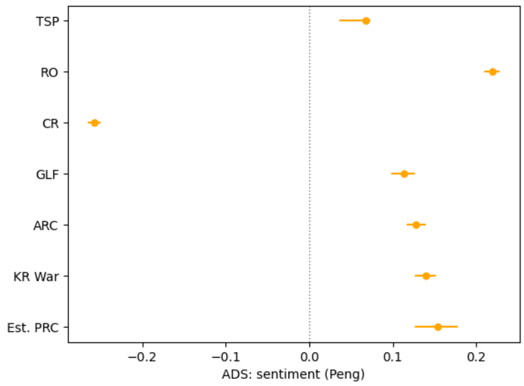

Figure 22: ADS (95% CI included) with difference in sentiment for different events (politician: *Peng Dehuai*, corpus: *People's Daily*)

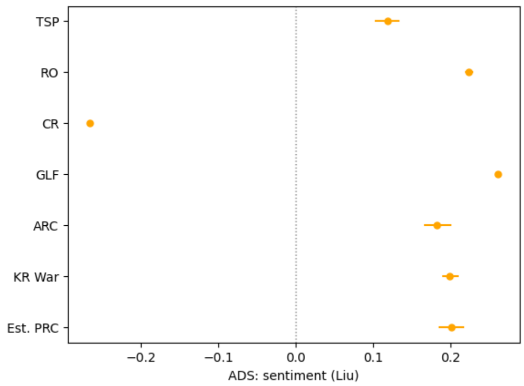

Figure 23: ADS (95% CI included) with difference in sentiment for different events (politician: *Liu Shaoqi*, corpus: *People's Daily*)

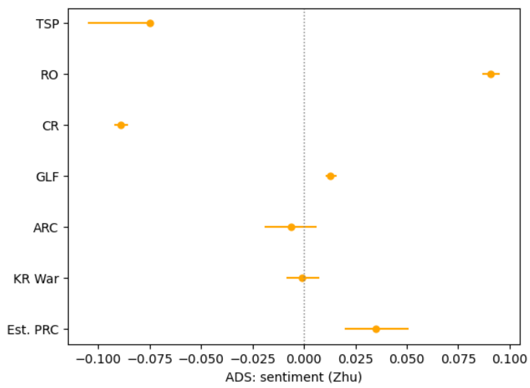

Figure 24: ADS (95% CI included) with difference in sentiment for different events (politician: *Zhu De*, corpus: *People's Daily*)

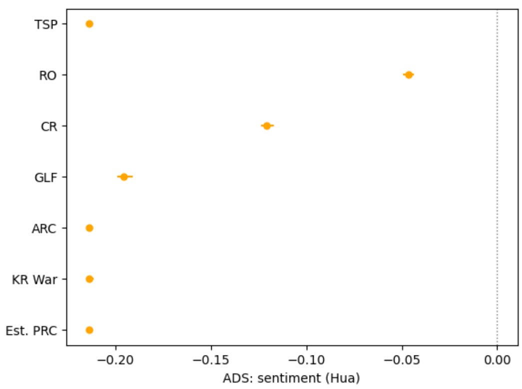

Figure 25: ADS (95% CI included) with difference in sentiment for different events (politician: *Hua Guofeng*, corpus: *People's Daily*)

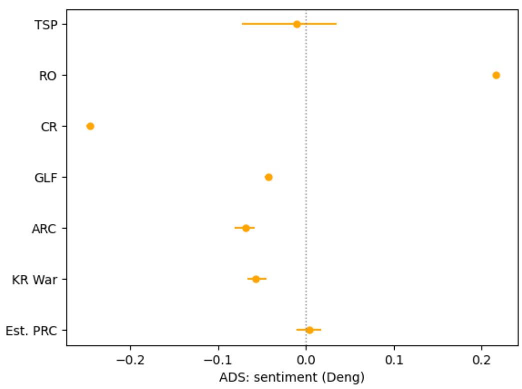

Figure 26: ADS (95% CI included) with difference in sentiment for different events (politician: *Deng Xiaoping*, corpus: *People's Daily*)

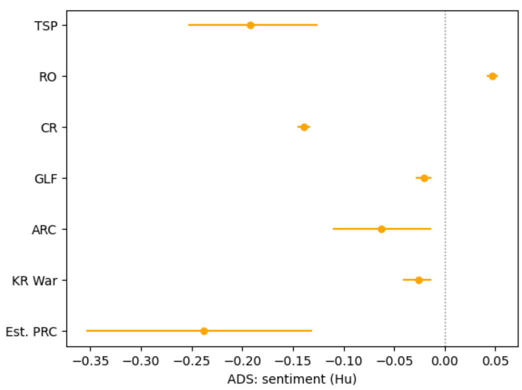

Figure 27: ADS (95% CI included) with difference in sentiment for different events (politician: *Hu Yaobang*, corpus: *People's Daily*)

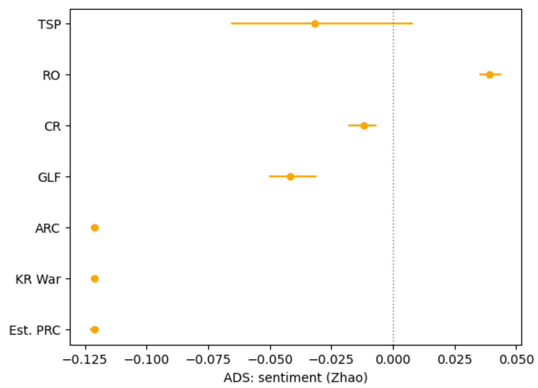

Figure 28: ADS (95% CI included) with difference in sentiment for different events (politician: *Zhao Ziyang*, corpus: *People's Daily*)

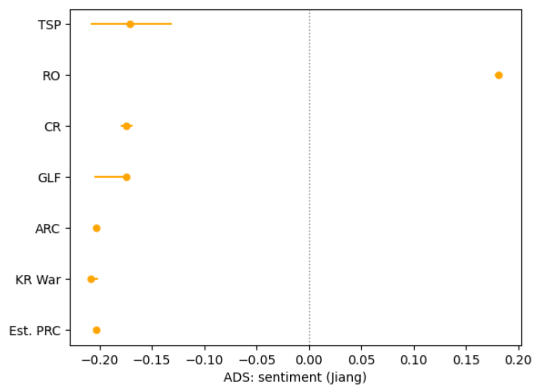

Figure 29: ADS (95% CI included) with difference in sentiment for different events (politician: *Jiang Zemin*, corpus: *People's Daily*)

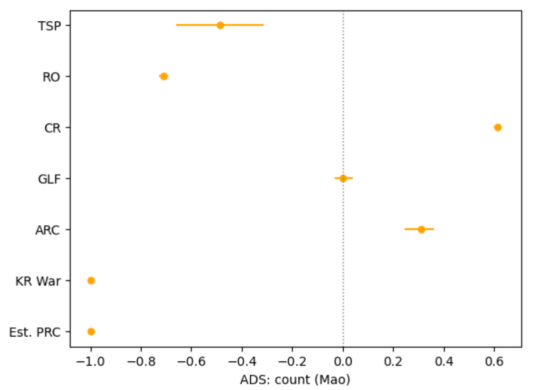

Figure 30: ADS (95% CI included) with difference in (normalized) name count for different events (politician: *Mao Zedong*, corpus: *Reference News*)

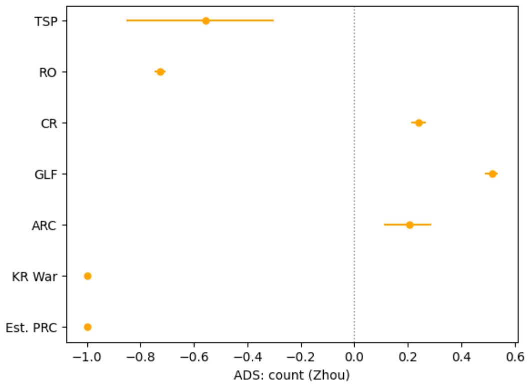

Figure 31: ADS (95% CI included) with difference in (normalized) name count for different events (politician: *Zhou Enlai*, corpus: *Reference News*)

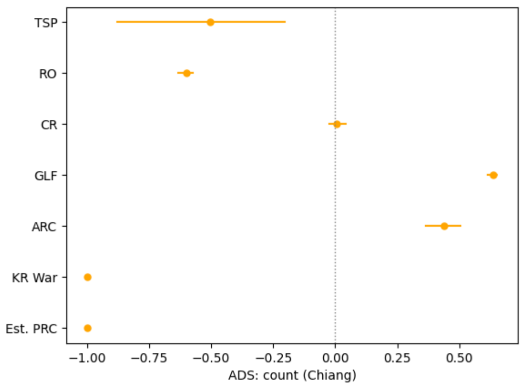

Figure 32: ADS (95% CI included) with difference in (normalized) name count for different events (politician: *Chiang Kai-shek*, corpus: *Reference News*)

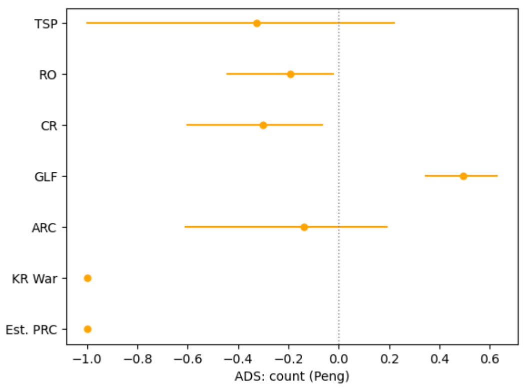

Figure 33: ADS (95% CI included) with difference in (normalized) name count for different events (politician: *Peng Dehuai*, corpus: *Reference News*)

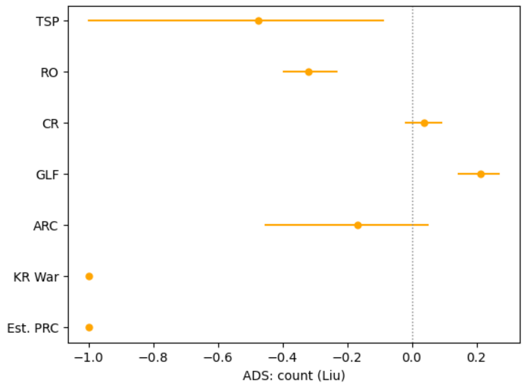

Figure 34: ADS (95% CI included) with difference in (normalized) name count for different events (politician: *Liu Shaoqi*, corpus: *Reference News*)

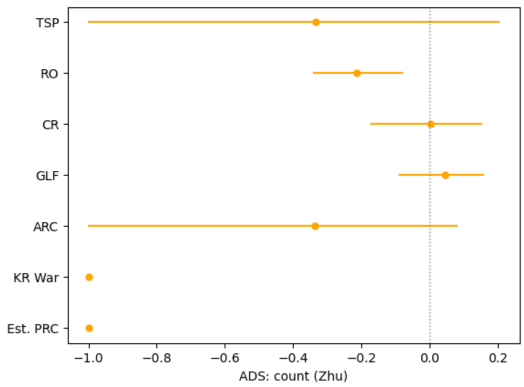

Figure 35: ADS (95% CI included) with difference in (normalized) name count for different events (politician: *Zhu De*, corpus: *Reference News*)

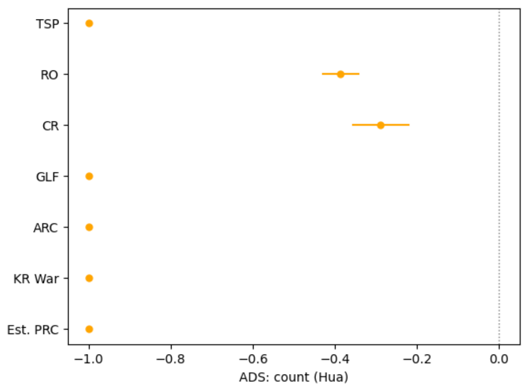

Figure 36: ADS (95% CI included) with difference in (normalized) name count for different events (politician: *Hua Guofeng*, corpus: *Reference News*)

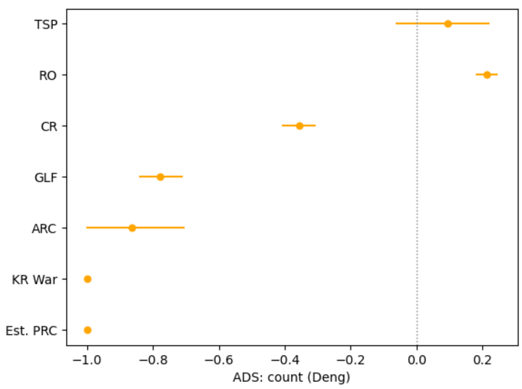

Figure 37: ADS (95% CI included) with difference in (normalized) name count for different events (politician: *Deng Xiaoping*, corpus: *Reference News*)

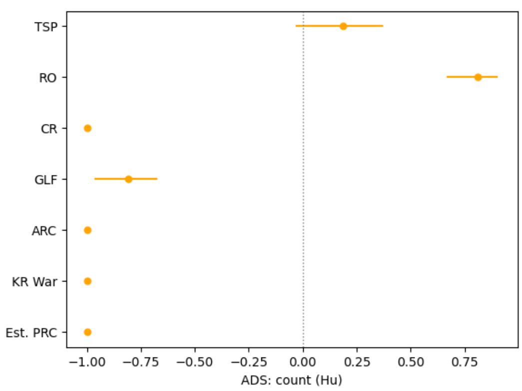

Figure 38: ADS (95% CI included) with difference in (normalized) name count for different events (politician: *Hu Yaobang*, corpus: *Reference News*)

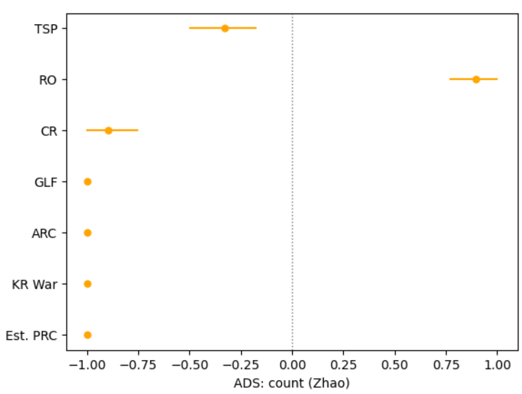

Figure 39: ADS (95% CI included) with difference in (normalized) name count for different events (politician: *Zhao Ziyang*, corpus: *Reference News*)

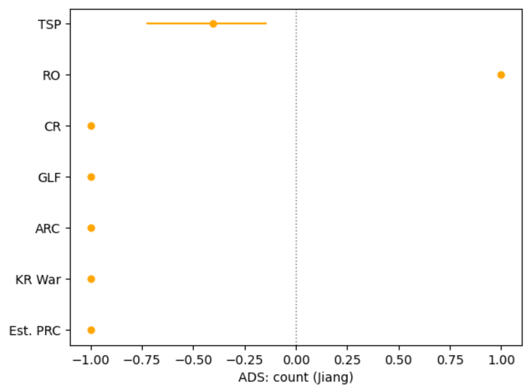

Figure 40: ADS (95% CI included) with difference in (normalized) name count for different events (politician: *Jiang Zemin*, corpus: *Reference News*)

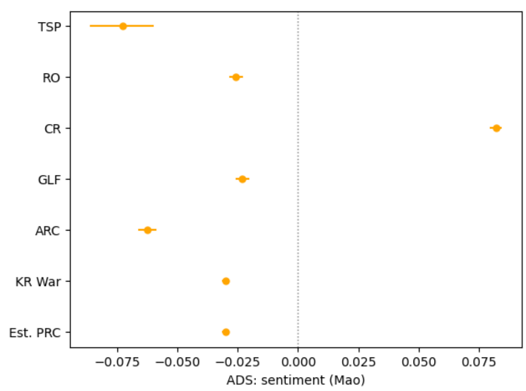

Figure 41: ADS (95% CI included) with difference in sentiment for different events (politician: *Mao Zedong*, corpus: *Reference News*)

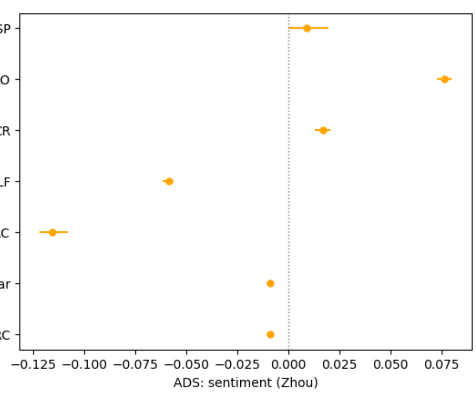

Figure 42: ADS (95% CI included) with difference in sentiment for different events (politician: *Zhou Enlai*, corpus: *Reference News*)

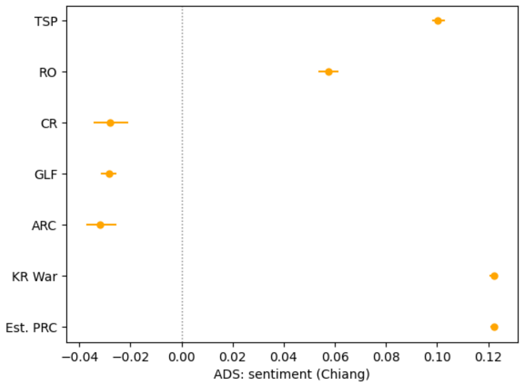

Figure 43: ADS (95% CI included) with difference in sentiment for different events (politician: *Chiang Kai-shek*, corpus: *Reference News*)

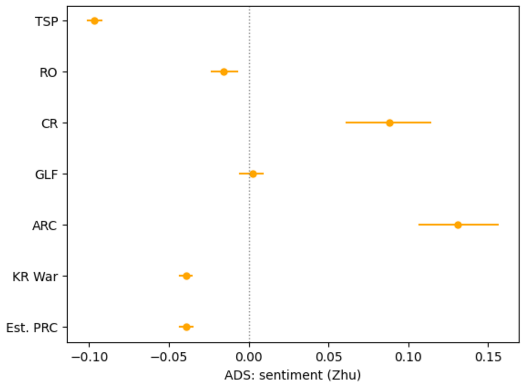

Figure 46: ADS (95% CI included) with difference in sentiment for different events (politician: *Zhu De*, corpus: *Reference News*)

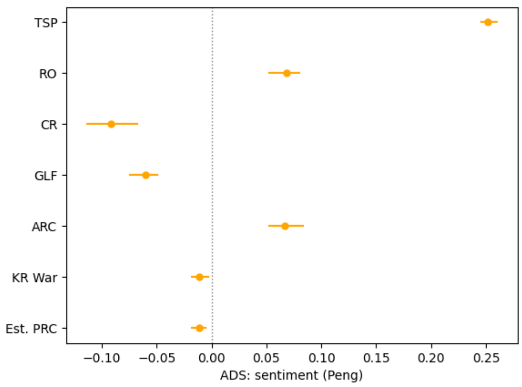

Figure 44: ADS (95% CI included) with difference in sentiment for different events (politician: *Peng Dehuai*, corpus: *Reference News*)

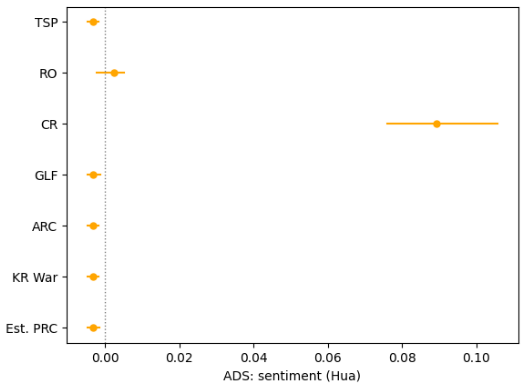

Figure 47: ADS (95% CI included) with difference in sentiment for different events (politician: *Hua Guofeng*, corpus: *Reference News*)

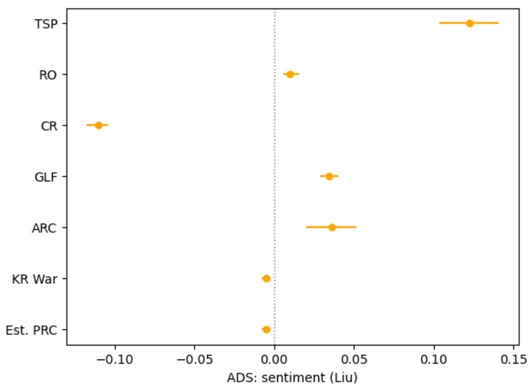

Figure 45: ADS (95% CI included) with difference in sentiment for different events (politician: *Liu Shaoqi*, corpus: *Reference News*)

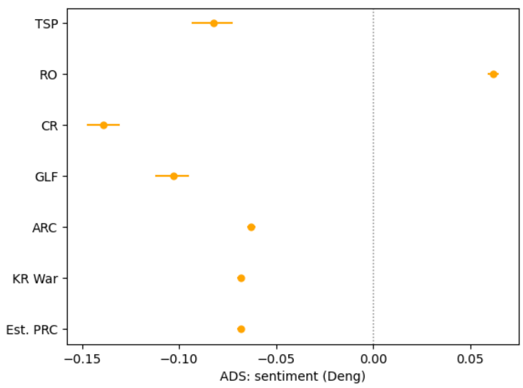

Figure 48: ADS (95% CI included) with difference in sentiment for different events (politician: *Deng Xiaoping*, corpus: *Reference News*)

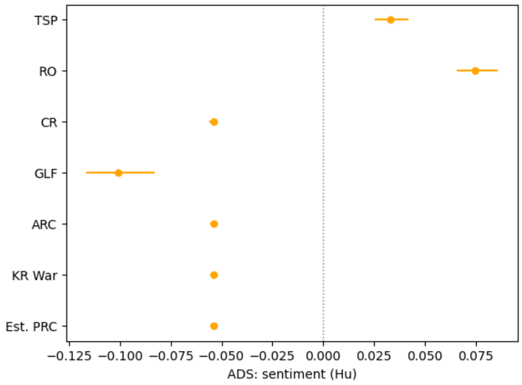

Figure 49: ADS (95% CI included) with difference in sentiment for different events (politician: *Hu Yaobang*, corpus: *Reference News*)

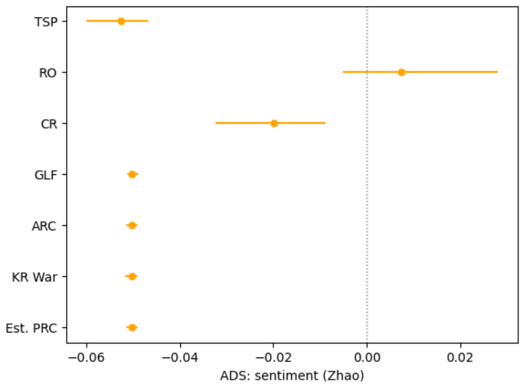

Figure 50: ADS (95% CI included) with difference in sentiment for different events (politician: *Zhao Ziyang*, corpus: *Reference News*)

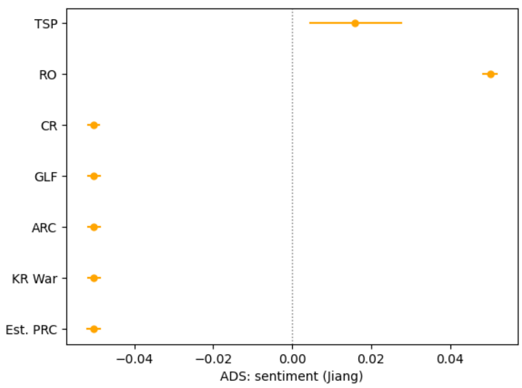

Figure 51: ADS (95% CI included) with difference in sentiment for different events (politician: *Jiang Zemin*, corpus: *Reference News*)

B Change Point Detection Results on *Reference News*

Note that *Reference News* starts to publish after the establishment of PRC so no change point shall be detected for EPRC.

Politician	**CP: linear**	**CP: rbf**
Mao Zedong	CR, DEA	CR
Zhou Enlai	DEA	CR, DEA
Deng Xiaoping	CR	CR
Peng Dehuai	CR	CR, DEA
Liu Shaoqi	DEA	CR
Zhu De	DEA	CR, DEA
Chiang Kai-shek	DEA	CR, DEA

Table 10: Change points detected on normalized name count over time on *Reference News*

Politician	**CP: linear**	**CP: rbf**
Mao Zedong	CR, DEA	CR
Zhou Enlai	DEA	DEA
Deng Xiaoping	CR	CR
Peng Dehuai	CR, DEA	CR, DEA
Liu Shaoqi	DEA	CR, DEA
Zhu De	DEA	DEA
Chiang Kai-shek	DEA	DEA

Table 11: Change points detected on average sentiment over time on *Reference News*

C Additional Results of Topic Modeling

Additional topic modeling results are in following pages:

Decade	**Top 5 Topic Words with Weights**				
1940	China 中国 0.1890	Soviet Union 苏联 0.0536	Delegate 代表 0.0517	Victory 胜利 0.0429	Work 工作 0.0425
1950	China 中国 0.1327	Work 工作 0.0646	Country 国家 0.0560	Revolution 革命 0.0436	Construction 建设 0.0415
1960	Revolution 革命 0.1749	Chairman Mao 毛主席 0.1055	China 中国 0.0971	The Masses 群众 0.0676	Thought 思想 0.0590
1970	Revolution 革命 0.1630	Chairman Mao 毛主席 0.1071	Struggle 斗争 0.0561	The Masses 群众 0.0528	Line 路线 0.0429
1980	China 中国 0.0895	Work 工作 0.0778	Party 党 0.0530	Comrade 同志 0.0489	Development 发展 0.0432
1990	China 中国 0.0937	Construction 建设 0.0756	Work 工作 0.0693	Development 发展 0.0594	Party 党 0.0371
2000	Development 发展 0.1162	Party 党 0.0914	Construction 建设 0.0792	Work 工作 0.0764	China 中国 0.0738

Table 12: Top topic words identified in different decades of articles mentioning **Mao Zedong** in *People's Daily*

Decade	**Top 5 Topic Words with Weights**				
1940	China 中国 0.1432	Delegate 代表 0.0994	Government 政府 0.0766	Meeting 会议 0.0494	KMT 国民党 0.0451
1950	China 中国 0.1665	Country 国家 0.0699	Government 政府 0.0603	Peace 和平 0.0522	Meeting 会议 0.0507
1960	China 中国 0.2592	Government 政府 0.0911	Revolution 革命 0.0694	Republic 共和国 0.0497	Struggle 斗争 0.0474
1970	China 中国 0.1589	Revolution 革命 0.0794	Government 政府 0.0514	Comrade 同志 0.0490	Chairman Mao 毛主席 0.0476
1980	China 中国 0.0952	Work 工作 0.0647	Comrade 同志 0.0482	Development 发展 0.0319	Party 党 0.0285
1990	China 中国 0.0949	Work 工作 0.0541	Development 发展 0.0373	Construction 建设 0.0277	Comrade 同志 0.0259
2000	China 中国 0.1050	Work 工作 0.0449	Development 发展 0.0395	Construction 建设 0.0279	Country 国家 0.0232

Table 13: Top topic words identified in different decades of articles mentioning **Zhou Enlai** in *People's Daily*

Decade	Top 5 Topic Words with Weights				
1940	Liberation Army 解放军 0.0784	Liberation 解放 0.0531	Victory 胜利 0.0476	China 中国 0.0456	Chiang Kai-shek 蒋介石 0.0427
1950	Ethnicity 民族 0.0823	Work 工作 0.0590	China 中国 0.0573	Country 国家 0.0472	Party 党 0.0407
1960	China 中国 0.0999	Revolution 革命 0.0667	CPC Central 中共中央 0.0594	Struggle 斗争 0.0455	Comrade 同志 0.0430
1970	Revolution 革命 0.1073	Chairman Mao 毛主席 0.0859	Struggle 斗争 0.0607	China 中国 0.0513	Line 路线 0.0425
1980	China 中国 0.0795	Work 工作 0.0687	Comrade 同志 0.0548	Construction 建设 0.0509	Development 发展 0.0495
1990	Development 发展 0.1054	Construction 建设 0.1003	Work 工作 0.1001	China 中国 0.0818	Economy 经济 0.0684
2000	Work 工作 0.1189	Development 发展 0.1173	Construction 建设 0.0944	Party 党 0.0798	China 中国 0.0739

Table 14: Top topic words identified in different decades of articles mentioning **Deng Xiaoping** in *People's Daily*

Decade	Top 5 Topic Words with Weights				
1940	Liberation Army 解放军 0.1271	China 中国 0.0929	Liberation 解放 0.0632	Government 政府 0.0613	CPC 中国共产党 0.0374
1950	China 中国 0.0991	Korea 朝鲜 0.0582	The U.S. 美国 0.0520	Peace 和平 0.0460	Country 国家 0.0361
1960	Chairman Mao 毛主席 0.2373	Revolution 革命 0.2059	Mao Zedong Thought 毛泽东思想 0.0826	Proletariat 无产阶级 0.0715	Line 路线 0.0694
1970	Chairman Mao 毛主席 0.1586	Revolution 革命 0.1465	Struggle 斗争 0.0607	Line 路线 0.0564	Party 党 0.0452
1980	Comrade 同志 0.0553	Work 工作 0.0374	China 中国 0.0298	Party 党 0.0278	Revolution 革命 0.0269
1990	Work 工作 0.0552	Comrade 同志 0.0406	Mao Zedong 毛泽东 0.0268	Central 中央 0.0261	Army 部队 0.0242
2000	Work 工作 0.0560	Party 党 0.0381	History 历史 0.0347	China 中国 0.0344	Construction 建设 0.0188

Table 15: Top topic words identified in different decades of articles mentioning **Peng Dehuai** in *People's Daily*

Decade	Top 5 Topic Words with Weights				
1940	China 中国 0.1123	Work 工作 0.0576	Delegate 代表 0.0553	Workers 工人 0.0436	Revolution 革命 0.0405
1950	China 中国 0.1106	Work 工作 0.0793	Country 国家 0.0553	Party 党 0.0419	Delegate 代表 0.0366
1960	Revolution 革命 0.1617	China 中国 0.0899	Chairman Mao 毛主席 0.0876	Thought 思想 0.0476	The Masses 群众 0.0445
1970	Revolution 革命 0.1526	Chairman Mao 毛主席 0.1091	The Masses 群众 0.0735	Production 生产 0.0682	Struggle 斗争 0.0554
1980	Work 工作 0.0632	China 中国 0.0506	Party 党 0.0495	Comrade 同志 0.0481	Revolution 革命 0.0366
1990	China 中国 0.0683	Work 工作 0.0468	Party 党 0.0348	Comrade 同志 0.0319	Construction 建设 0.0288
2000	China 中国 0.0783	Party 党 0.0335	Work 工作 0.0327	Development 发展 0.0306	Construction 建设 0.0292

Table 16: Top topic words identified in different decades of articles mentioning **Liu Shaoqi** in *People's Daily*

Decade	Top 5 Topic Words with Weights				
1940	China 中国 0.0949	Chiang Kai-shek 蒋介石 0.0836	The U.S. 美国 0.0506	The Masses 群众 0.0377	Government 政府 0.0303
1950	China 中国 0.1334	The U.S. 美国 0.1189	Country 国家 0.0490	Government 政府 0.0464	Taiwan 台湾 0.0360
1960	China 中国 0.1282	The U.S. 美国 0.0908	Revolution 革命 0.0803	US Imperialism 美帝国主义 0.0545	Struggle 斗争 0.0536
1970	Revolution 革命 0.1096	China 中国 0.1057	Chairman Mao 毛主席 0.0763	Struggle 斗争 0.0504	Country 国家 0.0273
1980	China 中国 0.0644	Revolution 革命 0.0603	Comrade 同志 0.0520	Work 工作 0.0435	Party 党 0.0336
1990	China 中国 0.1031	Revolution 革命 0.0345	Work 工作 0.0340	Comrade 同志 0.0291	History 历史 0.0235
2000	China 中国 0.0789	Construction 建设 0.0300	Work 工作 0.0268	History 历史 0.0260	Revolution 革命 0.0230

Table 17: Top topic words identified in different decades of articles mentioning **Chiang Kai-shek** in *People's Daily*

Decade	Top 5 Topic Words with Weights				
1950	China 中国 0.2894	Communist Party 共产党 0.1404	The U.S. 美国 0.0686	Country 国家 0.0644	Soviet Union 苏联 0.0513
1960	China 中国 0.3125	Revolution 革命 0.0939	Communist Party 共产党 0.0741	Struggle 斗争 0.0518	Soviet Union 苏联 0.0478
1970	China 中国 0.3283	The U.S. 美国 0.0687	Beijing 北京 0.0664	Revolution 革命 0.0569	Country 国家 0.0568
1980	China 中国 0.3062	Country 国家 0.0408	Economy 经济 0.0354	Beijing 北京 0.0337	The U.S. 美国 0.0296
1990	China 中国 0.2581	The U.S. 美国 0.0468	Beijing 北京 0.0352	Economy 经济 0.0349	Country 国家 0.0336
2000	China 中国 0.2247	The U.S. 美国 0.0648	Beijing 北京 0.0349	Country 国家 0.0297	Taiwan 台湾 0.0271

Table 18: Top topic words identified in different decades of articles mentioning **Mao Zedong** in *Reference News*

Decade	Top 5 Topic Words with Weights				
1950	China 中国 0.3424	Communist Party 共产党 0.1327	The U.S. 美国 0.0923	India 印度 0.0860	Government 政府 0.0673
1960	China 中国 0.3813	Beijing 北京 0.0860	India 印度 0.0841	Communist Party 共产党 0.0804	Soviet Union 苏联 0.0633
1970	China 中国 0.3561	Beijing 北京 0.1096	The U.S. 美国 0.0941	Japan 日本 0.0630	Visit 访问 0.0583
1980	China 中国 0.2948	The U.S. 美国 0.0516	Country 国家 0.0450	Relationship 关系 0.0390	Japan 日本 0.0388
1990	China 中国 0.2152	Japan 日本 0.0536	The U.S. 美国 0.0496	Beijing 北京 0.0374	Country 国家 0.0297
2000	China 中国 0.1563	The U.S. 美国 0.0571	Country 国家 0.0276	Japan 日本 0.0230	Beijing 北京 0.0227

Table 19: Top topic words identified in different decades of articles mentioning **Zhou Enlai** in *Reference News*

Decade	Top 5 Topic Words with Weights				
1950	China 中国 0.3266	Communist Party 共产党 0.2128	Beijing 北京 0.0956	Mao Zedong 毛泽东 0.0876	Policy 政策 0.0770
1960	China 中国 0.2183	Beijing 北京 0.1140	Soviet Union 苏联 0.0806	Communist Party 共产党 0.0698	Revolution 革命 0.0624
1970	China 中国 0.3768	The U.S. 美国 0.1075	Visit 访问 0.0728	Beijing 北京 0.0726	Soviet Union 苏联 0.0690
1980	China 中国 0.4115	The U.S. 美国 0.0661	Beijing 北京 0.0610	Economy 经济 0.0591	Relationship 关系 0.0493
1990	China 中国 0.3705	Economy 经济 0.1291	Hong Kong 香港 0.0708	Development 发展 0.0635	Reform 改革 0.0616
2000	China 中国 0.2442	The U.S. 美国 0.0674	Economy 经济 0.0484	Taiwan 台湾 0.0460	Shanghai 上海 0.0341

Table 20: Top topic words identified in different decades of articles mentioning **Deng Xiaoping** in *Reference News*

Decade	Top 5 Topic Words with Weights				
1950	China 中国 0.2227	The U.S. 美国 0.2070	KMT 国民党 0.1171	Taiwan 台湾 0.0862	Government 政府 0.0752
1960	China 中国 0.2707	The U.S. 美国 0.1523	Government 政府 0.0729	KMT 国民党 0.0697	Taiwan 台湾 0.0666
1970	China 中国 0.2586	The U.S. 美国 0.1599	Taiwan 台湾 0.1269	Japan 日本 0.0825	Government 政府 0.0688
1980	China 中国 0.1342	Taiwan 台湾 0.1127	KMT 国民党 0.0539	The U.S. 美国 0.0518	Beijing 北京 0.0324
1990	China 中国 0.1109	Taiwan 台湾 0.0934	The U.S. 美国 0.0790	Japan 日本 0.0467	KMT 国民党 0.0434
2000	Taiwan 台湾 0.1256	The U.S. 美国 0.0747	China 中国 0.0727	Japan 日本 0.0562	KMT 国民党 0.0386

Table 21: Top topic words identified in different decades of articles mentioning **Chiang Kai-shek** in *Reference News*

D Additional Results of Predicting Masked Names

The following pages show results of masked name predictions for different politicians:

Politician	Avg Probability
Chiang Kai-shek (蒋介石)	0.0650
Li Zongren (李宗仁)	0.0506
Chen Lifu (陈立夫)	0.0492
Mao Zedong (毛泽东)	0.0474
Zhang Zhizhong (张治中)	0.0247

Table 22: Top 5 politicians' names generated by BERT to replace the masked *Chiang Kai-shek's* name in articles from *People's Daily*

Politician	Avg Probability
Li Xiannian (李先念)	0.0543
Mao Zedong (毛泽东)	0.0486
Chen Yun (陈云)	0.0448
Peng Dehuai (彭德怀)	0.0443
Zhou Enlai (周恩来)	0.0393

Table 23: Top 5 politicians' names generated by BERT to replace the masked *Peng Dehuai's* name in articles from *People's Daily*

Politician	Avg Probability
Mao Zedong (毛泽东)	0.0757
Liu Shaoqi (刘少奇)	0.0679
Li Xiannian (李先念)	0.0500
Zhou Enlai (周恩来)	0.0454
Chen Yun (陈云)	0.0408

Table 24: Top 5 politicians' names generated by BERT to replace the masked *Liu Shaoqi's* name in articles from *People's Daily*

Politician	Avg Probability
Zhu De (朱德)	0.0768
Zhou Enlai (周恩来)	0.0673
Li Xiannian (李先念)	0.0562
Chen Yun (陈云)	0.0551
Mao Zedong (毛泽东)	0.0512

Table 25: Top 5 politicians' names generated by BERT to replace the masked *Zhu De's* name in articles from *People's Daily*

Politician	Avg Probability
Li Xiannian (李先念)	0.0789
Hua Guofeng (华国锋)	0.0586
Mao Zedong (毛泽东)	0.0506
Chen Yun (陈云)	0.0342
Zhou Enlai (周恩来)	0.0337

Table 26: Top 5 politicians' names generated by BERT to replace the masked *Hua Guofeng's* name in articles from *People's Daily*

Politician	Avg Probability
Deng Xiaoping (邓小平)	0.1066
Mao Zedong (毛泽东)	0.0845
Li Xiannian (李先念)	0.0554
Zhou Enlai (周恩来)	0.0447
Chen Yun (陈云)	0.0378

Table 27: Top 5 politicians' names generated by BERT to replace the masked *Deng Xiaoping's* name in articles from *People's Daily*

Politician	Avg Probability
Hu Yaobang (胡耀邦)	0.0625
Li Peng (李鹏)	0.0561
Chen Yun (陈云)	0.0520
Zhou Enlai (周恩来)	0.0321
Yang Shangkun (杨尚昆)	0.0294

Table 28: Top 5 politicians' names generated by BERT to replace the masked *Hu Yaobang's* name in articles from *People's Daily*

Politician	Avg Probability
Li Peng (李鹏)	0.0716
Chen Yun (陈云)	0.0541
Zhao Ziyang (赵紫阳)	0.0341
Zhou Enlai (周恩来)	0.0294
Yang Shangkun (杨尚昆)	0.0292

Table 29: Top 5 politicians' names generated by BERT to replace the masked *Zhao Ziyang's* name in articles from *People's Daily*

Politician	**Avg Probability**
Li Peng (李鹏)	0.0585
Jiang Zemin (江泽民)	0.0577
Xi Jinping (习近平)	0.0565
Hu Jintao (胡锦涛)	0.0368
Chen Yun (陈云)	0.0300

Table 30: Top 5 politicians' names generated by BERT to replace the masked *Jiang Zemin's* name in articles from *People's Daily*

E Results of Predicting Masked Names during Different Events

E.1 Masking *Peng Dehuai*

Politician	Avg Probability
Li Xiannian (李先念)	0.0628
Chen Yun (陈云)	0.0559
Liu Shaoqi (刘少奇)	0.0335
Yang Shangkun (杨尚昆)	0.0304
Zhou Enlai (周恩来)	0.0256

Table 31: Top 5 politicians' names generated by BERT to replace the masked *Peng Dehuai's* name in articles from *People's Daily* before the Establishment of PRC

Politician	Avg Probability
Kim Il-sung (金日成)	0.0909
Li Xiannian (李先念)	0.0887
Yang Shangkun (杨尚昆)	0.0323
Peng Dehuai (彭德怀)	0.0267
Liu Shaoqi (刘少奇)	0.0233

Table 32: Top 5 politicians' names generated by BERT to replace the masked *Peng Dehuai's* name in articles from *People's Daily* during the Korean War

Politician	Avg Probability
Li Xiannian (李先念)	0.0523
Peng Dehuai (彭德怀)	0.0429
Chen Yun (陈云)	0.0323
Yang Shangkun (杨尚昆)	0.0335
Zhou Enlai (周恩来)	0.0304

Table 33: Top 5 politicians' names generated by BERT to replace the masked *Peng Dehuai's* name in articles from *People's Daily* during the Anti-rightist Campaign

Politician	Avg Probability
Peng Dehuai (彭德怀)	0.0796
Li Xiannian (李先念)	0.0511
Chen Yun (陈云)	0.0362
Yang Shangkun (杨尚昆)	0.0361
Zhou Enlai (周恩来)	0.0309

Table 34: Top 5 politicians' names generated by BERT to replace the masked *Peng Dehuai's* name in articles from *People's Daily* during the Great Leap Forward

Politician	Avg Probability
Mao Zedong (毛泽东)	0.0785
Li Xiannian (李先念)	0.0448
Lin Biao (林彪)	0.0426
Chen Yun (陈云)	0.0372
Liu Shaoqi (刘少奇)	0.0273

Table 35: Top 5 politicians' names generated by BERT to replace the masked *Peng Dehuai's* name in articles from *People's Daily* during the Cultural Revolution

Politician	Avg Probability
Peng Dehuai (彭德怀)	0.0785
Li Xiannian (李先念)	0.0564
Chen Yun (陈云)	0.0553
Zhou Enlai (周恩来)	0.0505
Yang Shangkun (杨尚昆)	0.0364

Table 36: Top 5 politicians' names generated by BERT to replace the masked *Peng Dehuai's* name in articles from *People's Daily* after the Reform and Opening-up

Politician	Avg Probability
Zhou Enlai (周恩来)	0.1166
Lin Biao (林彪)	0.0998
Li Xiannian (李先念)	0.0837
Chen Yun (陈云)	0.0805
Yang Shangkun (杨尚昆)	0.0556

Table 37: Top 5 politicians' names generated by BERT to replace the masked *Peng Dehuai's* name in articles from *People's Daily* during the Tiananmen Square Protest

E.2 Masking *Liu Shaoqi*

Politician	Avg Probability
Zhou Enlai (周恩来)	0.0628
Liu Shaoqi (刘少奇)	0.0523
Li Xiannian (李先念)	0.0501
Mao Zedong (毛泽东)	0.0493
Chen Yun (陈云)	0.0491

Table 38: Top 5 politicians' names generated by BERT to replace the masked *Liu Shaoqi's* name in articles from *People's Daily* before the Establishment of PRC

Politician	Avg Probability
Li Xiannian (李先念)	0.0685
Zhou Enlai (周恩来)	0.0585
Chen Yun (陈云)	0.0509
Liu Shaoqi (刘少奇)	0.0400
Mao Zedong (毛泽东)	0.0395

Table 39: Top 5 politicians' names generated by BERT to replace the masked *Liu Shaoqi's* name in articles from *People's Daily* during the Korean War

Politician	Avg Probability
Liu Shaoqi (刘少奇)	0.0591
Li Xiannian (李先念)	0.0554
Zhou Enlai (周恩来)	0.0467
Mao Zedong (毛泽东)	0.0389
Chen Yun (陈云)	0.0385

Table 40: Top 5 politicians' names generated by BERT to replace the masked *Liu Shaoqi's* name in articles from *People's Daily* during the Anti-rightist Campaign

Politician	Avg Probability
Liu Shaoqi (刘少奇)	0.1408
Zhou Enlai (周恩来)	0.0832
Mao Zedong (毛泽东)	0.0779
Li Xiannian (李先念)	0.0563
Hu Yaobang (胡耀邦)	0.0301

Table 41: Top 5 politicians' names generated by BERT to replace the masked *Liu Shaoqi's* name in articles from *People's Daily* during the Great Leap Forward

Politician	Avg Probability
Mao Zedong (毛泽东)	0.1334
Liu Shaoqi (刘少奇)	0.0378
Li Xiannian (李先念)	0.0374
Chen Yun (陈云)	0.0313
Deng Xiaoping (邓小平)	0.0260

Table 42: Top 5 politicians' names generated by BERT to replace the masked *Liu Shaoqi's* name in articles from *People's Daily* during the Cultural Revolution

Politician	Avg Probability
Liu Shaoqi (刘少奇)	0.1089
Mao Zedong (毛泽东)	0.0530
Li Xiannian (李先念)	0.0499
Zhou Enlai (周恩来)	0.0448
Chen Yun (陈云)	0.0411

Table 43: Top 5 politicians' names generated by BERT to replace the masked *Liu Shaoqi's* name in articles from *People's Daily* after the Reform and Opening-up

Politician	Avg Probability
Zhou Enlai (周恩来)	0.1026
Li Xiannian (李先念)	0.0817
Chen Yun (陈云)	0.0637
Mao Zedong (毛泽东)	0.0485
Lin Biao (林彪)	0.0407

Table 44: Top 5 politicians' names generated by BERT to replace the masked *Liu Shaoqi's* name in articles from *People's Daily* during the Tiananmen Square Protest