

notebook_with_ref

December 13, 2023

1 Mapping the Transnational Public Sphere in modern China: A Bilingual Topic Modeling of the Republican Press (1919-1949)

1.0.1 Anonym

©. Published by De Gruyter in cooperation with the University of Luxembourg Centre for Contemporary and Digital History. This is an Open Access article distributed under the terms of the [Creative Commons Attribution License CC-BY-NC-ND](#)

This research has received funding from the Chiang Ching-kuo Foundation for International Scholarly Exchange (Project n°RG004-U-21) and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 788476).

modern China, elites, historical newspapers, topic modeling, Sino-American networks

This paper analyzes the formation of a transnational public sphere in China through the joint empirical study of two key institutions – a transnational organization (the Rotary Club) and the Shanghai periodical press, which has been recognized as the first mass medium for disseminating information in modern China. Methodologically, we rely on structural topic modeling (STM) to harness large corpora of digitized newspapers from a bilingual (Chinese/English) and dynamic perspective. We show that STM has the potential to encompass both the discursive and social aspects of the public sphere and to bridge the gap between previous theoretical and empirical studies. Thanks to STM, we are able to precisely classify and quantify the four main domains of the public sphere (meeting, self-organization, forum, community service) and to model their local and global dimensions. Furthermore, we reveal significant convergences between languages and a growing awareness of global issues during the troubled decades of the Republic (1919-1949). Methodologically, the contribution of this paper is threefold. First, we offer a simple yet efficient solution, based on concordance, to the problem of article segmentation in digitized newspapers. Second, this paper constitutes a rare instance of cross-lingual comparison involving a low-resource language, namely, the Chinese language during its transitional stage between classical and modern Chinese. Third, we demonstrate the value of using a multi-model approach to navigate between different levels of granularity, especially when dealing with corpora of different sizes and structures.

1.1 Introduction

The Chinese people had developed, the speaker quoted Dr. Wang as stating, the solidarity of family life. Its welfare was the summum bonum of one's life – it was what the individual was to live for. The word “family” in China not merely mean, as in the West, the father and mother and their children: it was all inclusive and extended often to the whole clan. But, just as the family life was developed, so the community life was starved. The individual Chinese until recently and only

in limited sense, always thought and acted in terms of the family. Therefore, the best things one formed in any city were family property: a family library; a family park; a family hall; a family art treasury, etc. It accounted in very substantial degree, for the lack of public spirit, and public service in Chinese community life. “I cannot think of any other agency as potential as our Rotary clubs to generate this spirit of public service,” he said. (The North China Herald 1935)

In his speech given at the Rotary Inter-City meeting held in Qingdao in 1935, the diplomat and political leader Wang Zhengting called for the formation of a modern public sphere, which he opposed to the traditional, family-centered vision of Chinese society. This excerpt, which was reprinted in several English and Chinese newspapers, raises three fundamental questions. To what extent did Wang’s idealism resonate with broader perceptions and practices of social organization and public discussion? How did the 1935 conference of the Rotary Club situate within the wider range of events and organizations which contributed to shaping the public sphere in modern China? How did the various historical actors used the press to disseminate their ideas and publicize their actions to a broader audience? The question of the emergence of a public sphere in China is an old debate. Early studies in the early 1990s were essentially concerned with theoretical issues, especially the transferability of Western concepts to China (*Huang 1993, Wakeman 1993, Rowe 1990, Metzger 1998, Zhongguo shehuikexue jikan 1993*, reprinted in *Deng Zhenglai 2021*). On the other hand, empirical studies of civil society organizations, such as local elites in Zhejiang or civil protest in Beijing, sought to illuminate how the public sphere was implemented in practice, but they were restricted to specific local settings (*Rankin 1990, Strand 1990*). Since the 2000s, the increasing accessibility of newspapers collections has opened new directions for the study of “public opinion” and has led to a new emphasis on the transnational dimension of the Chinese public sphere (*Wagner 2007, Goodman 2004, Wagner 1995*). This transnational perspective, which typically focuses on non-state actors and emphasizes circulations and connections across national boundaries, represents a direct challenge to the mainstream narrative of imperialism vs. nationalism, and to the China-centered approach that has dominated the scholarship during the last three decades. Scholars who have used the press to support their investigation of the public sphere have tended to focus on single publications, especially those published during the late Qing dynasty and the early Republic, and they have usually restricted their analysis to specific genres of articles, such as editorials or “free talk” columns (See for instance: (*Judge 1996*), *Veg 2021, Tang Xiaobing 2012*). Constrained by the limits of human reading capability, they have relied on small samples of manually selected articles, without providing the possibility to contextualize their findings and to assess whether/to what extent the selected texts or passages were representative of larger trends. To alleviate the risk of overstating or simply reinforcing prior assumptions based on elite views like Wang’s binary conceptualization of Chinese society, we need a different approach.

This paper brings the debate to a new level through the joint empirical study of two key institutions – a transnational organization, the Rotary Club, and the Shanghai press, which has long been recognized as the first mass medium for disseminating information in modern China, and which constituted the very venue through which non-state organizations like the Rotary Club shaped their social and discursive practices. In this paper, the public sphere is defined as “the space in which state and society, as well as different segments of society, articulate their interests and opinions within culturally and historically defined rules of rationality and propriety”. Following this definition, the public sphere has been conceived as an alternative to armed conflicts, “a key constituent of a social order whose members do not resort to violence in each instance when conflict occurs”. Furthermore, as Rudolf Wagner emphasized, the modern public sphere is not coterminous with the nation but is essentially transnational and international (*Wagner 2007, p.3*).

The Rotary Club perfectly matches with these criteria. One the one hand, it was a non-state organization aimed at the socialization and legitimization of business and professional elites. On the other hand, it was a truly transnational organization. Initially founded in Chicago in 1905, the Rotary expanded all over the world after World War I. Before the Communist revolution in 1949, there were no less than thirty-three clubs in China. Despite its elite nature, the Rotary was strongly committed to helping the local society through various philanthropic endeavors. At the same time, because of its foreign origins and its multinational membership, the Rotary Club was deeply concerned with international politics. Despite its self-claimed apolitical posture, it was dedicated to the promotion of international peace and goodwill between nations, especially in the troubled context of the interwar years (*Armand 2022*, p.233). It is these very tensions between business/politics, elite/society, and localism/internationalism that this paper seeks to investigate.

The second public sphere institution examined in this research is the Shanghai press. The general press represents an indispensable source for the history of modern China. It recorded the minute details of everyday life and linked the local with the global. As the first mass medium, the periodical press became an important venue through which a variety of historical actors shaped their public images and tried to influence public opinion. After the Opium wars (1839-), Shanghai became the uncontested center of the emerging modern press and the publishing industry in China. The foreign settlements established by the Nanjing Treaty (1842) served as exclaves facilitating the development of a free press beyond the control of the Chinese authorities (*Wagner 2007*, p.4). Increasingly dependent on advertising revenues, the so-called “commercial” press, of which the *Shenbao* and *North China Herald/Daily News* can be regarded as the most successful representatives, boomed during the Republic (*Armand , forthcoming*). These publications presented a broader swathe of opinions than intellectual pamphlets and political periodicals which flourished during the late years of the Qing dynasty, the New Culture movement, and the May Fourth movement. Fortunately, a significant proportion of the Shanghai press has been exceptionally well preserved until today. The leading periodicals published in Shanghai have been extensively digitized and recently made available in plain text, which creates the conditions for conducting quantitative and qualitative analyses that were not possible before.

Taking advantage of the massive, multilingual corpora recently made available in full text by the [ENP-China project](#) (Elites, Networks and Power in modern China), this paper relies on a computational technique called “topic modeling” to introduce a change of scale in the analysis of the historical press. These corpora include dozens of periodicals in both Chinese and English, representing thousands of pages and billions of tokens, spanning over a century (1850s-1950s). Topic modeling is a statistical method aimed at detecting hidden themes (topics) in large collections of unstructured texts, based on the co-occurrences of words in documents. We argue that this method is particularly suited for modeling the public sphere in the Shanghai press. While grounded on natural language, topic modeling is also scalable and can be applied to corpora of virtually any size. It can be conceived as a mixed-method approach that harnesses the power of unsupervised techniques to minimize the biases inherent in manual reading and subjective selection, while still relying on close reading and contextual knowledge to interpret the automatically inferred topics. Based on this hybrid approach, we can eventually bridge the gap between close and distant readings, quantitative and qualitative analyses, theoretical and empirical studies.

Historians have used topic modeling for a variety of purposes, such as refining corpus building, classifying documents, mapping discourse dynamics and ideational changes (*Oberbichler, Pfanzelter 2021, Andersen, Biryukov, Kalyakin, Wieneke 2020, Hengchen, Coeckelbergs, van Hooland, Verborgh, Steiner 2016, Bunout 2019, Yang, Torget, Mihalcea 2011, Indukaev 2021, Marjanen,*

Zosa, Hengchen, Pivovarova, Tolonen 2020). While this method has been increasingly used to harness the growing body of digitized newspapers, few studies have explicitly addressed the challenge of article separation. In most cases, the processing of digitized newspapers has been done at the page level, which barely constitutes a semantically coherent unit. Even when the layout analysis was performed at a lower level, specific sections and categories of articles such as “Men and Events” or “Brevities” may contain a range of information that have no relation with each other (figure 1). The topics inferred from these heterogenous articles will only reflect their messy contents, without being immediately relevant to our specific research question. To the best of our knowledge, only (*Yang, Torget, Mihalcea 2011*) and (*Van Galen, Nicholson 2018*) have addressed the crucial issue of article separation in historical newspapers, yet without offering scalable solutions. (*Eads, Schofield, Mahootian, Mimno, Wilderom 2021*) have also touched upon the issue of determining the size of text sections in more general terms. Until recently, topic modeling has been essentially implemented in monolingual contexts, primarily in English. Notable exceptions that apply multilingual topic modeling include (*Zosa, Granroth-Wilding 2019*), (*Mimno, Wallach, Naradowsky, Smith, McCallum 2009*), (*Yang, Boyd-Graber, Resnik 2019*) and (*Maier, Baden, Stoltzenberg, De Vries-Kedem, Waldherr 2022*). Yet because of the lack of training data, multilingual topic modeling has not been applied to historical newspapers so far. Topic modeling has also been applied to Chinese texts, but the existing case studies have focused on either contemporary or classical Chinese, especially classical literature (*Allen, Luo, Murdock, Pu, Wang, Zhai, Zhao 2017*) and in rare instances, historical records written in classical Chinese (*Miller 2013*). Due to the limited availability of digitized texts produced during the modern period, topic modeling has never been applied to modern Chinese newspapers so far.

```
[6]: library("IRdisplay")
display_png(file="./media/Fig.1a.png")
display_png(file="./media/Fig.1b.png")
```

Example document 1: *Shenbao*, 29 February 1936 (DocId: SPSP193602290401)

(...) 福州 省農村合作委員會為救濟仙遊糖產起見、特劃該縣為實施救濟縣份、貸款復典糖產、並令駐仙辦事處視察輔導、將來蔗糖運銷費用、即完全以蔗糖向銀行抵押借款、俟銷出後付返、藉使糖業發展、不致受困。 **福州 王正廷二十八日抵省談、此來純為推進扶輪社務、下午四時召開社董聯會、晚赴美領署應全社公宴、並演講、定一日返滬。** 福州 坊口利東匯分向永泰平潭二處逃逸、水陸軍已取四面包圍勢、被綑旅客傳囚於福清海島、正探救中、 廈門 日軍令部第三部長高須少將、二十八日乘福建丸經汕赴粵(...)

Total length of “article”: 9169 characters

Length of targeted segment: 56 characters

Example document 2: “Men and Events,” *The China Weekly Review*, 30 October 1926 (DocId: 1319907286)

Men and Events (...) Dr and Mrs Skevington are making tour of the Orient The Union Club of China is planning dinner and dance for Saturday evening Oct 30 in honor of Colin Scott of Butterfield and Swire In addition to this function the Club will hold its regular monthly dinner dance on the evening of Nov 2. daughter Pearl Adelaide was born to the Rev and Mrs II Swenson at Kuyuan Kansu Province on Oct 3, 192(>. Mr and Mrs Swenson are members of the Scandinavian Alliance Mission of North America attached to the Kansu station Leon Friedman general manager of China Motors Inc sailed on Oct 22 for Europe on vacation trip which will extend for several months After traveling through Europe Mr riedman will return to China next spring by way of the L'nted States Dr Myers of the faculty of the College of Agriculture Nanking University addressed the Shanghai Rotary Club on Thursday Oct 21 on the subject of Plant-Breeding experiments and developments at the College of Agriculture at Xanking Harry Gibson for years connected with the 1'oster-McClellan Company of Shanghai has resigned and has been succeeded by Dudley Ward-Smith who will have charge of all branches of the conn any in China Strike Settlements and Dutch East Indies Dr Anne Walter Iearn who is closing her hospital and residence at 30 Route Pichon has presented her grand piano to the Shanghai American School Students and faculty members gave special recital on I'riday morning Oct 22 when the gift was formally accepted (...)

Total length of “article”: 2091 words (12 467 characters)

Length of targeted segment: 38 words (230 characters)

The purpose of this research is twofold. Substantively, we aim to better understand how the public sphere operated in practice through the daily working of a non-state organization (Rotary Club), as it was recorded in the periodical press. Our key questions include: How did the Shanghai press reported on the Rotary Club? How did the organization mediate between elite/society, business/politics, localism/internationalism? What does this reveal about how the press itself functioned as a public sphere? How did the emerging public sphere change over time and vary across languages? Methodologically, we aim to design a robust, reliable method for conducting a bilingual, dynamic topic modeling approach of the historical press. More specifically, we address three major challenges: (1) to identify topics across multiple languages (in this paper, English and Chinese); (2) to trace topical changes over time; and (3) to adapt topic modeling to the heterogeneity of newspaper contents, particularly to brevity-style articles made up of short pieces of unrelated news.

This paper proceeds in three steps. The first section describes the corpora and the methodology used for preparing the text data and building the topic models. The two next sections exploit the results of topic modeling to map out the four modus operandi of the public sphere and to analyze how its local and global dimensions were articulated in the press across languages. Finally, the last section examines how the public sphere evolved during the troubled decades extending from the end of World War I to the Communist revolution (1919-1949).

START HERMENEUTICS ## Data and methodology END HERMENEUTICS

START HERMENEUTICS Our workflow follows three main steps. First, we create the text corpora. Second, we prepare the text data and build the topic models. Finally, we analyze and interpret the topics. While this is a standard workflow in any topic modeling approach, this research introduces specific adaptations for building the bilingual corpora and for pre-processing the text data to overcome the problem of article separation described earlier.

END HERMENEUTICS

1.1.1 Step 1: Corpus building

In the first step, we created two separate corpora, one for each language. The Chinese-language corpus is based on the *Shenbao*, a leading newspaper published in Shanghai between 1872 and 1949. Despite low literacy rates among the Chinese population, the *Shenbao* reached 150,000 copies in the 1930s, making it one of the two most widely circulated newspapers in China (On the *Shenbao*, see (*Mittler 2004*), (*Tsai 2008*), (*Liu Li 2014*). See also Christian Henriot's piece in this issue). Although it catered primarily to Shanghai intellectual, political, and business elites, its readership widened in the 1930s. The English-language corpus is based on the [ProQuest Chinese Newspapers Collection](#), which includes a dozen of periodicals running from 1832 to 1953, with varying circulation, periodicity, and duration. Despite the risk of overlap, we nonetheless chose to include all the periodicals included in the ProQuest collection to ensure the maximum coverage and to reduce the gap with the daily granularity of the *Shenbao*. Since the largest English periodicals were weekly publications, they were necessarily more selective than the *Shenbao*, with the risk of missing some events that were reported in the *Shenbao*, but not in one or the other English periodicals. The most widely distributed foreign periodical was the *North China Daily News*, peaking at 10,000 copies in the early 1930s, while its weekly edition, the *North China Herald*, distributed almost 2,000 copies per week. They were read not only by foreign expatriates but also increasingly by foreign-educated Chinese elites. Although they were printed in Shanghai, both the *Shenbao* and the *North China Herald* had a national, even international coverage, being circulated among overseas Chinese and foreign readers interested in Chinese affairs.

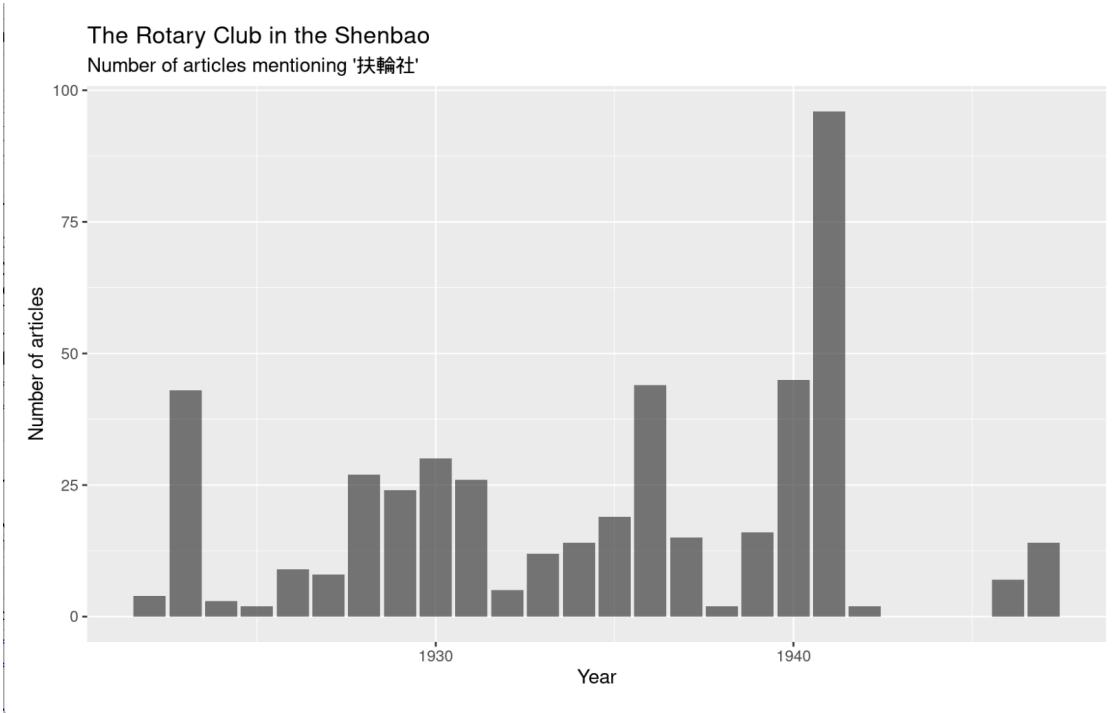
START HERMENEUTICS Since we are investigating a very specific organization with few possible homonyms and a low degree of ambiguity, we relied on simple keywords to build the corpora, namely “Rotary Club” in English and “ ” (*fulunshe*) in Chinese. We only excluded the quasi-homonym “ ” (*guoxue fulunshe*), which referred to a publishing enterprise established in the early 20th century with no connection with the Rotary Club. Additionally, we restricted the query to the period posterior to 1919, when the first Rotary Club in China was established in Shanghai. For the English-language corpus, we used the metadata to filter out irrelevant or content-poor categories of articles, such as “Advertisement”, “Classified Advertisement”, “General Information”, and “Table of Contents, Front Matter”. The full code for building the corpora with the [HistText R library](#) can be found [here](#) (Chinese) and [here](#) (English).

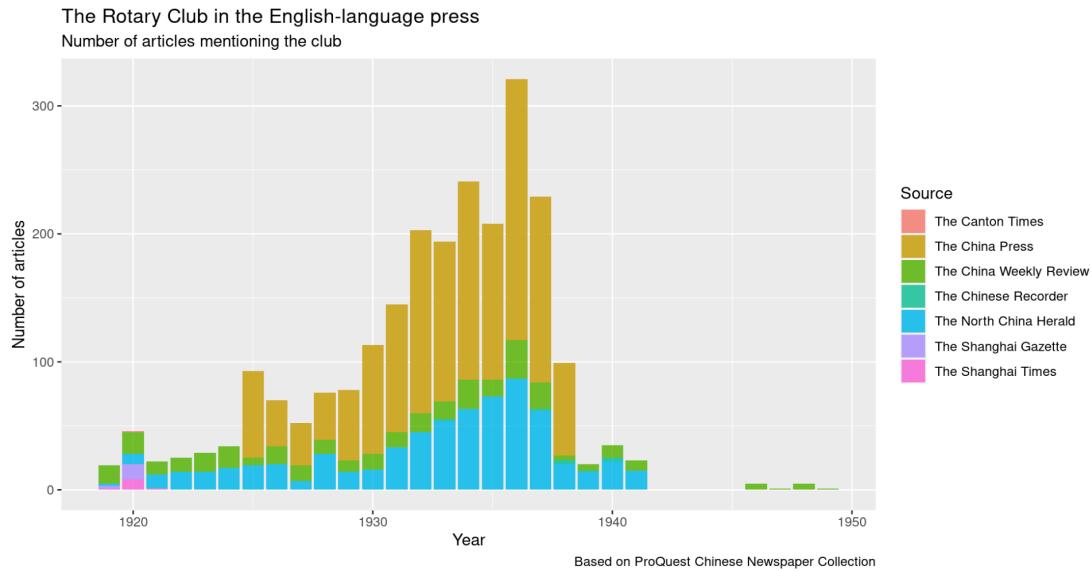
END HERMENEUTICS

START HERMENEUTICS Following this method, we obtained 467 documents spanning from 1922 to 1947 for the Chinese-language corpus (*Shenbao*) and 2387 documents published between 1919 and 1948 for the English-language corpus (ProQuest). The two plots below show the uneven distribution of documents over time in the two corpora (figure 2). For both, we observe a huge gap during the Sino-Japanese war (1937-1945), especially after 1941 when most periodicals ceased publication. The ProQuest corpus shows a clear peak in the 1930s, whereas the distribution was

less regular in the *Shenbao*. In addition, the ProQuest corpus was clearly dominated by three large periodicals, namely the *China Press* (Chinese-owned) (1380, 58 %), the *North China Herald* (British) (658, 28 %), and to a lesser extent, the *China Weekly Review* (American) (315, 13 %). Obviously, the two corpora are not balanced. Not only is English-language corpus larger than the Chinese one, but it contains several periodicals. This is not an issue, however, since the topics are inferred separately for each corpus based on their respective vocabulary. A major advantage of LDA-based probabilistic models, as we explained below, is to rely on the relative frequencies of words in each topic and on the relative proportions of topics in each corpus, which enables to compare corpora of their different sizes and structures. END HERMENEUTICS

```
[7]: display_png(file="./media/Fig.2a.png")
display_png(file="./media/Fig.2b.png")
```





1.1.2 Step 2: Text pre-processing

START HERMENEUTICS To address the challenge of article separation, we propose to use text segments, instead of entire documents, as the basic unit for inferring topics. As shown on the figure below, we designed a simple, three-step method based on Key Word in Context (KWIC) (figure 3). In the first step, we retrieved the key terms in their context using a concordance table (a). In the second step, we merged the queried terms (“Matched”) with the terms immediately preceding (“Before”) and immediately following (“After”) to reconstruct the sequence of words as a single string (b). Finally, in cases when the queried terms occurred multiple times in the same article, we reunited the article by merging its various instances in the same string (c). The full code for retrieving concordance can be found [here](#) (Chinese) and [here](#) (English). The main challenge with this method is to define the right context size for the concordance. After a close reading of a sample of articles representing different possible cases, we opted for the lowest possible context size in each language in order to minimize the risk of overlap, setting the upper limit to 400 characters for Chinese and 100 for English. In cases when the window size is too narrow, however, some information may be lost, whereas when it is too large, overlap and repetitions may occur. Although this solution is not ideal, we still believe it is our best option at this stage. To finely adjust the context size to each possible case would require a significant amount of manual work, which is well beyond the scope of this research. This is an issue we need to address in future research. END HERMENEUTICS

```
[8]: display_png(file="./media/Fig.3.png")
```

CONCORDANCE

DocId	Date	Title	Source	Before	Matched	After
1 SPSP194704140406	19470414	扶輪社年會閉幕	shunpao	國際	扶輪社	第九十七區域年會昨日已為最後一日，上午在美國教堂舉行禮拜，下午在永安公司總經理郭禮安私邸舉行圓遊會，
2 SPSP194704140406	19470414	扶輪社年會閉幕	shunpao	一日，上午在美國教堂舉行禮拜，下午在永安公司總經理郭禮安私邸舉行圓遊會，中外社友全體參加，盛況空前。	扶輪社	此次年會為期三”，各地社友來滙參加者達百餘人，對於擴大扶輪社範圍及增進國際友誼，已有詳細討論。
3 SPSP194704140406	19470414	扶輪社年會閉幕	shunpao	舉行圓遊會，中外社友全體參加，盛況空前。扶輪社此次年會為期三”，各地社友來滙參加者達百餘人，對於擴大	扶輪社	範圍及增進國際友誼，已有詳細討論。



Context size = 100 (Chinese), 400 (English)

MERGE

Before + Matched + After



RE-UNITE

1 document = 1 row

START HERMENEUTICS Next, we prepared the text data to make it processable by topic model algorithms. Standard pre-processing steps include tokenization, stemming and lemmatization (optional), and various filtering options (stop words, punctuation, rare words). Tokenization is the operation of segmenting the text into meaningful units, which in this context, are essentially words. For the Chinese text more specifically, we used [jiebaR](#), which is one of the most popular tokenizers today. Although it was primarily designed for tokenizing contemporary Chinese texts, it gave satisfactory results on the texts published during the period under study (1919-1949). For the English text, we relied on the standard tokenizer included in the pre-processing function of the [stm R package](#) (which we describe below). In the next steps, we chose not to stem (to reduce words to their root) and not to lemmatize word (to reduce words to their common form), because at this stage we preferred to maintain all the nuances conveyed in the original articles. Additionally, we provided a customized list of stop words, including the terms used for querying the corpus (" ", "Rotary", "club") and too common terms in this context, such as " ", " ", "China", and "Chinese". We also removed the words that contained less than two characters and occurred in less than two documents in Chinese, and those containing less than four characters and occurring in less than five documents in English. Given these parameters, five documents were removed from the Chinese corpus. The resulting corpora eventually contain 2378 terms in English and 921 in Chinese, representing their respective vocabulary. **END HERMENEUTICS**

1.1.3 Step 3: Model building

Various topic modeling algorithms have been implemented so far. For this research, we opted for a Latent Dirichlet Allocation (LDA) algorithm, which is one of the most popular methods today. LDA is a probabilistic model that treats topics as mixtures of words and documents as mixtures of topics. This implies that words can belong to different topics, while topics can be represented in multiple documents with varying proportions. More specifically, we rely on structural topic modeling (STM) to incorporate metadata such as the date of publication in order to investigate the effect of time on topical prevalence. From a technical perspective, we chose to use the [stm R](#)

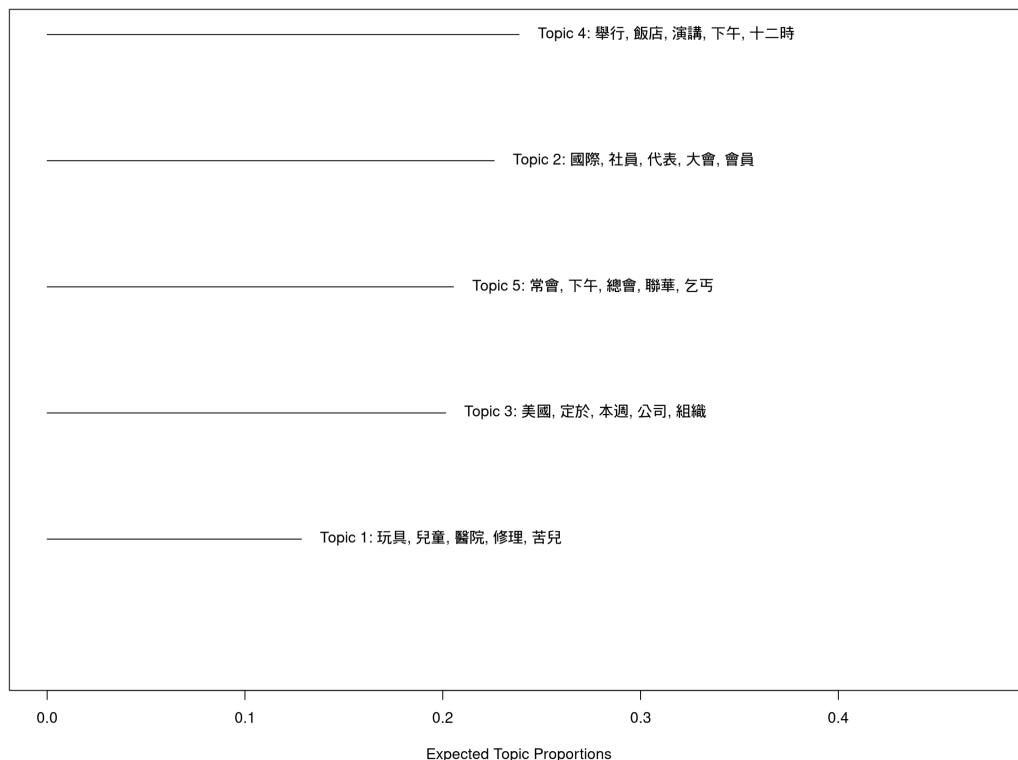
package which includes several built-in functions designed to facilitate the exploration of topics, including various visualizations and statistical outputs (*Roberts, Stewart, Tingley 2019*). The full code for building the topic models can be found [here](#) (Chinese) and [here](#) (English).

Choosing the right number of topics k remains a highly debated question. There is no definite answer. Most topic modeling tools, including the *stm* package, generally provide a set of metrics such as held-out likelihood, residual analysis, average exclusivity and semantic coherence, to help the researcher to determine the optimal number of topics for a given corpus. According to the creators of the *stm* package, the optimal number of topics for corpora that comprise between a few hundred to a few thousand documents, should range from 5 to 50 topics (*Roberts, Stewart, Tingley, Benoit 2020*). Ultimately, however, only the researcher's interpretational needs can determine the most appropriate number of topics for a given specific research. After several experiments, we decided to build three models with 5, 10, and 20 topics for each corpus. This would enable us to navigate between three different levels of granularity. Building several models with gradual numbers of topics also helps to overcome the issue of having corpora of different sizes and structures. Topics can then be aligned across models and similar topics do not necessarily need to be represented in the same models, with the exact same number of topics.

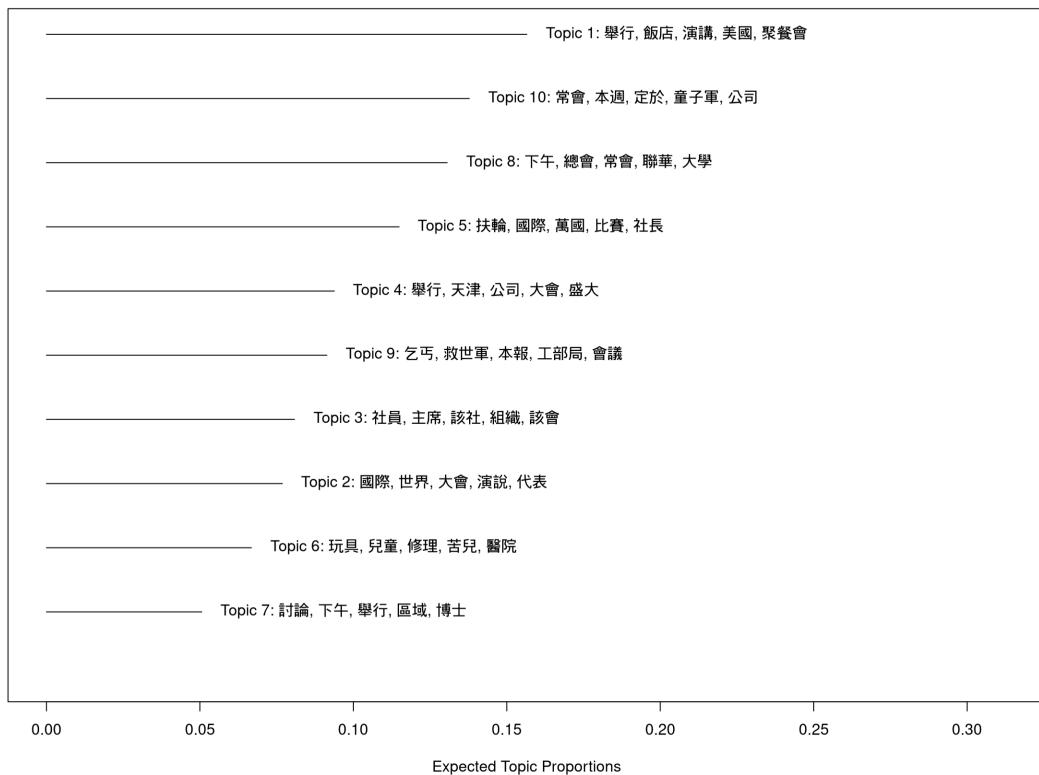
START HERMENEUTICS The results of topic modeling are usually presented in the forms of lists of words and probabilities. Each topic is defined by a series of words which most frequently co-occur together. Topics are also associated with their global proportion in the entire corpus and their relative proportion per document. The series of plots below shows the most frequent words for each topic and the relative proportion of the topics for each model (figure 4, figure 5). END HERMENEUTICS

```
[9]: display_png(file="./media/Fig.4a.png")
display_png(file="./media/Fig.4b.png")
display_png(file="./media/Fig.4c.png")
```

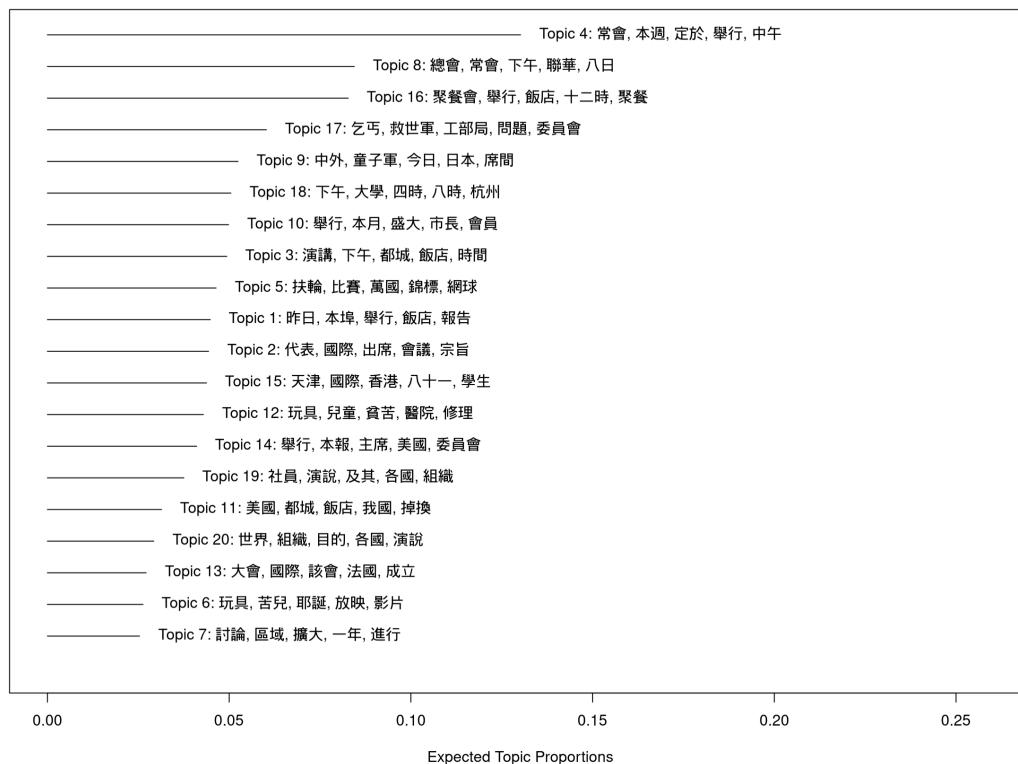
Shenbao 扶輪社 Corpus 5-topic model



Shenbao 扶輪社 Corpus 10-topic model

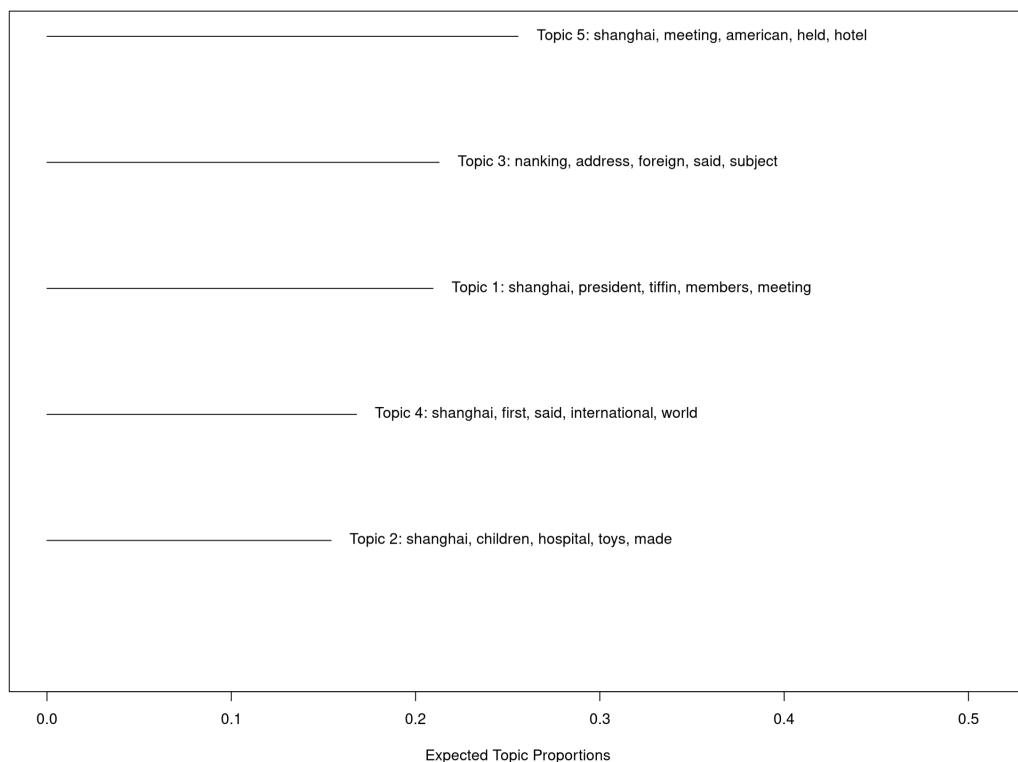


Shenbao 扶輪社 Corpus 20-topic model

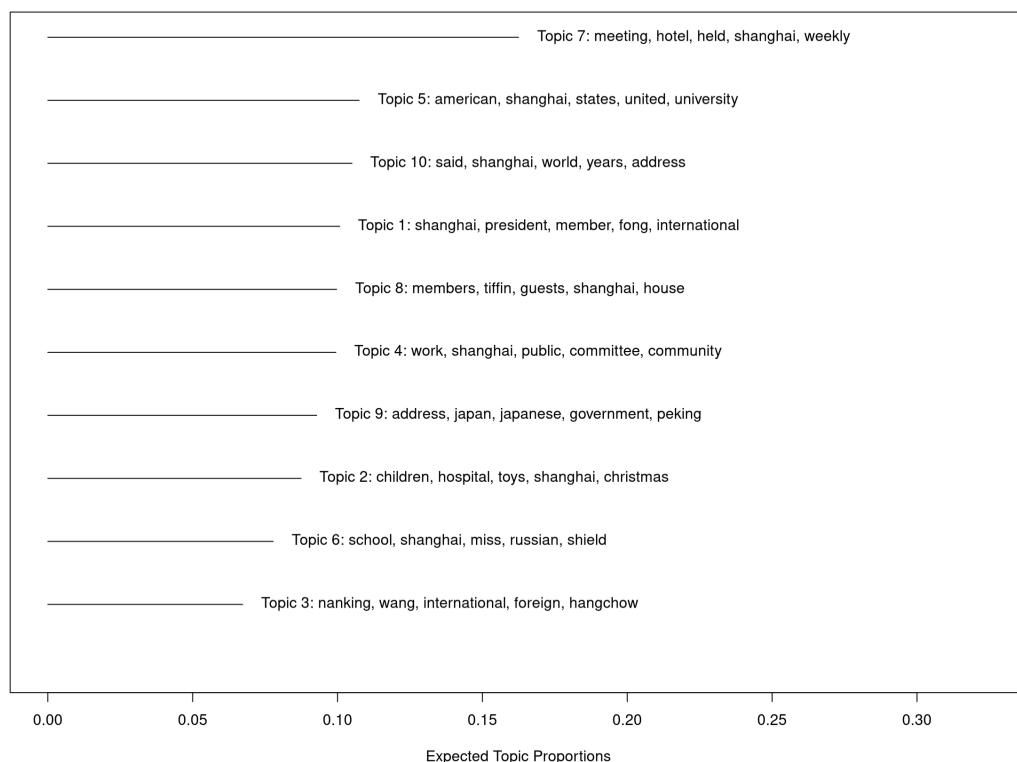


```
[10]: display_png(file="./media/Fig.5a.png")
display_png(file="./media/Fig.5b.png")
display_png(file="./media/Fig.5c.png")
```

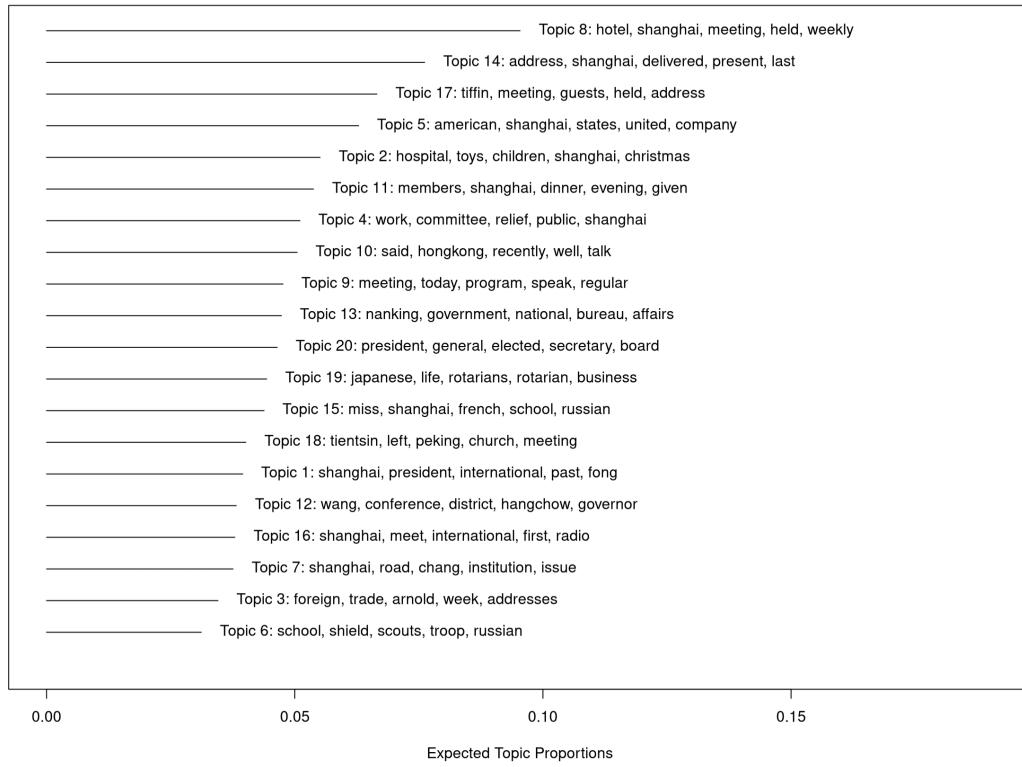
ProQuest Rotary Corpus 5-topic model



ProQuest Rotary Corpus 10-topic model



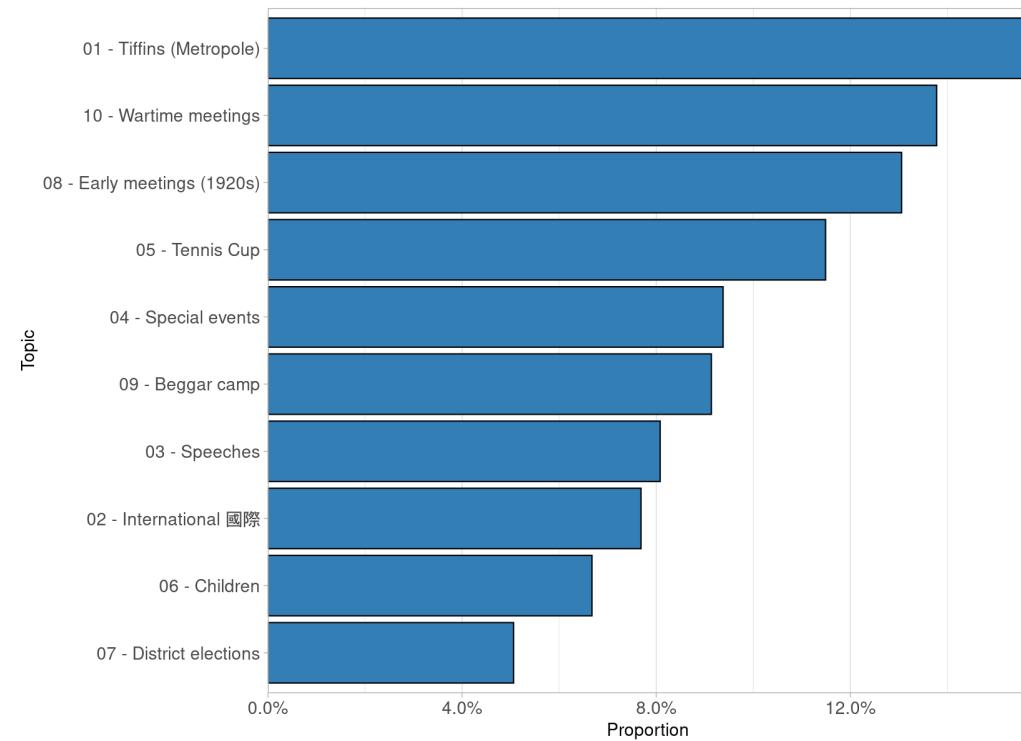
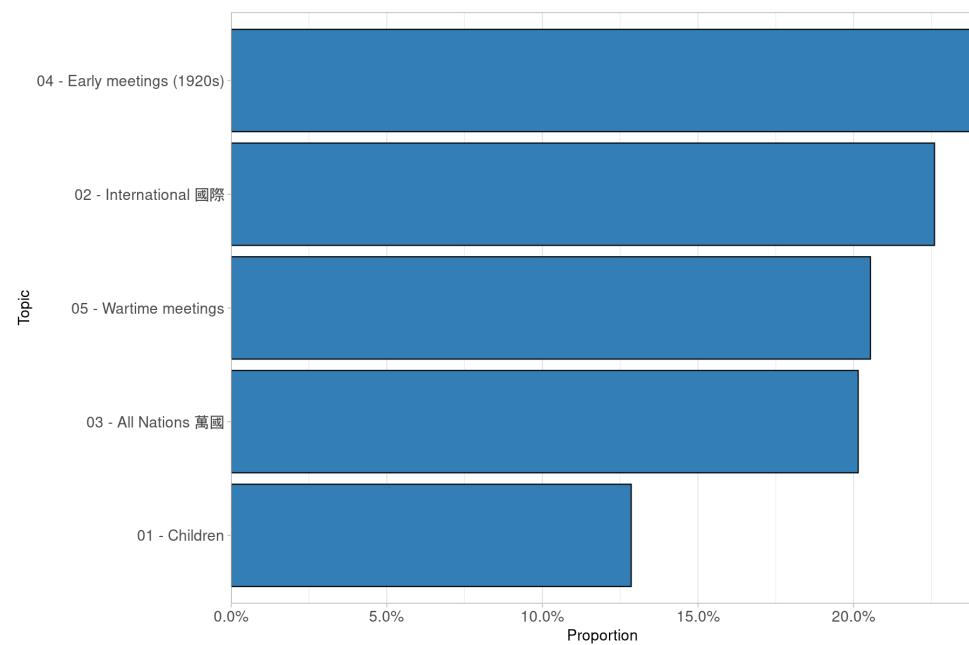
ProQuest Rotary Corpus 20-topic model

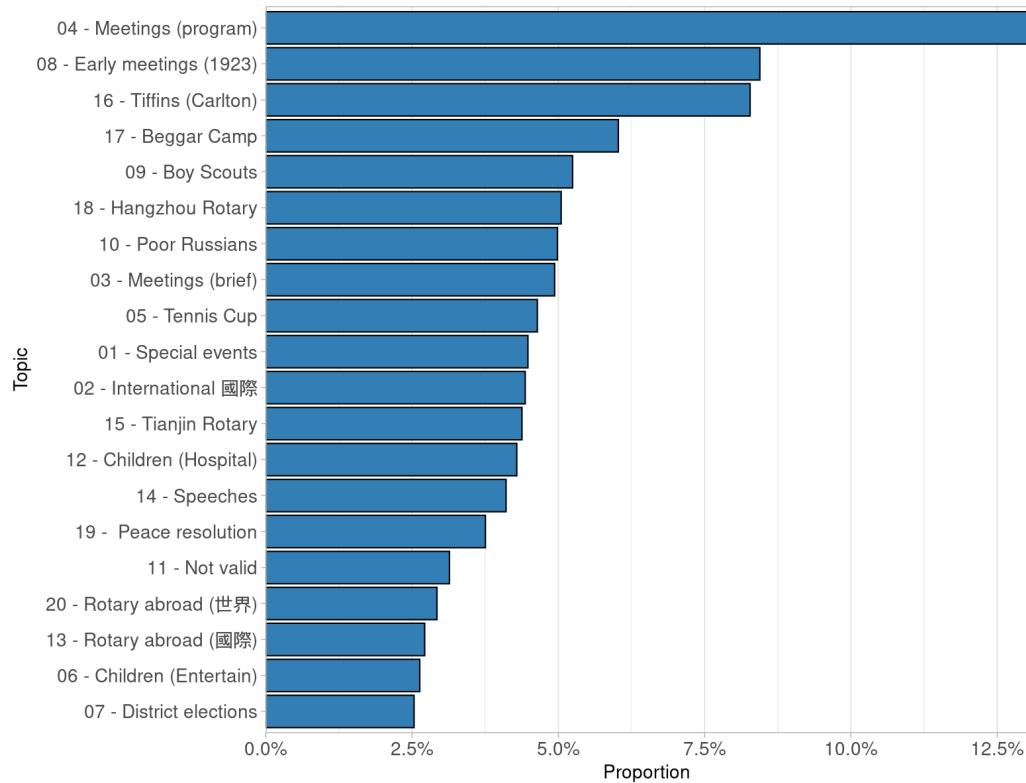


START HERMENEUTICS Based on the list of words, we labeled the topics in the most meaningful way as possible (figure 6, figure 7). More than often, especially in cases when the topics seemed unclear, we do not just rely on the list of words. We also closely examined a sample of highly representative texts for each topic to better understand how the topics translated into concrete words and sentences. Additionally, we relied on our previous knowledge of the club and of the broader historical context to enhance our final interpretation of the topics. END HERMENEUTICS

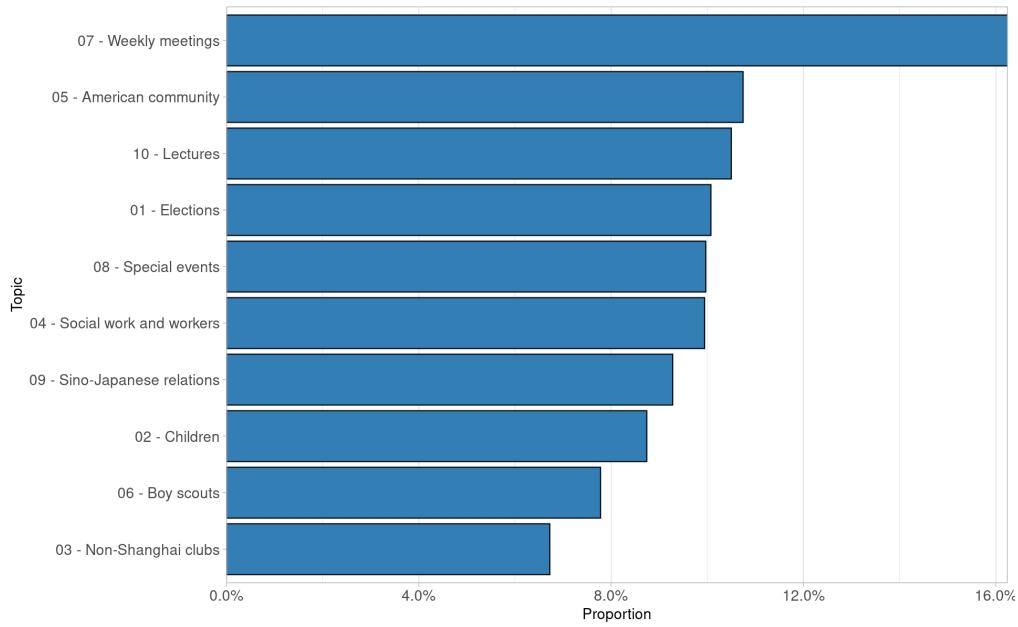
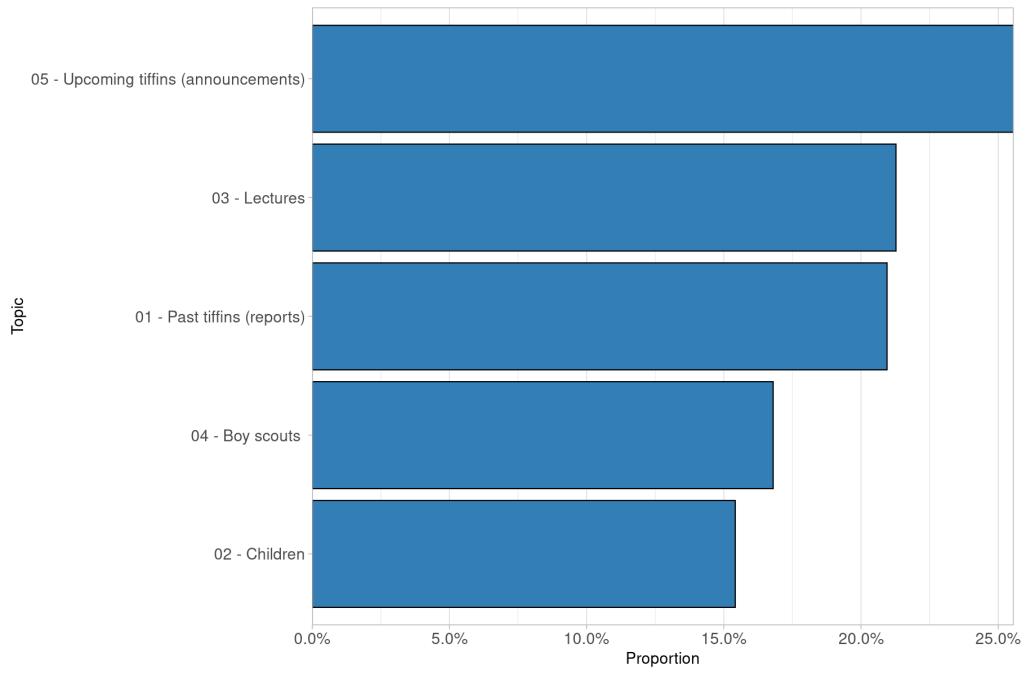
```
[11]: display_png(file="./media/Fig.6a.png")
display_png(file="./media/Fig.6b.png")
display_png(file="./media/Fig.6c.png")
```

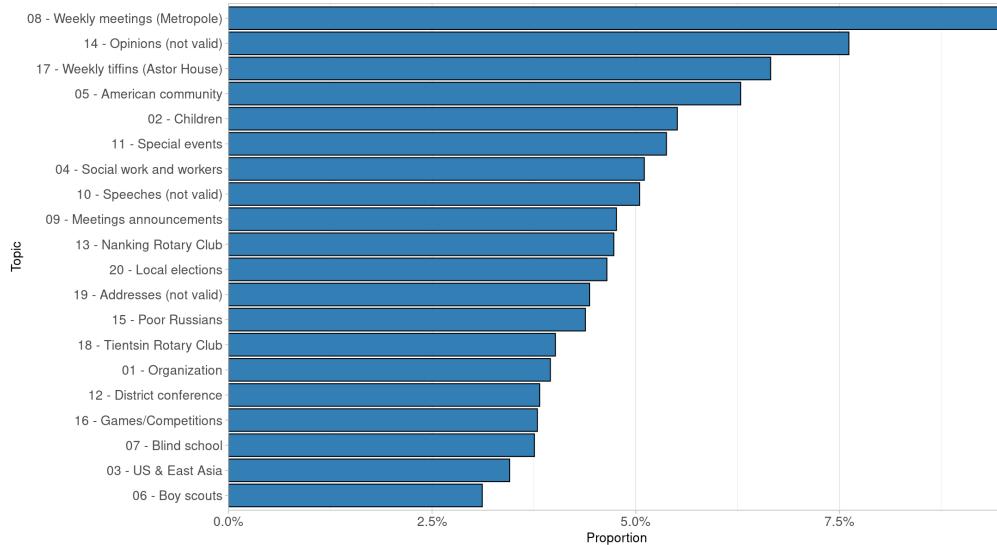
This plot shows the topic proportions over all documents.





```
[12]: display_png(file='./media/Fig.7a.png')
display_png(file='./media/Fig.7b.png')
display_png(file='./media/Fig.7c.png')
```



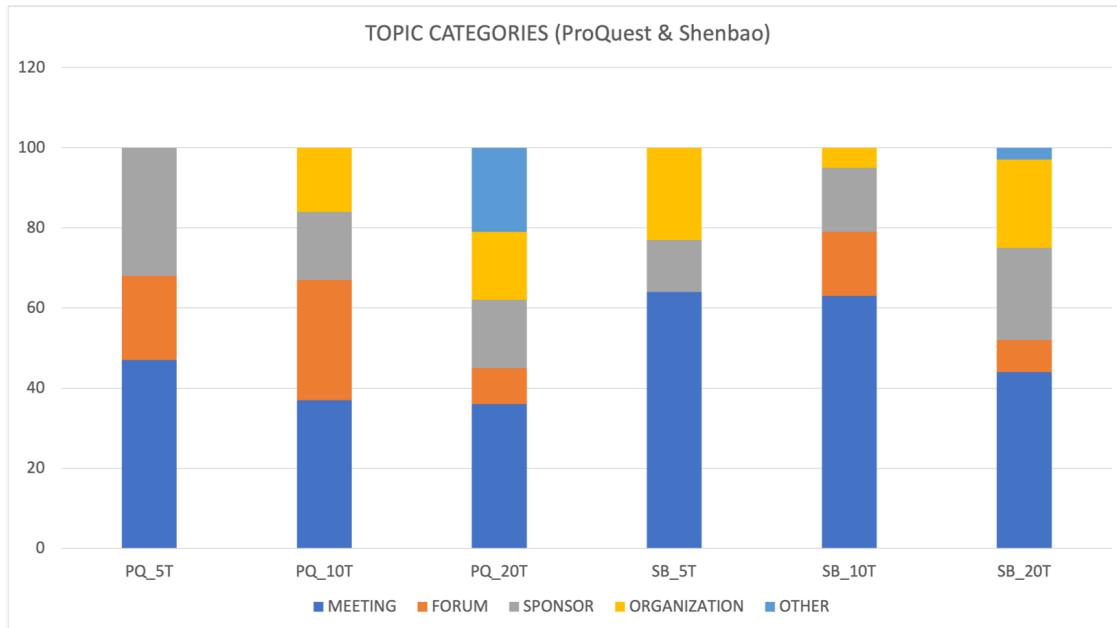


START HERMENEUTICS Tables 1 to 6 in the (Appendix) provide a summary of the topics for each model, including their label, the 10 most frequent words defining each topic, and their various attributes (topical group, dimension, proportion, and trend over time). Tables 7 and 8 show the topics aligned across models. Tables 9, 10, and 11 present the topics aligned across languages.
END HERMENEUTICS

1.2 The *modus operandi* of the public sphere

START HERMENEUTICS In the next step, we grouped the labeled topics into four broad categories (figure 8). These four meta-topics reflect the four main domains of activity of the Rotary Club, and by extension, to the four *modus operandi* of the Shanghai public sphere. END HERMENEUTICS

```
[13] : display_png(file=".~/media/Fig.8.png")
```



START HERMENEUTICS Meeting, socializing, entertaining (blue). This category refers to the weekly meetings (tiffins) held every Thursday by the Shanghai Rotary Club (lunch, tiffin), as well as special events (dinner dance, concert, garden party) and sports games, such as the Rotary Tennis Cup organized every year in the early 1930s. In the *Shenbao* corpus, this group includes topics 4 and 5 in the 5-topic model, topics 1, 10, 8, 5, and 4 in the 10-topic model, and topics 4, 8, 16, 3, 5, and 1 in the 20-topic model. In the ProQuest corpus, this group includes topics 5 and 1 in the 5-topic model, topics 7, 5, and 8 in the 10-topic model, and topics 8, 17, 5, 11, and 9 in the 20-topic model.

END HERMENEUTICS

START HERMENEUTICS Sponsorship and community service (grey) refers to various philanthropic projects aimed at both elite (students, boy scouts) and non-elite populations (poor Russian emigres, poor children, the blind, war refugees). In the *Shenbao*, this group includes topic 1 in the 5-topic model, topics 9 and 6 in the 10-topic model, topics 17, 9, 10, 12, and 6 in the 20-topic model. In the English-language press, this group includes topics 4 and 2 in the 5-topic model, topics 2 and 6 in the 10-topic model, and topics 10, 15, 7, and 6 in the 20-topic model.

END HERMENEUTICS

START HERMENEUTICS Organization (yellow) includes topics dealing with elections of local and district officers, Rotary International delegates, national and international conferences, relations with Rotary International, exchanges between Rotary clubs in China and other countries. In the *Shenbao*, this group is specifically represented by topic 2 (Rotary International) in the 5-topic model, by topic 7 (district elections) in the 10-topic model and becomes more prominent in the 20-topic model (topics 18, 2, 15, 20, 13, and 7). In ProQuest, organizational issues emerge as a topic in the 10-topic model (topic 10 on non-Shanghai clubs) and gain importance in the 20-topic model (topics 1, 12, 13, and 20). END HERMENEUTICS

START HERMENEUTICS Forum (orange) refers to lectures given during tiffin-meetings and

speeches delivered by official guests during special events. The discussions covered a variety of topics, including economic development, social welfare, health, technology, Chinese arts and cultural exchanges. In the *Shenbao*, this topical group emerges only in the 10-topic model (topics 2 and 3) and remains under-represented in the 20-topic model (topics 14 and 19). It is more prevalent in the English-language press, represented by topic 3 in the 5-topic model, topics 10, 4 (social work), and 9 (Sino-Japanese relations) in the 10-topic model, and by topics 4 (social work), 3 (American and East Asia), and several opinion articles (topics 14, 10, 19) in the 20-topic model.

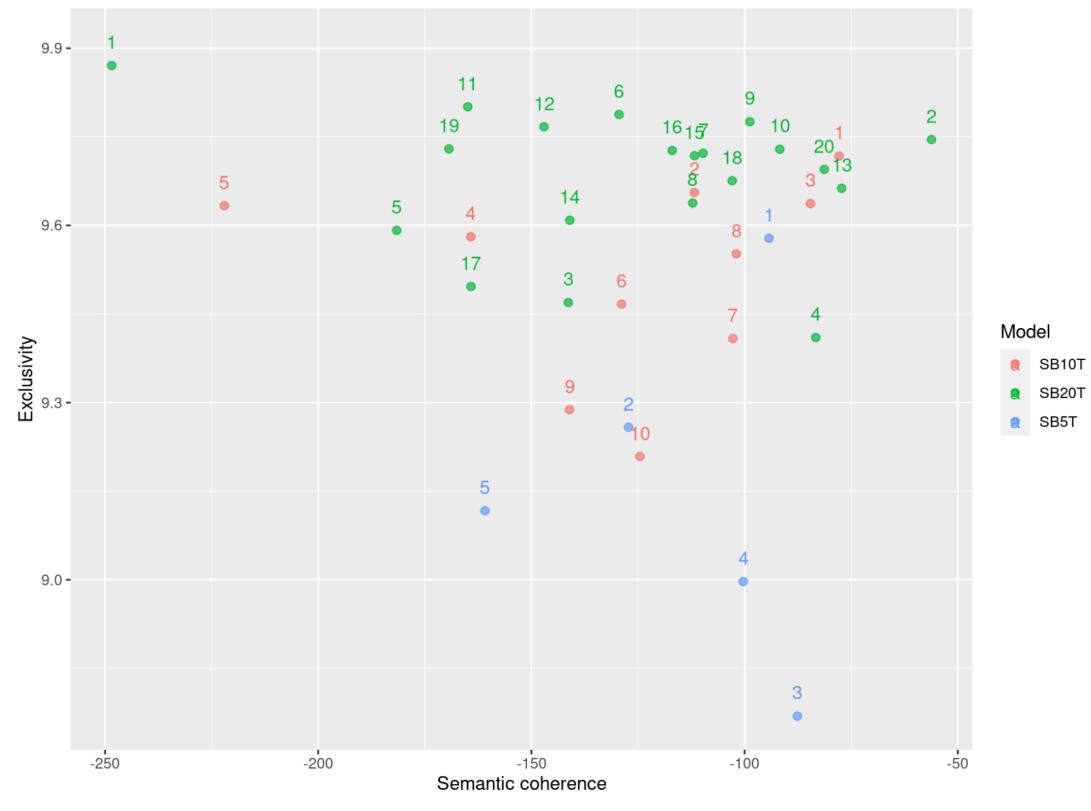
END HERMENEUTICS

START HERMENEUTICS **Other (pale blue)**. The remaining topics are classified as ambiguous or irrelevant. They often present a low semantic coherence, which reflects the polysemous nature of the words they include. Topic 11 in the 20-topic model of the *Shenbao* provides an example of a non-topic made up of three distinct issues: (1) Sino-American relations (speech of the Chinese general consul before the New York Rotary Club regarding Sino-American friendship, Rotary International scholarship for students in the United States), (2) a series of advertisements for a toy collection campaign, and (3) various meetings. The connection between (2) and (3) is made through the “Metropole Hotel” () which is used both as a place of meeting and a place for collecting toys, whereas the connection between (1) and (3) is made through the word “speech” (). In the English corpus, topics 14, 10, and 19 in the 20-topic model contain unrelated articles that share very general words, such as “life”, “business”, “said”, and “address”. In an improved version of the model, these generic words should be removed to increase the relevance of topics. As these examples suggest, invalid topics tend to become more frequent as we increase the number of topics in the model. This can be measured by plotting topic coherence against exclusivity, as shown on the plot (figure 9). The higher the number of topics, the lower the coherence, and the higher the exclusivity.

END HERMENEUTICS

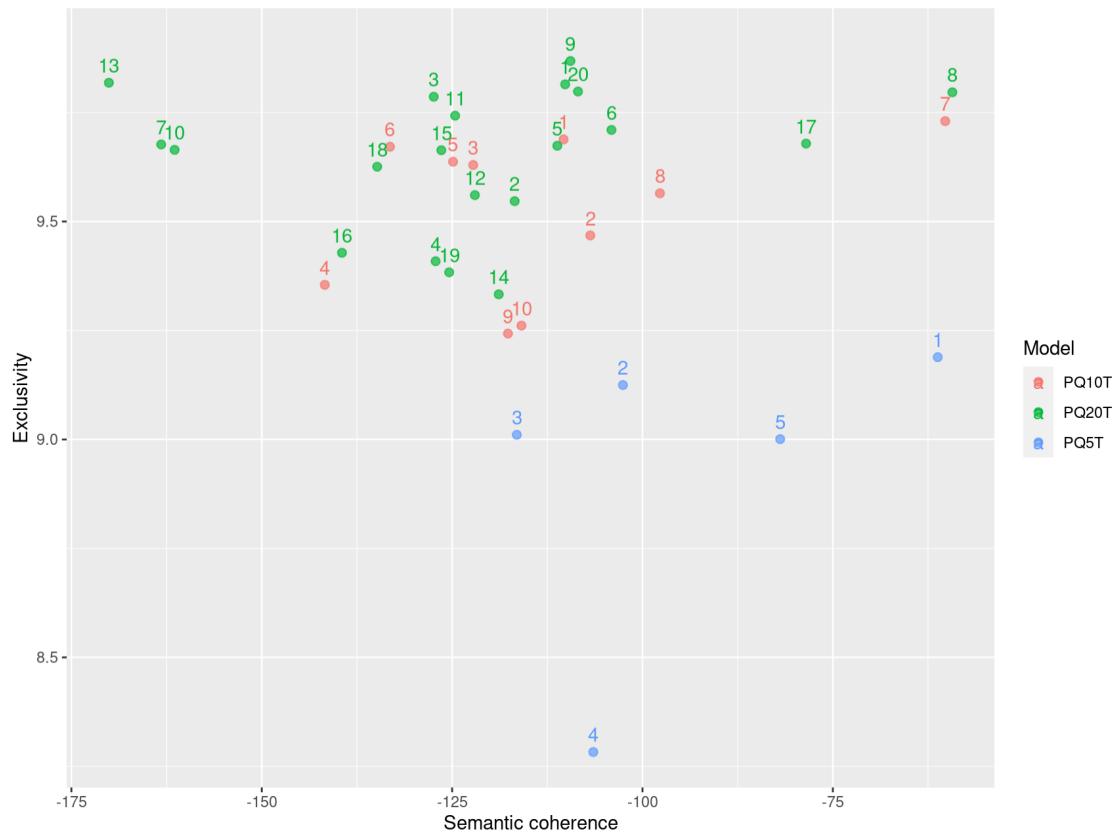
```
[14]: display_png(file="./media/Fig.9a.png")
display_png(file="./media/Fig.9b.png")
```

Comparing exclusivity and semantic coherence
Chinese-language corpus (Shenbao)



Comparing exclusivity and semantic coherence

English-language corpus (ProQuest)



START HERMENEUTICS Meeting-related topics prevail in both corpora (40 to 60%), meaning that socializing and entertaining were the primary functions of the club. Local meetings were particularly prominent in the *Shenbao*, which essentially reflects the daily periodicity of the newspaper. The *Shenbao* contains a fairly large number of repeated announcements of upcoming meetings, whereas the English-language weekly press featured less frequent but more substantial reports of past meetings. END HERMENEUTICS

START HERMENEUTICS Community service was the second most important activity of the club (10 to 30%). The two corpora shared a similar concern for children, reflected in the toy collection campaigns that the Rotary organized every year for Christmas (SB20T06), and the fundraising campaigns aimed at building a children hospital in Zhabei (north of the International Settlement) (SB20T12). In fact, children-oriented philanthropy represented the most prevalent and resilient topic in this category. The 20-topic models also shared a growing concern for helping poor Russian emigres in Shanghai (SB20T10, PQ20T15). Boy scouts featured more prominently in the English-language press (PQ5T04, PQ10T06, PQ20T06), whereas the *Shenbao* was more concerned with the construction of a “beggar camp” for refugees during the Sino-Japanese war (SB10T09, SB20T17). These divergences reflect both the national profiles of the individuals who were involved in the various charity projects, as well as the uneven impact of the Sino-Japanese war (1937-1945) on the Chinese and the English press. On the one hand, the boy scout movement in Shanghai was essentially supported by the British Rotarian Francis Charles Millington. On the

other hand, the English-language press was more seriously affected by the war than the *Shenbao*.
END HERMENEUTICS

The greater prevalence of organizational topics in the *Shenbao* (25% compared to 5% in the Pro-Quest corpus) reflects the general growth of Chinese members and the increasing autonomy of Chinese clubs in the global organization (*Armand 2022*). While the English press focused on local elections prior to the Sino-Japanese war (PQ10T01, PQ20T20), the Chinese press seemed more concerned with district elections and postwar reorganization (SB10T07, SB20T07). Both corpora reported on other Rotary clubs in China (PQ10T03), including Tianjin (PQ20T18, SB20T15), Nanjing (PQ20T13), after the Nationalists had established their capital in the city in 1927, and Hangzhou, where the first Chinese-speaking club was established in 1931 (SB20T18) (*Armand 2022*, p.254). While international conventions were systematically documented in the two corpora (SB5T02 , SB10T02, SB20T02, PQ20T01), the *Shenbao* was more likely to report on other Rotary clubs outside China (SB20T13, SB20T20).

Finally, the presence of forum-related topics (5 to 20%) in the two corpora illustrates a core characteristic of the public sphere as an alternative to armed conflicts. It is remarkable that in both languages, every topic is deprived of any words denoting conflict or violence. Instead, the words that dominate the six topic models all emphasize dialogue and discussion as the primary means of communication among Rotarians. These words also express the self-image that the Rotary aimed to present to the world. For instance, if we focus on the 10-topic model, topic 2 is defined by the words *yanshuo* (speech), *daibiao* (delegate or representative), *qinshan* (goodwill), and *jingshen* (spirit) in Chinese. These words serve to define other topics as well, and in fact they occur frequently across the entire corpus. In the English-language press, the words “said”, “address”, “speech”, and “talk” are highly frequent in topic 10 as well as in other topics. These topics epitomize the self-claimed apolitical nature of the Rotary club and its commitment to international peace and goodwill. Potentially conflictual issues were always discussed on a rational basis during meetings and special events, which offered a platform for guest speakers, including government officials, to defend their opinions and political programs in their keynote speeches. The Rotarians’ mode of communication and ways of managing conflicts contrasted sharply with the war of resistance advocated by other Chinese organizations during the same period. This emphasis on dialogue and non-violence applied not just to the Rotary Club but also to the press as an informational medium at large. As (*Wagner 2007*) stated: “The very institution of the newspaper came in this package of successful modernization as a core ingredient promising information flow in society as well as rational and effective handling of conflict.” (*Wagner 2007*, p.5). We would certainly obtain very different results if we applied topic modeling to more “political” publications and perhaps to other sections in the same periodicals, especially the sections devoted to national and international news during these troubled times.

START HERMENEUTICS The “forum” category nonetheless presents significant differences between the two corpora in terms of topic prevalence and topical contents. Forum-related topics were more prevalent in English periodicals than they were in the *Shenbao*. This again essentially reflects the weekly periodicity of English-language press, which allowed for lengthy and more detailed accounts of lectures and discussions, whereas the Chinese daily was more likely to publish brief accounts and announcements of lectures. The English press tended to focus on social/economic issues (PQ10T04 and PQ20T04) and international relations, especially the American presence in East Asia (PQ20T03). The two corpora shared a growing concern for Sino-Japanese relations after the Japanese invasion of Manchuria in September 1931. This critical topic, however, was not as prominent as expected and made only brief appearances in both corpora (PQ10T09, SB20T19).

END HERMENEUTICS

1.3 Negotiating between the local and the global

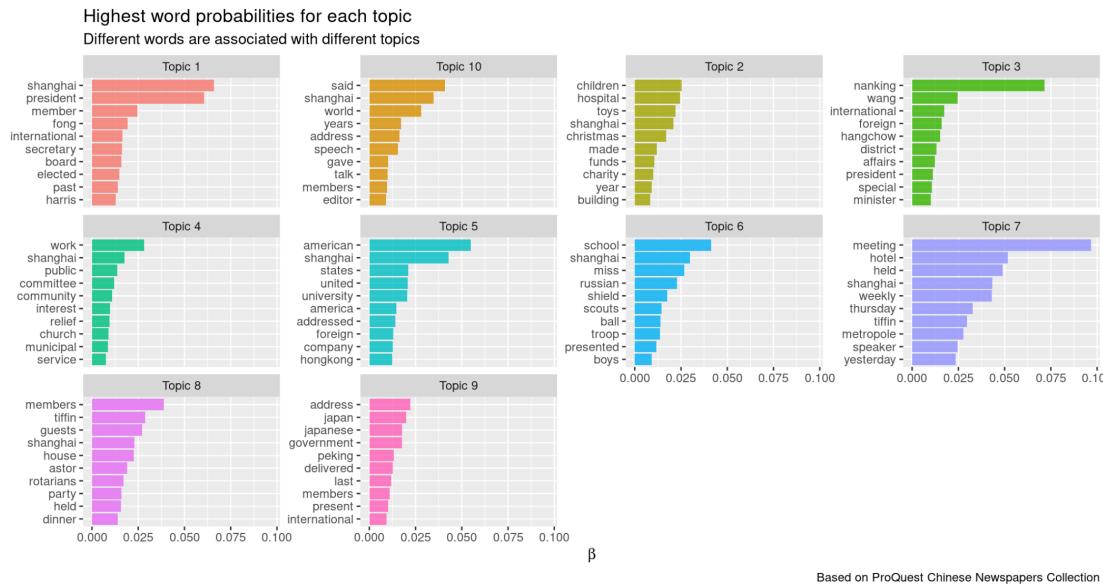
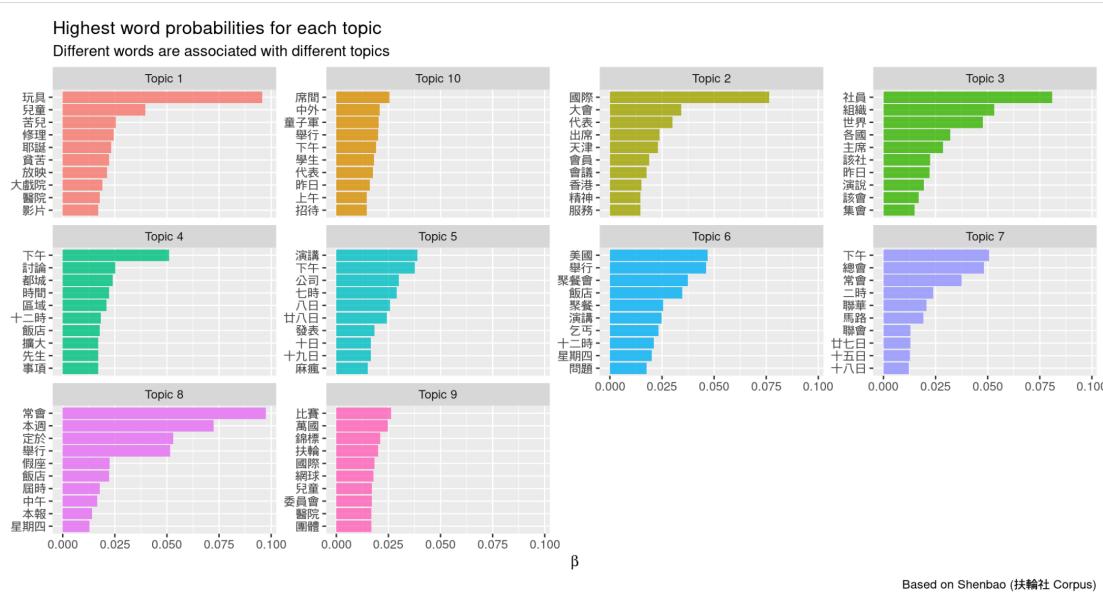
START HERMENEUTICS To investigate the transnational nature of the public sphere, we further classified the topics into local/non-local categories. In this research, the “local” reflects the Shanghai point of view. Since we are relying on periodicals published in Shanghai and catering primarily to a Shanghai-based audience, the “local” refers to the point of view of Shanghai newspaper readers. In the two languages, the local can be grasped empirically through terms referring to local places (street name, building or hotel in which the Rotary Club held its meetings), local time (day, date, exact time of meeting), local institutions (local companies or local branch of outside companies, municipal administration, civic associations, or local business organizations), and local positions (president, manager, director, board member). The local, however, is not defined by people (Rotary members or guests). We consider that human actors were intrinsically mobile, both geographically and professionally, and their mindset and influence extended far beyond their daily routine and local sphere of action.

END HERMENEUTICS

How did the transnational dimension of the Rotary Club perspire through its four operating spheres? First, as a transnational organization, the Rotary Club had local branches with a large degree of autonomy from its center in Chicago, which is reflected in several topics dealing with local elections, district elections and national conventions (SB10T07, PQ10T01, SB20T07, PQ20T01, PQ20T12, PQ20T20). Second, topics related to community service reveal the implications of global conflicts on local philanthropy, as in the case of war refugees (SB10T09, SB20T17). Topics related to lectures and speeches (forum) often dealt with global issues, such as international relations, political economy, and cultural exchanges across national boundaries (SB10T02, PQ10T09, PQ10T10, SB20T19, PQ20T03). Finally, topics related to meetings and socializing arenas reveal that Chinese Rotary clubs received many visitors from abroad, including Rotary International delegates and representatives of other clubs in Asia and Western countries (SB20T01, PQ20T05). From the perspective of the local press, Shanghai emerged as the new regional and global center in East Asia. Circulations went both ways. Not only did foreign visitors come to China, but Chinese Rotarians as well frequently visited other clubs in foreign countries.

START HERMENEUTICS The distinction between local and non-local topics, however, is not always straightforward. In LDA-based models, topics are mixtures of words which can be both localized and globalized. This can be best illustrated by focusing on the words that define each topic. The following examples are taken from the 10-topic model, which offers the best compromise. The 5-topic models contain topics that are too broad and difficult to interpret, whereas the 20-topic models include overlapping topics, non-topics, and topics that are exclusive to single documents. Eventually, the 10-topic model is more balanced and lends itself to more meaningful interpretation (figure 10). END HERMENEUTICS

```
[15] : display_png(file="./media/Fig.10a.png")
display_png(file="./media/Fig.10b.png")
```



START HERMENEUTICS Topics 1 and 3 in the Chinese corpus provide examples of mixed topics including both local and non-local terms (figure 11). Topic 1 includes words that refer to the United States () (American guests) as well as tiffin meetings () in local hotels (Metropole Hotel) at local time (yesterday, Thursday, noon). Topic 3 is defined by terms that refer to other countries () as well as local time (yesterday), local place (this port), local institutions (this club) and local positions (club member, chairman). Topics 6, 8, 9, and 10 are more clearly local (, , : local children hospital, : local movie projection, : this afternoon, : two o'clock, , : street names), whereas topics 2, 4, 5, 7 refer to more distant settings and events (international,

universal, world, delegate, goodwill, Japan, Tianjin, this month, district). END
HERMENEUTICS

[16]: `display_png(file="./media/Fig.11.png")`

Topic 1	Topic 3
<p>生是緊張的，時代也是偉大的，兩者都不容我不流，也不許我靜流。所以，我是一條激流，跳躍地。游藝界扶輪社主辦的徵集苦兒玩具電影新影——耶誕前輪流在各大影院公映——耶誕之前，上海照例有各團體發後，不知有著多少苦兒，他們終年是沒有享得娛樂的。因此為使這些苦兒在耶誕時節，也可娛樂起見，像青年會，扶輪社等各機關，照例在耶誕之前，要發起放映徵募苦兒玩具的電影。而目前，離耶誕新年為日無多，扶輪社主辦的徵集</p>	<p>未遷入、嗣經納稅華人會及華董杜君出任調停、雙方均圓解決。公董局已於昨日通知各攤基菜飯、定於明日日本埠扶輪社、昨日舉行盛大熱烈之聚餐會、席間由該會會長介紹萬國扶輪社特派員大衛生氏演說、氏首致謝忱、略謂、得在中、公董局已於昨日通知各攤基菜飯、定於明日日本埠扶輪社、昨日舉行盛大熱烈之聚餐會、席間由該會會長介紹萬國扶輪社特派員大衛生氏演說、氏首致謝忱、略謂、得在中國首創且會員國籍最多之扶輪社中演說、殊覺榮幸、扶輪社之組</p>
<p>扶輪社主辦的徵集苦兒玩具電影 新影 ——耶誕前輪流在各大影院公映—— 耶誕之前，上海照例有各團體發起後，不知有著多少苦兒，他們終年是沒有享得娛樂的。因此為使這些苦兒在耶誕時節，也可娛樂起見，像青年會，扶輪社等各機關，照例在耶誕之前，要發起放映徵募苦兒玩具的電影。而目前，離耶誕新年為日無多，扶輪社主辦的徵集</p>	<p>上海中西商所組織之上海扶輪社、於昨日中午、在利查飯店慶祝該會世界扶輪社二十三週年紀念、主席為孔雀影片公司經理朱錫君、報告該會組織 上海中西商所組織之上海扶輪社、於昨日中午、在利查飯店慶祝該會世界扶輪社二十三週年紀念、主席為孔雀影片公司經理朱錫君、報告該會組織至今已有二十三年、迄今世界各國、有扶輪社之</p>
<p>車六進一（勝） 一件玩具換一票 整個世界雖充滿了恐怖，時間却過得那麼快，耶誕又離得不遠了。上海扶輪社，照例在耶誕之前，要公映幾次徵集玩具影片。把徵集到的玩具，於耶誕時節，分發給終年沒有看見玩具的苦兒。善，倒是不拘的，倘若沒有玩具，那當天在劇院門口，也有玩具出售的。所有舊的玩具，集到後，立即送到扶輪社的玩具醫院去修理，務使耶誕時節，能夠發給全市慈幼機關中的苦兒，當做一份禮物。扶輪社預定的目標，可徵</p>	<p>濟南扶輪社二十一晚開正式註冊紀念會、到中外男女來賓百餘人、韓致祝詞、盼世界十七萬萬人盡成扶輪社員、則世界可無戰 濟南扶輪社二十一晚開正式註冊紀念會、到中外男女來賓百餘人、韓致祝詞、盼世界十七萬萬人盡成扶輪社員、則世界可無戰爭。（二十二日專電）濟扶輪社註冊紀念</p>

START HERMENEUTICS In the English corpus, mixed topics include topics 1, 5, and 10 (figure 12). Topics 1 deals with the elections of local club officers (board, elected, president, secretary,

Shanghai) with an international stature, such as Kuang Fuzhuo, an American-educated Chinese who served as president of the Shanghai Club as well as delegate of Rotary International, district governor and leader of the Young Men's Christian Association (Y.M.C.A.). Topic 5 deals with American residents and visitors in Shanghai, exemplifying the transnational flows of people coming and going every day through the treaty port. Topic 10 refers to speeches that addressed global issues during a local meeting in Shanghai. More clearly global topics includes topics 3 & 9 (about Japan), whereas topics 2, 4, 6, 7, and 8 are more clearly locally grounded. END HERMENEUTICS

```
[17]: display_png(file=".\\media\\Fig.12.png")
```

Topic 1	Topic 5	Topic 10
<p>n.-S Sanzelli LEAVING FOR EUROPE SOON - Mr Chairman of the Shanghai Rotary Club Dr Julian Arnold will leave Shanghai soon for Europe - Sanzelli said Dr Herbert Chay head of the Whangpoo Cemetery General Manager of the Chinese Department of the Rotary Club Mr Bertrand has been Dr Wiedens in China for many years Photo by Ah Fung we Ajitao Mr and Mrs Bertrand who have left here recently via Siberia to</p>	<p>Julian Arnold Addresses Rotary Club on China's Foreign Trade - Dr Julian Arnold, Vice Chairman, an address before the Shanghai Rotary Club by Dr Julian Arnold, Vice Chairman, was made available on the subject of How Can China's Julian Arnold Addresses Rotary Club on China's Foreign Trade The topic was chosen by Dr Julian Arnold, Vice Chairman, of the Shanghai Rotary Club by Julian Arnold American Commercial</p>	<p>JEWISH FLAG ANOTHER VIEW To the Editor of the 111 North-China Daily News Sir - Mr Bloch in addressing the meeting of the Shanghai Jewish Palestine said "The first flag in the world being flown by king Solomon and now recognized by the whole world and the world's peace" World Jewry will undoubtedly</p>
<p>Mr Sterling Fassenden Chairman of the Shanghai Municipal Council Mr Harper Commissioner of Public Works Dr Herbert Chay head of the Whangpo Cemetery General Manager of the Chinese Department of the Rotary Club Mr Bertrand has been Dr Wiedens in China for many years Photo by Ah Fung we Ajitao Mr and Mrs Bertrand who have left here recently via Siberia to</p>	<p>October 9 Sergeant Thomas Wilson of the United States Marines Legation Guard commanded aside at Peiping Dr Herbert Chay head of the Whangpoo retired Chinese woman physician addressed the Peiping Rotary Club Seat 21 on Chinese Women Old and New Photo by Ah Fung we Ajitao Mr and Mrs Express Company in Shanghai and Miss Helen Mary Haven were married at St Joseph's Church on September 26 Bamford</p>	<p>To bring peaceful solution through diplomatic means The other Chinese to sneak out was Gen Wu Te Lin of the Kunming General Hospital Dr Wang Shanglin who is the old time member of the Shanghai Rotary Club coming up presented the seal of the Chinese Red Cross Society and China and World Peace His address was particularly timely and was accorded an interested reception because of his</p>
<p>Dankworth Elected Manila Rotary Club Head According to information from the Rotary Club Mr H. C. Dankworth of Manila has been elected as the new club president Mr Dankworth Dankworth Elected Manila President of the Rotary Club after election from the Rotary Club of Manila Mr George Dankworth President of the Philippine Trust Company has been</p>	<p>nd was active in many Shanghai organizations He was a member of the American Chamber of Commerce Club the Cercle Sportif Francias and was past pres. of the American Chamber of Commerce and of the American Chamber of Commerce of Manchuria Club few months preceding his death China sends more students to American universities than does any other foreign country and it is consumed in the world</p>	<p>SPEECH IN BASIC ENGLISH Mr. Woodhead Talks Before Shanghai Rotary Club have the feeling that number of those who have been hearing my talk are quite numerous and I am sure that the English only said Mr. W. Woodhead in the course of his talk in Basic Eng. the number of those who have been hearing my talk in English only said Mr. W. Woodhead in the</p>
<p>Wai Goo Lee, The Director of Shanghai Bakery Association captured the may of sudden and mysterious on the board of directors of the Shanghai Rotary Club who were elected yesterday They are front row left to right Dr Macdonald Mr. J. C. L. Bertrand President Harry Price and Yuktadee Bachet back row left to right Percy Kwok Hasenfeld Gange Price and Shaeter</p>	<p>to American Consulate Chinese Bar of the Commercial Office of the American Consulate General The Star of the United States Court for China and the Far East and the Shanghai The Simplified Rotary Club MB 011 Dwight Purley Kuranishi Mr. and Mrs. John C. L. Bertrand Prices Ruth and Uriel Clarke 1,001 P. V. 11th Mr. and Mary Hasenfeld Mr. and Mrs. Hopkins J. Irvine George How Mr and Mrs.</p>	<p>sped colour even Gassanis could conceive This is not lie to write in caption can be taken by the press men against one presman The other day during the meeting of the annual gathering of the Rotary Club which we know confounds at the best of the in ample Emporis builders protested against the steady in induction of guitar players metes into Australia and am doing however the</p>
<p>who captured it in the Keeying mountains Information of the existence of the animal was reported to have been spread by Mr Chang's newspaper and the Chinese government took steps to stop SHANGHAI ROTARY CLUB Mr Hans Berentsen Dr. Herbert Chay and Dr. Julian Arnold were elected President of the Shanghai Rotary Club at Thursday business meeting Mr Tan was elected Vice President These together</p>	<p>The American Club George F. American University Club Capt. Shapley Navy Everett Postal Agent Sterling Fassenden Naval Bar Association Foreign Correspondent of the American Consulate in Peiping Club Dr. Ramsden American Company and Health Officer Nichols American Red Cross Webb and Davis The China Press, Haier Advertising Club in China Barnes Moore Advisor to the Com missioner of</p>	<p>To the Editor of the North-China Daily News Sir - Mr. Arnold editor of the Russian daily paper Shanghai Zaria delivered speech on Russia in Asia at the meeting of the Shanghai Rotary Club his speech Mr. Arnold made rightful statement that most of the countries truly still believe the products and services of the Soviet Union greatest achievements of Imperial Russia</p>

The fact that topics are often mixtures of local/non-local words complicates the categorizations designed in the beginning of this section. Furthermore, when taken out of their context, some words appear ambiguous. Deciding on their local/non-local dimensions usually requires additional, contextual knowledge. For example, the words “ ” (*ducheng*) in topic 1 and “Metropole” in topic 7 did not refer to some global metropolis but to the name of a local hotel in Shanghai (*ducheng fandian*), which the Rotary Club used for its weekly meetings in the 1930s. Alternative text analysis tools such as bigrams and collocations can be used in combination with topic modeling to help disambiguate such words. Ultimately, disambiguation requires a close reading of sample texts, usually the top first documents with the highest topic prevalence, to better understand how topics translate into concrete words and sentences, and to eventually validate the actual content of topics

(Nelson 2020).

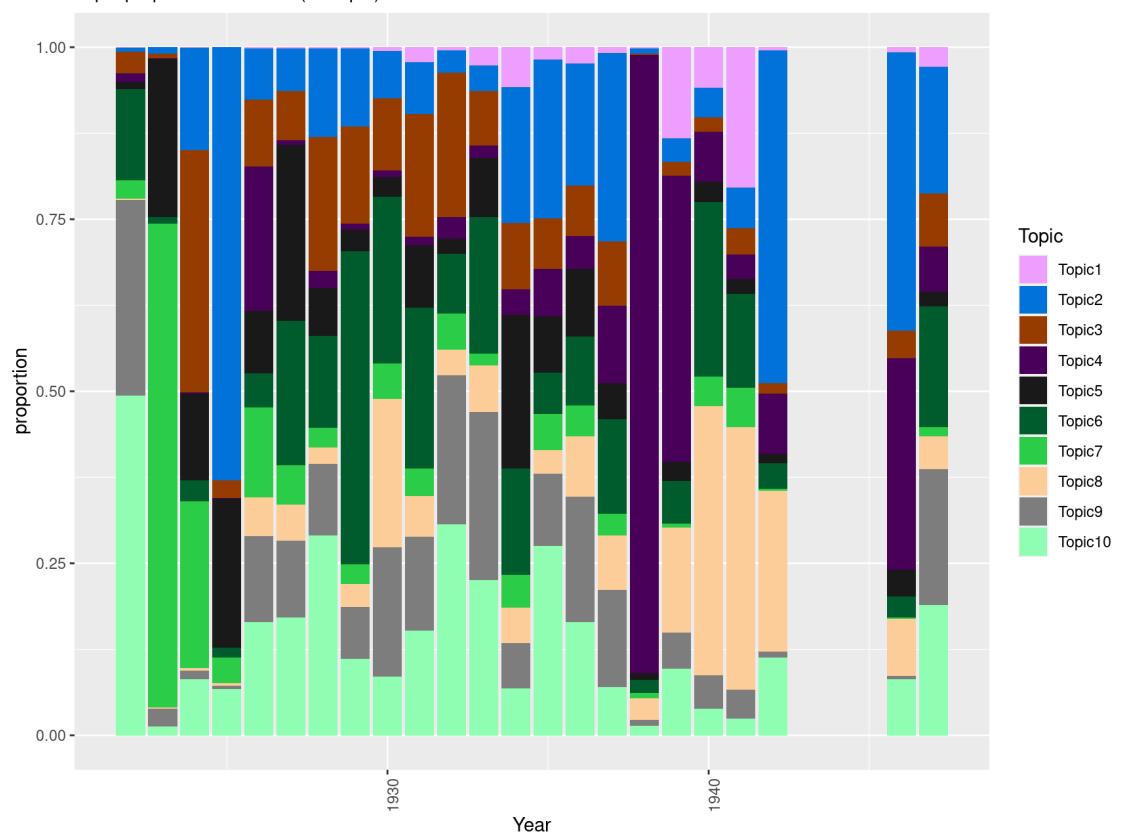
1.4 The public sphere over time

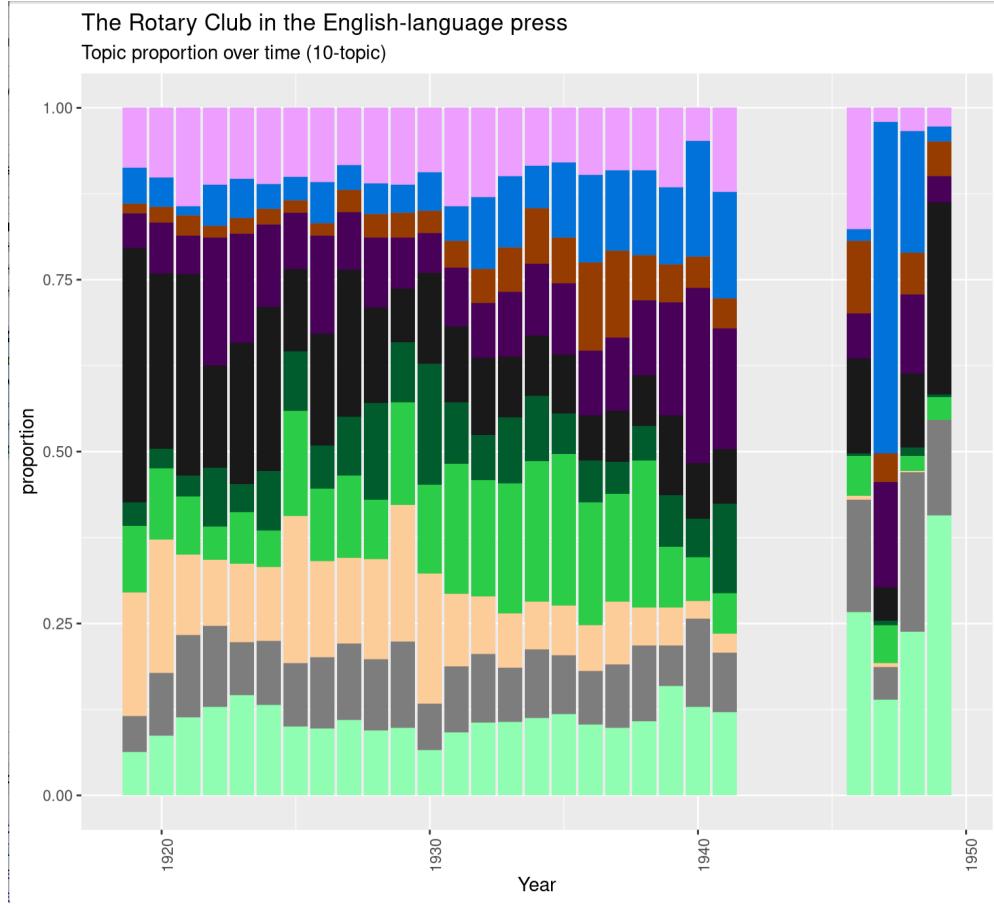
How did the public sphere evolve over time? Did the war reshape the four domains of the public sphere? Did the Shanghai press and the Rotary Club become more globally minded in times of international crisis? For this temporal analysis, we continue to rely on the 10-topic model. In the two corpora, local meetings tended to decline over time, whereas discussions and local philanthropy driven by international conflicts tended to increase during the same period. In the *Shenbao*, three topics in particular experienced a significant decline. These topics refer to special events and social gatherings aimed at entertaining members (topic 4) and guest speakers (topic 3). Topic 5 also includes sports games, especially the Rotary Tennis Cup which was exclusive to the 1930s decade. In the ProQuest corpus, weekly meetings, and special events (topics 6, 7, 8) also declined during the war and the postwar period. Either meetings ceased during the war, or they were no longer systematically reported in the press. On the opposite, topics 2 and 7 experienced a notable increase in the Chinese corpus. Topic 2 reveals that transnational flows of people did not stop during the war. What we learn from this topic is that Rotary clubs in China continued to host visitors from abroad, including Rotary International delegates and fellow Rotarians from other countries. Topic 7 relates to the postwar reorganization and the elections of regional delegates at the district level. In the ProQuest corpus, rising topics essentially relate to children welfare during the war (topic 2), Japan (topic 9), and various lectures about global issues (figure 13).

```
[18]: display_png(file="./media/Fig.13a.png")
       display_png(file="./media/Fig.13b.png")
```

The Rotary Club (扶輪社) in the Shenbao

Topic proportion over time (10-topic)





The remaining topics seem less stable. Topics 1, 6, 8, 9, and 10 in the Chinese corpus are examples of such unstable topics. Most refer to tiffin meetings (announcements or reports), which constitute a regular feature of the Rotary Club and can be defined as a “background topic” for the Rotary Club. Topics dealing with children peaked during the war and vanished afterwards. Topics dealing with the beggar camp and the refugee problem also peaked during the war. Their early appearances are misleading, however, resulting from polysemous terms used to define different topics (such as *xiwang* “hope”).

In the English corpus, temporally unstable topics include topics 1, 3, 4 (social work), and 5. They consist in fact of less consistent topics addressing different issues defined through similar terms. Topic 1 relates to various elections and organizational issues. Topic 3 deals with Rotary clubs outside of Shanghai, appearing mostly in the mid- and late 1930s. Topic 4 addresses a range of issues related to social work and workers, labor strikes in the 1920s, the development of Greater Shanghai and the new Chinese municipality in the 1930s, as well as epidemics and public health administration during the war. Finally, topic 5 focuses on the American community, Sino-American relations, and the coming and going of American visitors in China. This topic emerges as another background topic for several reasons. With the growing influence of the United States in the Pacific region after WWI, an increasing number of American businessmen and officials came to work in China in the 1920s-1940s. After the Sino-Japanese war, the United States reinforced their influence in the region in the context of the Chinese civil war and the Cold War on the global stage. Finally,

this topic reflects the permanence of “People and Events” sections in English-language periodicals, a particular genre of brief news articles focusing on people’s life events. Eventually, from the 1930s onwards, it becomes increasingly difficult to disentangle local and global topics, because of China’s increasing participation in the globalized world, especially during the Sino-Japanese war (1937-1945).

1.5 Conclusion

This paper has contributed to a better understanding of how the Shanghai public sphere functioned in practice in a bilingual environment, and how it evolved during the two last decades of the Republic (1919-1949). In this research, we have relied on topic modeling to map the social and discursive practices of the Rotary Club as a key transnational public sphere institution operating through four main spheres of actions: socializing and entertaining (meeting), self-regulation (organization), community service (philanthropy), lectures and discussions (forum). While this is not an entirely new finding, we are now able to precisely classify and quantify the relative importance of each domain, to model how they differ across languages, and to trace how they changed over time. By extension, we have modeled how the Shanghai press itself operated as a public sphere, not only through the events it reported, but also through its particular way of reporting. More specifically, the topics inferred from the two corpora reveal the predominant use of non-violent words over conflictual and emotional language, supporting previous research that has emphasized the role of the modern press as an arena for the rational handling of conflicts.

Our topic modeling approach further supports previous findings regarding the transnationality of the Shanghai public sphere. As shown in this paper, each of the four spheres identified has a transnational dimension. Foreign guests were regularly invited and entertained at local meetings. Several locally elected officers such as Kuang Fuzhuo had a truly international stature. Local clubs echoed the philosophy of Rotary international and were incorporated into broader organizational units at the district and regional levels. International politics and global issues were vividly discussed during local meetings, while local philanthropy and community service were shaped by transnational movements (boy scouts) and increasingly driven by international conflicts (war, refugees). By extension, the topics also reflect the transnational nature of the Shanghai press itself, featuring multiple points of view and different scales of observations. While the constant intermingling of the local and the global is not an entirely new finding, it has been for the first time empirically grounded in natural language, embodied in the mixture of words defining each of the topics identified in this research.

Cross-lingual comparisons have demonstrated large spheres of intersection between Chinese and English-language periodicals. This general convergence points to the intense circulation of information and journalistic practices within the press industry. The main observed differences result from variations in the periodicity and in the particular history of each publication. The daily *Shenbao* featured short announcements of meetings, whereas the English weeklies published less frequent but more detailed reports of meetings. More subtle divergences reflect different concerns and perspectives. While both corpora reported on Rotary clubs outside Shanghai, the *Shenbao* placed a stronger emphasis on hemispheric circulations and cooperation within the non-Western world. Further investigation is needed to identify more subtle differences in the ways the same topics were presented in different languages, or different views on the same events.

Methodologically, the contribution of this paper is threefold. First, it has offered a simple yet efficient solution, based on concordance, to the problem of article segmentation in digitized newspapers.

This preliminary method needs to be refined and adjusted to different types of texts depending on their varying length and relevance to the research question. Second, this paper constitutes a rare instance of cross-lingual comparison involving the Chinese language during its transitional stage between classical and modern Chinese. As a low-resource language, pre-modern Chinese presents significant challenges for the application of natural language processing tools and computer-assisted text analysis. As (*Kaske 2007*) has demonstrated, the Chinese language was highly unstable during the century under study, which was a pivotal phase in the creation of a standard vernacular (*baihua*) national language (*guoyu*). In his data-driven study of the *Shenbao*, (*Magistry 2019*) has shown that the Chinese language actually evolved through six main stages between 1872 and 1949. Further research should pay greater attention to the decisions made during the pre-processing phase and design strict protocols for evaluating the impact of tokenization and language variation on the resulting topics. From a multilingual perspective, future research could also benefit from more sophisticated techniques for the automatic alignment of topics across languages (*Mimno, Wallach, Naradowsky, Smith, McCallum 2009*). It could also investigate the differences between the various English periodicals included in the ProQuest collection, especially between the British *North China Herald*, the American *China Weekly Review*, and the Chinese-owned *China Press*. Third, this research has demonstrated the value of combining different models with different k number of topics, instead of focusing on a single, definitive model. This multi-model approach is particularly appropriate when dealing with corpora of different sizes and with different structures. This multi-scalar reading of corpora enables scholars to navigate between different levels of granularity and to select in each model the topics that are the most relevant to the research question.

In this paper, we have maintained a critical attitude towards topic modeling. We have pointed out the limitations of topic modeling as a method for mapping complex texts such as historical newspapers. As a statistical model based on word co-occurrences and relying on a rather static definition of words, it may produce incoherent topics and sometimes struggle for handling polysemous words in changing contexts. Topic modeling nonetheless remains an irreplaceable tool for structuring vast collections of texts in multiple languages and to model their changing contents over long periods of time, which is well beyond human reading capability. Eventually, we have adopted a dual approach to topic modeling, oscillating between “topic realism” – equating topics with actual themes - and “topic instrumentalism” – treating topics as proxies to gain a better understanding of digitized artifacts and to evaluate their potential biases on historical interpretations (*Pääkkönen, Ylikoski 2021*). We concur with previous literature who considers that topic modeling is best used as an exploratory tool and in combination with other methods. As shown in this paper, topic modeling is never fully automated. To properly label, interpret, and validate the topics, we have relied on the close reading of representative articles, informed by our previous archival research on the Rotary Club. This external knowledge has also helped us to highlight some important topics that went missing or were underrepresented in the press, such as Sino-Japanese tensions and language barriers to the inclusion of Chinese members in the organization.

START HERMENEUTICS The results of this topic modeling exercise can be used as a starting point for addressing more specific research questions. The inferred topics point to the existence of two main categories of articles that can be further investigated using adequate methods. On the one hand, topics related to meetings, organization, and philanthropy are generally rich with names of individuals, organizations, and locations. Named entity recognition (NER) and network analysis can then be utilized to automatically extract the names of these actors and further analyze their connections. On the other hand, topics related to lectures and discussions (forums), which are richer in semantic contents, lend themselves to a deeper examination of the discourses articulated by the various actors, using methods such as semantic and sentiment analyses. Finally, while

the Rotary Club has served as a test case in this paper, our methodology can be expanded to investigate other public sphere institutions and more abstract concepts related to the public sphere. Furthermore, with some adjustments, it can be readily transposed to similar digitized texts in English, Chinese, and possibly other languages, beyond the specific corpora utilized in this research.

END HERMENEUTICS

1.6 Appendix

START HERMENEUTICS In the following tables, *SB* stands for *Shenbao*, *PQ* = ProQuest. 5T = 5-topic model, 10T = 10-topic model, 20T = 20-topic model. The number immediately following indicates the number of the topic in each topic model.

END HERMENEUTICS

START HERMENEUTICS Tables 1 to 6 provide a summary of the topics for each model, including their label, the 10 most frequent words defining each topic, and their various attributes (topical group, dimension, proportion, and trend over time). Tables 7 and 8 display the topics aligned across models. Tables 9, 10, and 11 display the topics aligned across languages.

END HERMENEUTICS

1.6.1 Summary of topic models

START HERMENEUTICS

TopicCo	TopicLabel	TopWords	Proportion	Proportional	Functional	Trend
SB5T01	Children	, , , , , , , , ,	0.129	13%	Local	SPONSOR Increase (dramatic)
SB5T02	International	, , , , , , , , ,	0.226	23%	Non-local	ORGANIZATION
SB5T03	All Nations	, , , , , , , , ,	0.201	20%	Non-local	MEETING Increase (moderate)
SB5T04	Early meetings (1920s)	, , , , , , , , ,	0.239	24%	Local	MEETING Increase (moderate)
SB5T05	Wartime meetings	, , , , , , , , ,	0.205	21%	Local	MEETING Decline (dramatic)

END HERMENEUTICS

_____ | | **PQ5T01** | Past tiffins (reports) | shanghai, president, tiffin, members, meeting, held, yesterday, hotel, weekly, rotarians | 0.209 | 21% | Local | MEETING | Stable | | **PQ5T02** | Children | shanghai, children, hospital, toys, made, school, christmas, work, committee, russian | 0.154 | 15% | Local | SPONSOR | Increase (dramatic) | | **PQ5T03** | Lectures | nanking, address,

foreign, said, subject, shanghai, delivered, speech, told, government | 0.213 | 21% | Non-local
FORUM | Increase (moderate) | | **PQ5T04** | Boy scouts | shanghai, first, said, international,
world, shield, present, work, local, great | 0.168 | 17% | Local | SPONSOR | Decrease (moderate)
| | **PQ5T05** | Upcoming tiffins (announcements) | shanghai, meeting, american, held, hotel,
thursday, members, weekly, next, today | 0.256 | 26% | Local | MEETING | Decrease (dramatic) |
END HERMENEUTICS

START HERMENEUTICS	TopicCode	TopicLabel	TopWords	Proportion	Proportion%
Functions	Local	Trend			
(Metropole)	,	,	,	0.157	15.7%
International	,	,	,	0.077	7.7%
Speeches	,	,	,	0.081	8.1%
Special events	,	,	,	0.094	9.4%
Tennis Cup	,	,	,	0.115	11.5%
MEETING	Local	Decrease (moderate)			
Children	,	,	,		
SPONSOR	Local	Increase (dramatic)			
District elections	,	,	,		
ORGANIZATION	Non-local	Increase (dramatic)			
Beggar camp	,	,	,		
Wartime meetings	,	,	,		
END HERMENEUTICS					

START HERMENEUTICS	TopicCode	TopicLabel	TopWords	Proportion	Proportion%
Functions	Local	Trend			
	PQ10T01	Elections	shanghai, president, member, fong, international, secretary, board, elected, past, harris	0.101 10.1% Mixed ORGAN Decrease (moderate)	PQ10T02
		Children	children, hospital, toys, shanghai, christmas, made, funds, charity, year, building	0.087 8.7% Local SPONSOR Increase (dramatic)	PQ10T03
		Non-Shanghai Clubs	nanking, wang, international, foreign, hangchow, district, affairs, president, special, minister	0.067 6.7% Non-Local ORGAN Increase (dramatic)	PQ10T04
		Social work and workers	work, shanghai, public, committee, community, interest, relief, church, municipal, service	0.099 9.9% Local FORUM Increase (moderate)	PQ10T05
		American community	american, shanghai, states, united, university, america, addressed, foreign, company, hongkong	0.107 10.7% Mixed MEET Decrease (dramatic)	PQ10T06
		Boy scouts	school, shanghai, miss, russian, shield, scouts, ball, troop, presented, boys	0.078 7.8% Local SPONSOR Decrease (moderate)	PQ10T07
		Weekly meetings	meeting, hotel, held, shanghai, weekly, thursday, tiffin, metropole, speaker, yesterday	0.162 16.2% Local MEET Increase (moderate)	PQ10T08
		Special events	members, tiffin, guests, shanghai, house, astor, rotarians, party, held, dinner	0.1 10% Local MEET Decrease (dramatic)	PQ10T09
		Sino-Japanese relations	address, japan, japanese, government, peking, delivered, last, members, present, international	0.093 9.3% Non-Local FORUM Stable	PQ10T10
		Lectures	said, shanghai, world, years, address, speech, gave, talk, members, editor	0.105 10.5% Non-Local FORUM Stable	

END HERMENEUTICS

| 01 - Special events | , , , , , , , | 0.045 | 4.5% | Mixed | MEETING | Decline
(dramatic) || **SB20T02** | 02 - International | , , , , , , , | 0.044 | 4.4% | Non-local
| ORGANIZATION | Stable || **SB20T03** | 03 - Meetings (brief) | , , , , , , , |
0.049 | 4.9% | Local | MEETING | Increase (dramatic) || **SB20T04** | 04 - Meetings (program) | ,
, , , , , , | 0.13 | 13% | Local | MEETING | Increase (dramatic) || **SB20T05** | 05-
Tennis Cup | , , , , , , , | 0.046 | 4.6% | Local | MEETING | Decline (moderate) ||
SB20T06 | 06 - Children (Entertain) | , , , , , , , | 0.026 | 2.6% | Local | SPONSOR
| Increase (dramatic) || **SB20T07** | 07 - District elections | , , , , , , , | 0.025 | 2.5%
| Non-local | ORGANIZATION | Increase (dramatic) || **SB20T08** | 08 - Early meetings (1923) | ,
, , , , , , | 0.084 | 8.4% | Local | MEETING | Decline (dramatic) || **SB20T09** | 09

END HERMENEUTICS

START HERMENEUTICS	TopicCode	TopicLabel	TopWords	Proportion	Proportion%
Local	Functions	Trend			
	PQ20T01	01 - Organization	shanghai, president, international, past, fong, member, local, harris, fitch, george	0,04	4%
	PQ20T02	02 - Children	hospital, toys, children, shanghai, christmas, charity, made, year, funds, ward	0,055	6%
	PQ20T03	03 - US & East Asia	foreign, trade, arnold, week, addresses, julean, various, commercial, members, commissioner		
	PQ20T04	04 - Social work and workers	0,035	4%	
	PQ20T05	05 - American community	Non-local FORUM Increase (dramatic)		
	PQ20T06	06 - Boy scouts	Local FORUM Increase (moderate)		
	PQ20T07	07 - Blind school	0,051	5%	
	PQ20T08	08 - Weekly tiffins (Metropole)	american, shanghai, states, united, company, member, addressed, thursday, university, commerce		
	PQ20T09	09 - Meetings announcements	0,063	6%	
	PQ20T10	10 - Speeches (not valid)	Mixed MEETING Decrease (dramatic)		
			shield, scouts, troop, russian, shanghai, president, jamboree, camp, scout	0,031	3%
			Local SPONSOR Decrease (dramatic)		
			PQ20T07 07 - Blind school shanghai, road, chang, institution, issue, work, building, charge, official, appeal	0,038	4%
			Local SPONSOR Decrease (dramatic)		
			PQ20T08 08 - Weekly tiffins (Metropole) hotel, shanghai, meeting, held, weekly, metropole, thursday, members, yesterday, speaker	0,095	10%
			Local MEETING Increase (dramatic)		
			PQ20T09 09 - Meetings announcements meeting, today, program, speak, regular, held, next, closed, tomorrow, shang	0,048	5%
			Local MEETING Increase (moderate)		
			PQ20T10 10 - Speeches (not valid) said, hongkong, recently, well, talk, pacific,		

institute, great, read, speech | 0,05 | 5% | Non-local | OTHER | Increase (moderate) | | **PQ20T11**
| 11 - Special events | members, shanghai, dinner, evening, given, party, clock, ladies, night, afternoon | 0,054 | 5% | Local | MEETING | Decrease (dramatic) | | **PQ20T12** | 12 - District conference | wang, conference, district, hangchow, governor, members, international, president, tsinan, held | 0,038 | 4% | Non-local | ORGANIZATION | Increase (dramatic) | | **PQ20T13** | 13 - Nanking Rotary Club | nanking, government, national, bureau, affairs, soochow, special, central, capital, minister | 0,047 | 5% | Non-local | ORGANIZATION | Increase (dramatic) | | **PQ20T14** | 14 - Opinions (not valid) | address, shanghai, delivered, present, last, speech, editor, years, subject, history | 0,076 | 8% | Non-local | OTHER | Increase (dramatic) | | **PQ20T15** | 15 - Poor Russians | miss, shanghai, french, school, russian, children, girls, society, race, donation | 0,044 | 4% | Local | SPONSOR | Increase (dramatic) | | **PQ20T16** | 16 - Games/Competitions | shanghai, meet, international, first, radio, local, team, tennis, american, presented | 0,038 | 4% | Local | MEETING | Decrease (moderate) | | **PQ20T17** | 17 - Weekly tiffins (Astor House) | tiffin, meeting, guests, held, address, members, gave, weekly, yesterday, interesting | 0,067 | 7% | Local | MEETING | Decrease (dramatic) | | **PQ20T18** | 18 - Tientsin Rotary Club | tientsin, left, peking, church, meeting, house, astor, union, service, shanghai | 0,04 | 4% | Non-local | OTHER | Stable | | **PQ20T19** | 19 - Addresses (not valid) | japanese, life, rotarians, rotarian, business, members, years, first, great, said | 0,044 | 4% | Non-local | OTHER | Increase (moderate) | | **PQ20T20** | 20 - Local elections | president, general, elected, secretary, board, shanghai, year, meeting, directors, chairman | 0,046 | 5% | Local | ORGANIZATION | Decrease (moderate) |

END HERMENEUTICS

1.6.2 Cross-model alignment

START HERMENEUTICS Tables 7 and 8 display the topics aligned across models.

END HERMENEUTICS

END HERMENEUTICS

| **PQ20T08** | 08 - Weekly tiffins (Metropole) | | **PQ5T01** | Past tiffins (reports) | **PQ10T07** |
 Weekly meetings | **PQ20T17** | 17 - Weekly tiffins (Astor House) | | **PQ5T02** | Children | **PQ10T02**
 | Children | **PQ20T02** | 02 - Children | | **PQ5T03** | Lectures | **PQ10T09** | Sino-Japanese relations
 | **PQ20T19** | 19 - Addresses (not valid) | | **PQ5T03** | Lectures | **PQ10T10** | Lectures | **PQ20T10**
 | 10 - Speeches (not valid) | | **PQ5T04** | Boy scouts | **PQ10T06** | Boy scouts | **PQ20T06** | 06
 - Boy scouts | | **PQ5T05** | Upcoming tiffins (announcements) | **PQ10T07** | Weekly meetings |
PQ20T09 | 09 - Meetings announcements | | | | **PQ10T01** | Elections | **PQ20T01** | 01 - Orga-
 nization | | | | **PQ10T01** | Elections | **PQ20T12** | 12 - District conference | | | | **PQ10T01** |
 Elections | **PQ20T20** | 20 - Local elections | | | | **PQ10T03** | Non-Shanghai Clubs | **PQ20T13**
 | 13 - Nanking Rotary Club | | | | **PQ10T03** | Non-Shanghai Clubs | **PQ20T18** | 18 - Tientsin
 Rotary Club | | | | **PQ10T04** | Social work and workers | **PQ20T04** | 04 - Social work and workers
 | | | | **PQ10T05** | American community | **PQ20T05** | 05 - American community | | | | **PQ10T08**
 | Special events | **PQ20T11** | 11 - Special events | | | | **PQ20T03** | 03 - US & East Asia | | | |
 | **PQ20T07** | 07 - Blind school | | | | **PQ20T14** | 14 - Opinions (not valid) | | | | **PQ20T15**
 | 15 - Poor Russians | | | | **PQ20T16** | 16 - Games/Competitions | END HERMENEUTICS

1.6.3 Cross-language alignment

START HERMENEUTICS | CHINESE PRESS (SHENBAO) | | ENGLISH PRESS (PRO-
 QUEST) | | | :-----:-:-----:-:-----:-:-----:
 :-| SB5T01 | Children | PQ5T02 | Children | | SB5T02 | International | | | | SB5T03 | All
 Nations | | | | SB5T04 | Early meetings (1920s) | PQ5T01 | Past tiffins (reports) | | SB5T04 |
 Early meetings (1920s) | PQ5T05 | Upcoming tiffins (announcements) | | SB5T05 | Wartime meet-
 ings | PQ5T01 | Past tiffins (reports) | | SB5T05 | Wartime meetings | PQ5T05 | Upcoming tiffins
 (announcements) | | | | PQ5T03 | Lectures | | | | PQ5T04 | Boy scouts | END HERMENEUTICS

START HERMENEUTICS | CHINESE PRESS (SHENBAO) | | ENGLISH PRESS (PROQUEST)
 | | | :-----:-:-----:-:-----:-:-----:-| | SB10T06 |
 Children | PQ10T02 | Children | | SB10T07 | District elections | PQ10T01 | Elections | | SB10T04 |
 Special events | PQ10T08 | Special events | | SB10T010 | Tennis Cup | PQ10T08 | Special events |
 | SB10T08 | Early meetings (1923) | PQ10T07 | Weekly meetings | | SB10T01 | Tiffins (Metropole)
 | PQ10T07 | Weekly meetings | | SB10T10 | Wartime meetings | PQ10T07 | Weekly meetings | |
 SB10T03 | Speeches | PQ5T10 | Lectures | | SB10T03 | Speeches | PQ10T09 | Sino-Japanese rela-
 tions | | SB10T02 | International | PQ10T09 | Sino-Japanese relations | | SB10T09 | Beggar camp |
 PQ10T04 | Social work and workers | | SB10T02 | International | PQ10T05 | American community
 | | | | PQ10T03 | Non-Shanghai Clubs | | | | PQ10T06 | Boy scouts | END HERMENEUTICS

START HERMENEUTICS | CHINESE PRESS (SHENBAO) | | ENGLISH PRESS (PRO-
 QUEST) | | | :-----:-:-----:-:-----:-:-----:
 -----| | SB20T01 | 01 - Special events | PQ20T11 | 11 - Special events | | SB20T02 | 02 - International
 | PQ20T01 | 01 - Organization | | SB20T03 | 03 - Meetings (brief) | PQ20T09 | 09 - Meetings
 announcements | | SB20T04 | 04 - Meetings (program) | PQ20T08 | 08 - Weekly tiffins (Metropole)
 | | SB20T05 | 05 - Tennis Cup | PQ20T16 | 16 - Games/Competitions | | SB20T06 | 06 - Children
 (Entertain) | PQ20T02 | 02 - Children | | SB20T07 | 07 - District elections | PQ20T12 | 12 - District
 conference | | SB20T07 | 07 - District elections | PQ20T20 | 20 - Local elections | | SB20T08 | 08
 - Early meetings (1923) | PQ20T09 | 09 - Meetings announcements | | SB20T09 | 09 - Boy Scouts
 | PQ20T06 | 06 - Boy scouts | | SB20T10 | 10 - Poor Russians | PQ20T15 | 15 - Poor Russians |
 | SB20T11 | 11 - Not valid | PQ20T10 | 10 - Speeches (not valid) | | SB20T11 | 11 - Not valid |
 PQ20T14 | 14 - Opinions (not valid) | | SB20T11 | 11 - Not valid | PQ20T19 | 19 - Addresses (not

valid) | | SB20T12 | 12 - Children (Hospital) | PQ20T02 | 02 - Children | | SB20T13 | 13 - Rotary abroad () | | | | SB20T14 | 14 - Speeches | | | | SB20T15 | 15 - Tianjin Rotary | PQ20T18 | 18 - Tientsin Rotary Club | | SB20T16 | 16 - Tiffins (Carlton) | PQ20T17 | 17 - Weekly tiffins (Astor House) | | SB20T17 | 17 - Beggar Camp | | | | SB20T18 | 18 - Hangzhou Rotary | PQ20T13 | 13 - Nanking Rotary Club | | SB20T19 | 19 - Peace resolution | | | | SB20T20 | 20 - Rotary abroad () | | | | | PQ20T03 | 03 - US & East Asia | | | | PQ20T04 | 04 - Social work and workers | | | | PQ20T05 | 05 - American community | | | | PQ20T07 | 07 - Blind school | | | | PQ20T13 | 13 - Nanking Rotary Club | END HERMENEUTICS

1.7 Bibliography

Allen, C., et al.. “Topic Modeling the *Hàn Diǎn* Ancient Classics (handian Guji).”. Journal of Cultural Analytics, vol. 1, no. 2, p. 11882.

Andersen, E., et al.. “How to Read the 52.000 Pages of the British Journal of Psychiatry? A Collaborative Approach to Source Exploration”. Journal of Data Mining and Digital Humanities, vol. 2020, pp. 1–4.

Armand, C.. “Foreign Clubs with Chinese Flavor: The Rotary Club of Shanghai and the Politics of Language”. Knowledge, Power, and Networks: Elites in Transition in Modern China, edited by Cécile Armand et al., Brill, pp. 233–59.

Armand, C.. La Publicité En Chine: Espaces Et Pouvoirs (1905-1949) [between the Public and the Market: The Politics of Advertising in Modern China 1905-1949)]. De Gruyter.

Bunout, E.. “Nibbling at Text: Identifying Discourses on Europe in a Large Collection of Historical Newspapers Using Topic Modelling”. C2DH | Luxembourg Centre for Contemporary and Digital History, <https://www.c2dh.uni.lu/thinkering/nibbling-text-identifying-discourses-europe-large-collection-historical-newspapers-using>.

Deng Zhenglai , ., editor. *Guojia yu shehui: zhongguo shimin shehui yanjiu* (state and society: research on Chinese civil society). Edited by Deng Zhenglai , Beijing Book Co. Inc..

Eads, A., et al.. “Separating the Wheat from the Chaff: A Topic and Keyword-based Procedure for Identifying Research-relevant Text”. Poetics, vol. 86, pp. 1–9, <https://doi.org/10.1016/j.poetic.2020.101527>.

Goodman, B.. “Semi-colonialism, transnational networks and news flows in early republican Shanghai”. China review, vol. 4, no. 1, pp. 55–88.

Hengchen, S., et al.. “Exploring Archives with Probabilistic Models: Topic Modelling for the Valorisation of Digitised Archives of the European Commission”. 2016 IEEE International Conference on Big Data (big Data), pp. 3245–49.

Huang, P. C. C.. ““public sphere”/“civil society” in China? the third realm between state and society”. Modern China modern China, vol. 19, no. 2, pp. 216–40.

Indukaev, A.. “Studying Ideational Change in Russian Politics with Topic Models and Word Embeddings”. The Palgrave Handbook of Digital Russia Studies, edited by Daria Gritsenko et al., Springer, pp. 443–64.

Judge, J.. Print and politics: ‘shibao’ and the culture of reform in late Qing China. Stanford University Press.

- Kaske, E.. *The politics of language in chinese education, 1895-1919*. Brill.
- Liu Li , .. *Zhongguo jindai baoye caifang shilun: yi shenbao wei zhongxin de kaocha* : (histoire du journalisme moderne en chine : une étude centrée sur le shenbao). Anhui daxue chubanshe .
- Magistry, P.. “Languages(s) of the Shun-pao, a Computational Linguistics Account”. 10th International Conference of Digital Archives and Digital Humanities.
- Maier, D., et al.. “Machine Translation Vs. Multilingual Dictionaries Assessing Two Strategies for the Topic Modeling of Multilingual Text Collections”. *Communication Methods and Measures*, vol. 16, no. 1, pp. 19–38, <https://doi.org/10.1080/19312458.2021.1955845>.
- Marjanen, J., et al.. “Topic Modelling Discourse Dynamics in Historical Newspapers”. Arxiv:2011.10428 [cs], <http://arxiv.org/abs/2011.10428>.
- Metzger, T. A.. The Western Concept of the Civil Society in the Context of Chinese History. Hoover Institution on War, Revolution and Peace, Stanford University.
- Miller, I. M.. “Rebellion, Crime and Violence in Qing China, 1722–1911: A Topic Modeling Approach”. *Poetics*, vol. 41, no. 6, pp. 626–49.
- Mimno, D., et al.. “Polylingual Topic Models”. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2 - Volume 2, Association for Computational Linguistics, pp. 880–89.
- Mittler, B.. *A Newspaper for China? Power, Identity, and Change in Shanghai’s News Media, 1872-1912*. Harvard University Press.
- Nelson, L. K.. “Computational grounded theory: A methodological framework”. *Sociological methods & research* sociological methods & research, vol. 49, no. 1, pp. 3–2.
- Oberbichler, S.and E. Pfanzelter. “Topic-specific Corpus Building: A Step Towards a Representative Newspaper Corpus on the Topic of Return Migration Using Text Mining Methods”. *Journal of Digital History*, vol. 1, no. 1, <https://journalofdigitalhistory.org/en/article/4yxHGiqXYRbX>.
- Pääkkönen, J.and P. Ylikoski. “Humanistic Interpretation and Machine Learning”. *Synthese*, vol. 199, no. 1, pp. 1461–97, <https://doi.org/10.1007/s11229-020-02806-w>.
- Rankin, M. B.. “The Origins of a Chinese Public Sphere. Local Elites and Community Affairs in the Late Imperial Period”. *Études Chinoises*, vol. 9, no. 2, pp. 13–60.
- Roberts, M. E., et al.. “Stm: An R Package for Structural Topic Models”. *Journal of Statistical Software*, vol. 91, pp. 1–0, <https://doi.org/10.18637/jss.v091.i02>.
- Roberts, M., et al.. Stm: Estimation of the Structural Topic Model. 1.3.6, <https://CRAN.R-project.org/package=stm>.
- “Rotary’s Progress in China”. “Rotary’s Progress in China”. *The North China Herald*, p. 343.
- Rowe, W. T.. “The public sphere in modern china”. *Modern china*, vol. 16, no. 3, pp. 309–29.
- “Shimin Shehui Yu Guojia: Xueli Shang De Fenyel Liang Zhong Jiagou – (civil Society and China: Theoretical Divide and Dual Structure)”. “Shimin Shehui Yu Guojia: Xueli Shang De Fenyel Liang Zhong Jiagou – (civil Society and China: Theoretical Divide and Dual Structure)”. *Zhongguo Shehuikexue Jikan* , no. 3, pp. 5–2.

- Strand, D.. “Protest in Beijing: Civil Society and Public Sphere in China”. Problems of Communism, vol. 39, p. 1, <https://heinonline.org/HOL/Page?handle=hein.journals/probscmu39&id=245&div=&collection=>.
- Tang Xiaobing , .. Xiandai zhongguo de gonggong yulun: yi dagongbao “xingqi lunwen” he shenbao “ziyoutan” wei li : “ ”“ ”“ ”“ ” (public opinion in modern china: dagongbao’s “weekly essay” and shenbao’s “free conversation” columns). Sheke wenxian chubanshe.
- Tsai, W.. Reading shenbao: nationalism, consumerism and individuality in china, 1919-37. Palgrave Macmillan.
- Van Galen, Q.and B. Nicholson. “In Search of America”. Digital Journalism, vol. 6, no. 9, pp. 1165–85, <https://doi.org/10.1080/21670811.2018.1512879>.
- Veg, S.. “Creating Public Opinion, Advancing Knowledge, Engaging in Politics: The Local Public Sphere in Chengdu, 1898–1921”. The China Quarterly, vol. 246, pp. 331–53, <https://doi.org/10.1017/S0305741021000217>.
- Wagner, R. G., editor. Joining the global public: word, image, and city in early chinese newspapers, 1870-1910. Edited by Rudolf G Wagner, State University of New York Press.
- Wagner, R. G.. “The Role of the Foreign Community in the Chinese Public Sphere”. The China Quarterly, no. 142, pp. 423–43, <https://www.jstor.org/stable/655422>. JSTOR.
- Wakeman, F.. “The Civil Society and Public Sphere Debate: Western Reflections on Chinese Political Culture”. Modern China, vol. 19, no. 2, pp. 108–38. JSTOR.
- Yang, T.-I., et al.. “Topic Modeling on Historical Newspapers”. Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, Association for Computational Linguistics, pp. 96–104, <https://aclanthology.org/W11-1513>.
- Yang, W., et al.. “A Multilingual Topic Model for Learning Weighted Topic Links Across Corpora with Low Comparability”. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, pp. 1243–48, <https://doi.org/10.18653/v1/D19-1120>.
- Zosa, E.and M. Granroth-Wilding. “Multilingual Dynamic Topic Model”. Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019), INCOMA Ltd., pp. 1388–96, https://doi.org/10.26615/978-954-452-056-4_159.